Handbooks in Finance

HANDBOOK of FINANCIAL MARKETS
DYNAMICS AND EVOLUTION

Thorsten Hens
Klaus Reiner Schenk-Hoppé
Editors

North-Holland

North-Holland is an imprint of Elsevier

Radarweg 29, PO Box 211, 1000 AE
Amsterdam, The Netherlands

Linacre House, Jordan Hill
Oxford, OX2 8DP, UK

30 Corporate Drive, Suite 400
Burlington, MA 01803, USA

525 B Street, Suite 1900
San Diego, California 92101-4495, USA

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Application submitted.

ISBN: 978-0-12-374258-2

For information on all Elsevier publications,
visit our Web site at *www.books.elsevier.com*.

Printed in the United States of America
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

List of Contributors

Larry Blume, Department of Economics, Cornell University, Ithaca, NY 14850, USA; lb19@cornell.edu

Jean-Philippe Bouchaud, Science & Finance, Capital Fund Management, 6 Blvd Haussmann, 75009 Paris, France; jean-philippe.bouchand@cea.fr

Carl Chiarella, School of Finance and Economics, University of Technology-Sydney, Broadway NSW 2007, Australia, Carl.Chiarella@uts.edu.au

Roberto Dieci, Dipartimento di Matematica per le Scienze Economiche e Sociali, University of Bologna, Bologna, Italy; rdieci@rimini.unibo.it

David Easley, Department of Economics, Cornell University, Ithaca, NY 14850, USA; dae3@cornell.edu

Igor V. Evstigneev, Economic Studies, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK; Igor.Evstigneev@manchester.ac.uk

J. Doyne Farmer, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA; jdf@santafe.edu

Xue-Zhong He, School of Finance and Economics, University of Technology-Sydney, Broadway, NSW 2007, Australia; Tony.He@uts.edu.au

Thorsten Hens, Swiss Banking Institute, University of Zurich, CH-8032 Zurich, Switzerland; thens@isb.uzh.ch

David Hirshleifer, The Paul Merage School of Business, University of California, Irvine, Irvine, CA 92697, USA; David.H@usi.edu

Cars Hommes, CeNDEF, University of Amsterdam, NL-1018 WB Amsterdam, The Netherlands; C.H.Hommes@uva.nl

Mordecai Kurz, Department of Economics, Stanford University, Stanford, CA 94305, USA; mordecai@stanford.edu

Fabrizio Lillo, Dipartimento di Fisica e Tecnologie Relative, University of Palermo, 90128 Palermo, Italy; lillo@lagash.dft.unipa.it

Thomas Lux, Department of Economics, University of Kiel, D-24118, Kiel, Germany; lux@bwl.uni-kiel.de

Klaus Reiner Schenk-Hoppé, The University of Leeds, Business School and School of Mathematics, Leeds, LS2 9JT, UK; K.R.Schenk-Hoppe@leeds.ac.uk

Siew Hong Teoh, The Paul Merage School of Business, University of California–Irvine, Irvine, CA 92697, USA; steoh@uci.edu

Florian Wagener, CeNDEF, University of Amsterdam, NL-1018 WB Amsterdam, The Netherlands; F.O.O.Wagener@uva.nl

Jan Wenzelburger, Economic and Management Studies, Keele University, Staffs, ST5 5BG, UK; j.wenzelburger@econ.keele.ac.uk

Preface

The aim of this handbook is to provide readers with an overview of cutting-edge research on the dynamics and evolution of financial markets. While the insights offered in this book will be valuable for the future development of finance theory, we are convinced they are also of vital importance to today's financial practitioners. All chapters are written exclusively for this handbook with the goal of being accessible to the nonspecialist reader, may they be asset managers or researchers from other disciplines.

The view of financial markets promoted here goes far beyond traditional finance approaches to asset management. The classic credo is still to buy and hold a market portfolio or, in more sophisticated versions, to place bets on the convergence of asset prices to some equilibrium. In contrast, the models presented in this book aim to explain the market dynamics of asset prices based on the heterogeneity of investors. This can offer insights for asset management approaches including market timing, which is potentially very fruitful but also very difficult without a clear understanding of the various interactions in a financial market.

Although this handbook is not the only work in finance highlighting the importance of dynamics and heterogeneity for financial markets, it is unique because it is the most recent and most encompassing account of this literature. Other important contributions to the general theme are, for example, Shefrin's excellent book, *A Behavioral Approach to Asset Pricing*, and Volume 2 of the *Handbook of Computational Economics* edited by Tesfatsion and Judd, both published by Elsevier. As compared to this one, the two other contributions have a different focus; however, Shefrin's book is "less dynamic" because it is fully based on general equilibrium, and the Tesfatsion and Judd chapters of the book dealing with finance focus more on illustrating the dynamics of heterogeneous agents models by computational simulations.

The importance of this work for the development of finance theory is best explained by contrasting it to the main paradigm in finance: optimization and rational expectations as theoretical underpinnings of the efficient market hypothesis. The prevalent view of traditional finance is that of any point in time all traders make use of all available information; and as a consequence, any predictable pattern, such as a price trend must already be anticipated and reflected in current prices. Only the arrival of new information can lead to price changes. In 1964 Cootner formulated the conjecture that period-by-period price changes are random movements statistically independent of each other. This stochastic price mechanism is at the heart of many of the key theoretical models in finance such as optimal portfolio rules inspired by the work of Markowitz and Merton from 1952 onward; the static and intertemporal capital asset pricing models of Sharpe, Lintner, Mossin, and Merton from the 1960s; and models for the pricing of contingent claims beginning in the 1970s with the work of Black and Scholes.

The two main theoretical justifications of the traditional finance view—optimization and rational expectations—have been under heavy attack for some time and will clearly not emerge unscathed. One may argue that people think twice when money is involved, coming to a conclusion that is void of any biases or mistakes. Empirical evidence for this view is weak at best. To the contrary, high monetary gains (or losses) are often observed to trigger emotions that severely distort traders' decisions. The second argument is that the market itself will take care of irrational behavior and erase it through the force of market selection. This conjecture—made by Cootner, Friedman, and Fama—is challenged on the basis of theoretical and practical work (see, for example, Blume and Easley's contribution, "Market Competition and Selection," to the *New Palgrave Dictionary of Economics*). Work by Shleifer and others has shown that too many irrational investors are a risk to rational investors because they cannot be "arbitraged away," at least in the short run. One of the contributions of this handbook is to show the state of the current debate on the market-selection hypothesis—a debate that still has not come to a definite conclusion.

Recent empirical and experimental work challenged the traditional view of efficient markets and the long-sustained belief in market rationality; see, for example, the excellent surveys on asset pricing in the *Journal of Finance* by Campbell (2000) and Hirshleifer (2001). Indeed a new paradigm based on behavioral models of decision under risk and uncertainty is beginning to crowd out the traditional view based on complete rationality of all market participants. The traditional and the behavioral finance models, however, share one important feature: They are both based on the notion of a representative agent—although this mythological figure is dressed differently. While traditionally he had rational preferences, expectations, and beliefs, he is currently a prospect theory maximizer, unable to carry out Bayesian updating and likely to fall into framing traps.

The chapters in this book, in contrast, suggest models of portfolio selection and asset price dynamics that are explicitly based on the idea of heterogeneity of investors. They are descriptive and normative as well, answering which set of strategies one would expect to be present in a market and how to find the best response to any such market. The models presented are successful as a descriptive approach because they are able to explain facts of asset prices such as fat tails in the return distribution, stochastic and clustered volatility, and bubbles and crashes—facts that are anomalies or puzzles in the traditional finance world. On the second issue, the main observation is that there is nothing like "the" best strategy because the performance of any strategy will depend on all strategies in the market. Rationality therefore is to be seen as conditional on the market ecology. The key to investment success, thus, is understanding the interaction of the various strategies.

This handbook has nine chapters on topics within the emerging field of dynamics and evolution in financial markets. They aim to explore the preceding ideas in consistent and adequate models with the goal of contributing to a better understanding of the dynamics of financial markets. The collection of chapters reflects the diversity of evolutionary approaches in terms of both conceptual and methodological aspects. On the conceptual level, readers will be exposed to several different modeling approaches: temporary and general equilibrium models are considered; dynamic systems theory, as well as game-theoretic reasoning, is applied; traders' behavior originates from expected utility maximization, genetic learning, or is only restricted by being adapted to the information filtration; and fundamentalists and noise traders also enter the stage. On the methodological level, readers will see analytical, empirical, and numerical techniques applied by this book's authors. In the best tradition of the *Handbooks in Finance* series, all chapters share a thorough and formal treatment of the issue under consideration. As with every growing area of research, we expect to see further progress and fruitful applications in this exciting field.

Chapter 1, "Thought and Behavior Contagion in Capital Markets" by David Hirshleifer and Siew Hong Teoh, surveys more than 200 theoretical and empirical papers that emphasize the social interaction of traders. The authors argue that the analysis of thought contagion and the evolution of financial ideologies, and their effects on markets, is a missing chapter in modern finance, including behavioral finance. While financial practitioners always emphasize that their decisions are influenced not only by fundamentals and price movements but also by opinions expressed (e.g., in the media), theorists have done little to provide them with models that can check the consistency of these claims, and moreover help them to better understand in which direction financial markets might move. Given the progress in information technology that allows researchers to categorize and to rapidly put into context any piece of news, we can expect profitable trading strategies to evolve from the novel research on behavior contagion in capital markets.

Chapter 2, "How Markets Slowly Digest Changes in Supply and Demand" by Jean-Philippe Bouchaud, J. Doyne Farmer, and Fabrizio Lillo, is a beautiful piece on the market microstructure of financial markets that makes the "econophysics" approach accessible to a wide audience. It exemplifies how natural scientists do research: In contrast to economics and finance, observations are more important than theories. It is argued convincingly that the standard dichotomy of finance between informed and uninformed traders can neither be supported empirically nor is it useful theoretically because it leads to overly complicated models. The authors suggest, instead, distinguishing between speculators and liquidity traders. They develop a theory of market liquidity and show that block trades can be traced back in markets for several days. Moreover, this chapter's authors have proven in practice that their insights are very valuable—measured in real money.

Chapter 3, "Stochastic Behavioral Asset-Pricing Models and the Stylized Facts" by Thomas Lux, acknowledges that traditional finance in the form of the efficient market hypothesis still plays a dominant role in explaining the first moment of asset returns by the martingale property, but that it fails to explain robust stylized facts concerning higher moments such as fat tails of the return distribution and stochastic and clustered volatility. The chapter outlines recent models of stochastic interaction of traders using simple behavioral rules that can explain these stylized facts as emergent properties of interactions and dispersed activities of a large ensemble of agents populating the market place. Understanding the properties of higher moments of asset returns is very profitable, since using derivatives allows the exploitation of predictability of any degree. Moreover, showing that market behavior emerges as a fundamentally different behavior than individual behavior avoids wasting money on simple analogies often used in standard finance, such as the "representative agent," according to which market behavior is of the same type as individual behavior.

Chapter 4, "Complex Evolutionary Systems in Behavioral Finance" by Cars Hommes and Florian Wagener, provides inspiring theoretical, empirical and experimental results. The theoretical results are outlined by a simple adaptive beliefs system based on trading strategies derived from mean–variance analysis with heterogeneous beliefs. In the most simple setting, beliefs are of two types: fundamentalists and trend followers. The population weights are driven by the success of the strategies. The model results range from perfect foresight equilibria to chaotic dynamics. This model has become the leading paradigm of heterogeneous agents models, and it has been generalized in many directions, one of which is the large type limit case—an approximation of a market with many different trader types—which is also outlined in this chapter. The model does well on yearly data going back to 1871. Finally, the main results of the model are validated by forecasting experiments.

Chapter 5, "Heterogeneity, Market Mechanism, and Asset Price Dynamics" by Carl Chiarella, Roberto Dieci, and Xue-Zhong He, shows the similarities and the differences in market dynamics between models populated by constant absolute and constant relative risk-averse agents. In both cases, agents have heterogeneous price expectations, some of which are formed by simple rules of thumb, and markets clear through a Walrasian auctioneer or through a market maker. The resulting dynamics are nonlinear and stochastic and differ according to the market-clearing mechanism assumed. As the authors show empirically, the framework nicely explains asset-price features such as fat tail behavior, volatility clustering, power-law behavior in returns, and bubbles and crashes.

Chapter 6, "Perfect Forecasting, Behavioral Heterogeneities, and Asset Prices" by Jan Wenzelburger, develops an intertemporal CAPM with heterogeneous expectations that lies between models in which agents have perfect foresight and models with exogenous and ad hoc expectation rules. The twist of this contribution is to properly define the expectations of rational agents, which consider that other agents may be irrational and that their own behavior has some market impact. Such perfect forecasts need to solve for the temporary equilibria of each period. These issues are discussed for the CAPM, with detailed derivation of the asset-price dynamics and conditions for market selection and survival of agents using various expectations.

Chapter 7, "Market Selection and Asset Pricing" by Lawrence Blume and David Easley, focuses on the old but nevertheless unsettled and very important debate about whether markets select for rational agents. The framework is traditional in its choice of a dynamic general equilibrium model populated by infinitely lived subjective expected utility maximizers. This chapter provides a very important link between finance and mainstream economics because the main topics of this handbook are exemplified in a model setup that every classically trained finance or economics student knows by heart. The main result discussed is that if equilibrium allocations are Pareto-efficient, markets select for rational agents (i.e., the market selects for those traders whose subjective beliefs are closest to the objective probabilities with which the states of the world occur). The chapter also provides insights into the deeper workings of market selection in this modeling framework by describing the discipline imposed by the market. A relevant and interesting issue is the analysis of the relationship between market selection and the noise trader literature.

Chapter 8, "Rational Diverse Beliefs and Market Volatility" by Mordecai Kurz, outlines a model in which all market participants do the very best they can—but not better. Every agent keeps track of all publicly available information and builds expectations that are consistent with this observation. This modeling approach leaves sufficient heterogeneity to explain asset-price features that, according to the rational expectations literature, are "anomalous" or "puzzling": excess volatility of asset returns, high- and time-varying risk premia, high volume of trade, and so on. Rational diverse beliefs turn out to provide a realistic and flexible paradigm between the two extremes—rational expectations and arbitrary ad hoc beliefs.

Chapter 9, "Evolutionary Finance" by Igor V. Evstigneev, Thorsten Hens, and Klaus Reiner Schenk-Hoppé, studies market selection among traders following behavioral rules that may not necessarily be generated by utility maximization. The model allows for complete and incomplete markets and for short-and long-lived assets. The surprising finding of the literature surveyed is that, even though the pool of behavioral rules is quite large and the model is fairly general, a simple fundamental trading strategy—investing proportional to the expected relative dividends—"does the trick" (i.e., achieves the highest expected growth rate) or is at least the unique evolutionary stable trading strategy. This trading strategy can be seen as the Kelly Rule—betting your beliefs—applied in a market that generates returns endogenously from the interaction of trading strategies. One possible application of this result is to explain the success of value investing. Moreover, aggressive betting strategies in markets with endogenous odds, such as stock markets, can be derived from these results.

This handbook's nine chapters can be characterized by several attributes, one of which is the time scale of the dynamics studied. Chapters 6 and 8 study expectation dynamics (i.e., the adjustment and learning process of boundedly rational investors). For most investors these dynamics happen on a medium time scale (months, quarters, even years). Both chapters develop new expectation hypotheses that are somewhere in between simple ad hoc heuristics and rational expectations. This interpretation can also be given for Chapter 4, in particular since the asset-pricing application is based on annual data.

Chapter 2 goes to a much smaller time scale: intraday dynamics where the market microstructure—and in particular the market-clearing mechanism—plays a crucial role. Chapter 5 is concerned with these issues as well. Chapter 3 is somewhere in between the high-frequency intraday scale and medium-term dynamics, as can be seen from the attempt to explain daily return data. Finally, Chapters 7 and 9 consider long-term dynamics because they study market selection determined by the evolution of wealth. Chapter 1 surveys models across the board.

Alternatively, the nine chapters can also be ordered according to the degree of rationality of the traders considered. Chapter 7 is closest to the traditional view of complete rationality since agents maximize subjective expected utility and have correct price expectations. Chapters 6 and 8 define notions of rationality that are still quite demanding but more realistic: in Chapter 6 the problem of what a completely rational agent should expect in a market with irrational agents is solved, while Chapter 8 defines a notion of rationality that uses all available information but not more than that. Further "down the road" to a smaller degree of rationality, we find the modeling approach that Chapters 1 to 5 outline. Agents maximize but they may not have completely rational price expectations. Finally, the approach of Chapter 9 dismisses all assumptions on rationality by moving to a purely behavioral model of investment.

It is our hope that this handbook, which encompasses several directions of current developments in dynamic and evolutionary models of financial markets, will serve interested readers by providing insight and inspiration.

Thorsten Hens
Swiss Banking Institute, University of Zurich

Klaus Reiner Schenk-Hoppé
Business School and School of Mathematics,
University of Leeds

Introduction to the Series

Advisory Editors

The *Handbooks in Finance* are intended to be a definitive source for comprehensive and accessible information. Each volume in the series presents an accurate, self-contained survey of a subfield of finance, suitable for use by finance and economics professors and lecturers, professional researchers, and graduate students and as a teaching supplement. The goal is to have a broad group of outstanding volumes in various areas of finance.

William T. Ziemba
University of British Columbia

CHAPTER 1

Thought and Behavior Contagion in Capital Markets

David Hirshleifer and Siew Hong Teoh
Merage School of Business
University of California–Irvine

1.1.	Introduction	2
1.2.	Sources of Behavioral Convergence	5
1.3.	Rational Learning and Information Cascades: Basic Implications	7
1.4.	What Is Communicated or Observed?	9
	1.4.1. Observation of Past Actions Only	10
	1.4.2. Observation of Consequences of Past Actions	15
	1.4.3. Conversation, Media, and Advertising	16
1.5.	Psychological Bias	17
1.6.	Reputation, Contracts, and Herding	18
1.7.	Security Analysis	20
	1.7.1. Investigative Herding	20
	1.7.2. Herd Behavior by Stock Analysts and Other Forecasters	21
1.8.	Herd Behavior and Cascades in Security Trading	24
	1.8.1. Evidence on Herding in Securities Trades	24
	1.8.2. Financial Market Runs and Contagion	27
	1.8.3. Exploiting Herding and Cascades	28
1.9.	Markets, Equilibrium Prices, and Bubbles	29
1.10.	Cascades and Herding in Firm Behavior	36
	1.10.1. Investment and Financing Decisions	36
	1.10.2. Disclosure and Reporting Decisions	38
1.11.	Contagion of Financial Memes	39
1.12.	Conclusion	44
	References	46

Note: We thank Jason Chan, SuJung Choi, and Major Coleman for their valuable research assistance.

HANDBOOK OF FINANCIAL MARKETS: DYNAMICS AND EVOLUTION

Abstract

Prevailing models of capital markets capture a limited form of social influence and information transmission, in which the beliefs and behavior of an investor affect others only through market price, information transmission and processing is simple (without thoughts and feelings), and there is no localization in the influence of an investor on others. In reality, individuals often process verbal arguments obtained in conversation or from media presentations and observe the behavior of others. We review here evidence about how these activities cause beliefs and behaviors to spread and affect financial decisions and market prices; we also review theoretical models of social influence and its effects on capital markets. To reflect how information and investor sentiment are transmitted, thought and behavior contagion should be incorporated into the theory of capital markets.

Keywords: capital markets, thought contagion, behavioral contagion, herd behavior, information cascades, social learning, investor psychology, accounting regulation, disclosure policy, behavioral finance, market efficiency, popular models, memes

1.1. INTRODUCTION

The theory of capital market trading and pricing generally incorporates only a limited form of social interaction and information transmission, wherein the beliefs and behavior of an investor affect other investors only through market price. Furthermore, in standard capital market models, there is no localized contagion in beliefs and trading. Trading behaviors do not move from one investor to other investors who are proximate (geographically, socially, professionally, or attentionally through connectivity in the news media). Even most recent models of herding and information cascades in securities markets involve contagion mediated by market price so that there are no networks of social interaction. Furthermore, existing behavioral models of capital market equilibrium do not examine how investors form naïve popular ideas about how capital markets work and what investors should do, and how such popular viewpoints spread.[1]

The theory of investment has incorporated social interactions somewhat more extensively, both in the analysis of increasing returns and path dependence (see Arthur, 1989) and in models of social learning about the quality of investment projects (discussed in Section 1.10.1). However, traditional models of corporate investment decisions do not examine the process of contagion among managers of ideas about investment, financing, disclosure, and corporate strategy.

In reality, individuals often observe others' behavior and obtain information and ideas through conversation and through print and electronic media. Individuals process this information through both reasoning and emotional reactions rather than performing

[1] Section 1.9 discusses work on social networks and securities trading (DeMarzo, Vayanos, and Zwiebel, 2001, and Ozsoylev, 2005) and learning in standard capital market models. The work of Robert Shiller and coauthors on "popular models" in finance is discussed in Section 1.11.

the simple Bayesian or quasi-Bayesian updating of standard rational or behavioral models. Popular opinions about investment strategies and corporate policies evolve over time, partly in response to improvements in scientific understanding and partly as a result of psychological biases and other social processes. We are influenced by others in almost every activity, and price is just one channel of influence. Such influence can occur through rational learning (see, e.g., Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992) or through through irrational mechanisms (see Section 1.5), the latter including the urge to conform (or deviate) and contagious emotional responses to stressful events.

This essay reviews theory and evidence about the ways beliefs about and behaviors in capital markets spread. We consider here decisions by investors about whether to participate in the stock market and what stocks to buy; decisions by managers about investment, financing, reporting, and disclosure; and decisions by analysts and media commentators about what stocks to follow, what stocks to recommend, and what forecasts to make. We also consider the effects of contagion on market prices, regulation, and welfare as well as policy implications.

We argue that in actual capital markets, in addition to learning from price, a more personal form of learning is also important: from quantities (individual actions), from performance outcomes, and from conversation—which conveys private information, ideas about specific assets, and ideas about how capital markets work. Furthermore, we argue that learning is often local: People learn more from others who are proximate, either geographically or through professional or other social networks. We therefore argue that social influence is central to economics and finance and that contagion should be incorporated into the theory of capital markets.

Several phenomena are often adduced as evidence of irrational conformism in capital markets, such as anecdotes of market price movements without obvious justifying news; valuations which, with the benefit of hindsight, seem like mistakes (such as the valuations of U.S. Internet stocks in the late 1990s or of mortgage-backed securities in recent years); the fact that financial activity such as new issues, IPOs, venture capital financing, and takeovers move in general or sector-specific waves (see, e.g., Ritter and Welch, 2002; Rau and Stouraitis, 2008). Observers are often very quick to denounce alleged market blunders and conclude that investors or managers have succumbed to contagious folly.

There are two problems with such casual interpretations. First, sudden shifts do not prove that there was a blunder. Large price or quantity movements may be responses to news about important market forces. Second, even rational social processes can lead to dysfunctional social outcomes.

With respect to the first point, market efficiency is entirely compatible with massive *ex post* errors in analyst forecasts and market prices and with waves in corporate transaction actions in response to common shifts in fundamental conditions.

With respect to the second point, the theory of information cascades (defined in Section 1.2) and rational observational learning shows that some phenomena that seem irrational can actually arise naturally in fully rational settings. Such phenomena include (1) frequent convergence by individuals or firms on mistaken actions based on little investigation and little justifying information; (2) fragility of social outcomes with respect to seemingly small shocks; and (3) the tendency for individuals or firms to delay decision for extended periods of time and then, without substantial external trigger, suddenly to act simultaneously. Furthermore, theoretical work has shown that reputation-building incentives on the part of managers can cause convergent behavior (Item 1) and has also offered explanations for why some managers may deviate from the herd as well. So care is needed in attributing either corporate event clustering or large asset price fluctuations to contagion of irrational errors.[2]

In addition to addressing these issues, we consider a shift in analytical point of view from the individual to the financial idea or meme. A *meme*, first defined by Dawkins (1976), is a mental representation (such as an idea, proposition, or catchphrase) that can be passed from person to person. Memes are therefore units of cultural replication, analogous to the gene as a unit of biological heredity. The field of memetics views cultural units as *replicators*, which are selected upon and change in frequency within the population. Just as changes in gene frequency imply evolution within biologically reproducing populations, changes in meme frequency imply cultural evolution. We argue that certain investment theories have properties that make them better at replicating (more contagious or more persistent), leading to their spread and survival.

Furthermore, we argue that through cumulative evolution, financial memes combine into coadapted assemblies that are more effective at replicating their constituent memes than when the components operate separately. We call these assemblies *financial ideologies*. Memetics offers an intriguing analytical approach to understanding the evolution of capital market (and other) popular beliefs and ideologies.

Only a few finance scholars have emphasized the importance of popular ideas about markets (especially Robert Shiller, as mentioned in Footnote 1), and there has been very little formal analysis of the effects and spread of popular financial ideas. We argue here that the analysis of thought contagion and the evolution of financial ideologies, as well as their effects on markets, constitute a missing chapter in modern finance, including behavioral finance.

Our focus is on contagion of beliefs or behavior rather than defining contagion as occurring whenever one party's payoff outcomes affect another's. Therefore we do not review systematically the literature on contagion in bankruptcies or international crises in which fundamental shocks and financial constraints cause news about one firm or region to affect the payoffs of another.

Section 1.2 discusses learning and the general sources of behavioral convergence. Section 1.3 discusses basic implications of rational learning and information cascades. Section 1.4 discusses basic principles of rational learning models and alternative scenarios of information transfer by communication or observation. Section 1.5 examines psychological bias and herding. Section 1.6 describes agency and reputation-based herding models. Section 1.7 describes theory and evidence on herding and cascades in security analysis. Section 1.8 describes herd behavior and cascades in security trading.

[2]Recent reviews of theory and evidence of both rational observational learning and other sources of behavioral convergence in finance include Devenow and Welch (1996), Hirshleifer (2001), Bikhchandani and Sharma (2001), and Daniel, Hirshleifer, and Teoh (2002).

Section 1.9 describes the price implications of herding and cascading. Section 1.10 discusses herd behavior and cascading in firms' investment, financing, and disclosure decisions. Section 1.11 examines the popular models or memes about financial markets. Section 1.12 concludes.

1.2. SOURCES OF BEHAVIORAL CONVERGENCE

An individual's thoughts, feelings, and actions are influenced by other individuals by several means: verbal communication, observation of actions (e.g., quantities such as supplies and demands), and observation of the consequences of actions (such as payoff outcomes or market prices). Our interest is in convergence or divergence brought about by direct or indirect social interactions (herding or dispersing). So we do not count random groupings that arise solely by chance as herding, nor do we count mere clustering, wherein individuals act in a similar way owing to the parallel independent influence of a common external factor.

Following Hirshleifer and Teoh (2003a) we define *herding/dispersing* as any behavior similarity or dissimilarity brought about by the direct or indirect interaction of individuals.[3] Possible sources include the following:

1. *Payoff externalities* (often called *strategic complementarities* or *network externalities*). For example, there is little point to participating in Facebook unless many other individuals do so as well.
2. *Sanctions upon deviants*. For example, critics of a dictatorial regime are often punished.
3. *Preference interactions*. For example, a teenager may want an iPhone mainly because others talk about the product, though a few mavericks may dislike a product for the same reason.
4. *Direct communication*. This is simply telling; however, "just telling" often lacks credibility.
5. *Observational influence*. This is an informational effect wherein an individual observes and draws inferences from the actions of others or the consequences of those actions.

We can distinguish an *informational hierarchy* and a *payoff hierarchy* in sources of convergence or divergence (see also Hirshleifer and Teoh, 2003a). The most inclusive category, *herding/dispersing*, includes both informational and payoff interaction sources of herding as special cases.

Within herding/dispersing, the informational hierarchy is topped by *observational influence*, a dependence of behavior on the observed behavior of others or the results of their behavior. This influence may be either rational or irrational. A subcategory is

[3] The interaction required in our definition of herding can be indirect. It includes a situation in which the action of an individual affects the world in a way that makes it more advantageous for another individual to take the same action, even if the two individuals have never directly communicated. But mere clustering is ruled out.

rational observational learning, which results from rational Bayesian inference from information reflected in the behavior of others or the results of their behavior. A further refinement of this subcategory consists of *information cascades*, wherein the observation of others (their actions, payoffs, or statements) is so informative that an individual's action does not depend on his own private signal.[4]

Imitation, broadly construed, includes both information cascades and subrational mechanisms that produce conformity with the behavior of others. A crucial benefit of imitation is the exploitation of information possessed by others. When an insider is buying, it may be profitable to buy even without knowing the detailed reason for the purchase. There is also contagion in the emotions of interacting individuals (see, e.g., Barsade, 2002). The benefits of imitation are so fundamental that the propensity to follow the behaviors of others has evolved by natural selection. Imitation has been extensively documented in many animal species, both in the wild and experimentally.[5]

In an information cascade, since an individual's action choice does not depend on his signal, his action is uninformative to later observers. Thus, cascades are associated with *information blockages* (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992), and, as we will see, with *fragility* of decisions (Bikhchandani, Hirshleifer, and Welch, 1992). Information blockages are caused by an informational externality: An individual chooses her actions for private purposes, with little regard for the potential information benefit to others.[6]

A *payoff interaction hierarchy* provides a distinct hierarchy of types of herding or dispersing that intersects with the categories in the information hierarchy. The first subcategory of the catch-all herding/dispersing category is *payoff and network externalities*. This consists of behavioral convergence or divergence arising from the effects of an individual's actions on the payoffs to others of taking that action. Direct payoff externalities have been proposed as an explanation for bank runs (Chari and Jagannathan, 1988; Diamond and Dybvig, 1983), since a depositor who expects other depositors to withdraw has a stronger incentive to withdraw, and clumping of stock trades by time (Admati and Pfleiderer, 1988) or exchange (Chowdhry and Nanda, 1991), since uninformed investors have an incentive to try to trade with each other instead of with the informed.

In several models, a desire for good reputation causes payoffs to depend on whether individual behaviors converge.[7] Thus, a subcategory of the *payoff and network externalities* category is *reputational herding and dispersion*, wherein behavior converges

[4]See Bikhchandani, Hirshleifer, and Welch (1992); Welch (1992). Banerjee (1992) uses a different terminology for this phenomenon.

[5]See, e.g., Gibson and Hoglund (1992), Giraldeau (1997), and Dugatkin (1992). Some authors use definitions of imitation that require substantial understanding on the part of the imitator, in which case such imitation is rare among nonhumans.

[6]Chamley (2004b) and Gale (1996) review models of social learning and herding in general. For presentation of information cascades theory and discussion of applications, tests, and extensions, see Bikhchandani, Hirshleifer, and Welch (1998, 2008a). Bikhchandani, Hirshleifer, and Welch (2008b) provides an annotated bibliography of research relating to cascades.

[7]See Scharfstein and Stein (1990), Rajan (1994), Trueman (1994), Brandenburger and Polak (1996), Zwiebel (1995), and Ottaviani and Sørensen (2006).

or diverges owing to the incentive for a manager to maintain a good reputation with some observer. When individuals care about their reputations, reputational herding and information cascades can both easily occur, since an individual who seeks to build a reputation as a good decision maker may rely on the information of earlier decision makers (Ottaviani and Sørensen, 2000).

1.3. RATIONAL LEARNING AND INFORMATION CASCADES: BASIC IMPLICATIONS

If many individuals possess conditionally independent signals about which choice alternative is better, their information could be aggregated to determine the right decision with arbitrarily high precision. Information cascades lead to information blockage, which reduces the quality of later decisions. This blockage also has several other implications for the contagion and stability of financial decisions, some of which hold even in rational learning settings in which cascades proper do not occur.

Consider a sequence of individuals who face *ex ante* identical choices (e.g., investment projects), observe conditionally independent and identically distributed private information signals, and observe the actions but not the payoffs of predecessors. Suppose that individual i is in a cascade and that later individuals understand this. Then individual $i + 1$, having learned nothing from the choice of i, is in an informationally identical position to that of i. So $i + 1$ also makes the same choice regardless of his private signal. By induction, this reasoning extends to all later individuals; the pool of information implicit in the past actions of individuals stops growing when a cascade begins. Indeed, in the simplest possible cascades setting, at this point the quality of decisions never improves again.

When the assumptions are modified slightly, information is not blocked forever. If individuals are not identical *ex ante*, then the arrival of an individual with deviant information or preferences can dislodge a cascade. For example, an individual with a highly precise signal will act independently, which conveys new information to later individuals. Furthermore, the arrival of public news, either spontaneously and independently of past choices or as payoff outcomes from past choices, can dislodge a cascade. The more generic implication of the cascades approach is that the quality of decisions improves much more slowly than would be the case under ideal information aggregation. Information blockages can last for substantial periods of time; as we will see, at such times social outcomes are often fragile.

Information cascades are a special case of *behavioral coarsening*, defined as any situation in which an individual takes the same action for multiple signal values. When there is behavioral coarsening, as in an information cascade, an individual's action does not fully convey his information signal to observers. So, where a cascade causes (at least temporarily) a complete information blockage, behavioral coarsening leads to partial blockage. A surprising aspect of the theory of information cascades is that in a natural setting the most extreme form of behavioral coarsening occurs.

Since information is aggregated poorly in an information cascade, the quality of decisions is reduced. Rational individuals who are in a cascade understand that the public pool of information implicit in predecessors' actions is not very precise. As a result, even a rather small nudge, such as a minor public information disclosure, can cause a well-established and thoroughly conventional behavior pattern to switch.

The arrival of a meaningful but inconclusive public information disclosure can, paradoxically, reduce the average quality of individuals' decisions. Other things can equal, a given individual is better off receiving the extra information in the disclosure. However, additional information will sometimes cause individuals to cascade earlier, aggregating the information of fewer individuals. On balance, the public signal can induce a less informative cascade (Bikhchandani, Hirshleifer, and Welch, 1992). Of course, if highly conclusive public information arrives, rational individuals will make very accurate decisions.

The dangers of a little learning are created in other information environments as well. In cascade models, the ability of individuals to observe payoff outcomes in addition to past actions, or to more precisely make a noisy observation of past actions can reduce the average accuracy of decisions (Cao and Hirshleifer, 1997, 2002). Also, the ability to learn by observing predecessors can make the decisions of followers noisier by reducing their incentives to collect (perhaps more accurate) information themselves (Cao and Hirshleifer, 1997). Furthermore, even if an unlimited number of payoff outcome signals arrive, the choices that individuals can make may limit the resulting improvement in the information pool. For example, there can be a positive probability that a mistaken cascade will last forever (Cao and Hirshleifer, 2002).

Often individuals choose not only whether to adopt or reject a project but *when* to do so. As a result, the timing and order of moves, which are given in the basic cascade model, are endogenously determined. In models of the option to delay investment choices,[8] there can be long periods with no investment, followed by sudden spasms in which the adoption of the project by one firm triggers investment by others.

Most of the conclusions described here generalize to other social learning settings in which cascades proper do not occur. Even when information blockage is not complete, information aggregation is limited by the fact that individuals privately optimize rather than taking into account their effects on the public information pool. This creates a general tendency for information aggregation to be self-limiting. At first, when the public pool of information is very uninformative, actions are highly sensitive to private signals, so actions add a lot of information to the public pool.[9] As the public pool of information grows, individuals' actions become less sensitive to private signals.

In the simplest versions of the cascade model, behavioral coarsening occurs in an all-or-nothing fashion so that there is either full use of private signals or no use of private signals (as in Banerjee, 1992, and the binary example of Bikhchandani, Hirshleifer, and Welch, 1992). In more general settings, coarsening occurs by degrees, but complete blockage eventually occurs (see the cascades model with multiple signal

[8]See Chamley and Gale (1994); see also Hendricks and Kovenock (1989); Bhattacharya, Chatterjee, and Samuelson (1986); Zhang (1997) and Grenadier (1999); and Chamley (2004a, 2004b).

[9]The addition can be directly through observation of past actions or indirectly through observation of consequences of past actions, as in public payoff information that results from new experimentation with various choice alternatives.

values of Bikhchandani, Hirshleifer, and Welch, 1992). In some settings, coarsening can gradually proceed without ever reaching a point of complete blockage, though the probability that an individual uses his own signal asymptotes toward zero, a phenomenon called “limit cascades” (Smith and Sørensen, 2000). Or, if there is observation noise, the public pool of information can grow steadily but more and more slowly (Vives, 1993).

So whether information channels become gradually or quickly clogged and whether the blockage is partial or complete depends on the economic setting, but the general conclusion that there can be long periods in which individuals herd upon poor decisions is robust. In addition, there tends to be too much copying or behavioral convergence; someone who uses her own private information heavily provides a positive externality to followers, who can draw inferences from her action.

Information cascades result from the individual’s private signal being overwhelmed by the growing public pool of information. Such an outcome is impossible in a setting where there is always a chance that an individual will receive a signal that is conclusive or arbitrarily close to conclusive. However, if near-conclusive signals are rare, the public information pool can grow very slowly, in which case “information cascade” can be a good approximation. Indeed, as the quality of the public information pool improves, the likelihood that an individual will receive a signal powerful enough to oppose it declines.

To summarize, the information cascade model and some related rational learning theories provide a few key general implications. The first and central implication is *idiosyncrasy*, or poor information aggregation. Cascades tend to emerge rapidly, so the signals of a relatively small number of early individuals dominate the behavior of numerous followers.

The second is *fragility*, or fads. The blockage of information aggregation that is characteristic of cascades makes behavior sensitive to small shocks. We are accustomed to thinking of sensitivity to shocks as a rare circumstance, as when a flipped coin lands on its side. The tendency of cascades to form suggests that real life is somewhat like Hollywood thrillers in which the chase scene inevitably ends with the hero’s car teetering precariously at the edge of a precipice.

The third is *simultaneity*, or delay followed by sudden joint action. Such effects are sometimes referred to as *chain reactions*, *stampedes*, or *avalanches*. Endogenous order of moves, heterogeneous preferences, and precisions can exacerbate these problems.

The fourth is *paradoxicality*, or adverse effects on decision accuracy or welfare of informational improvements; the fifth is *path dependence*, or outcomes depending on the order of moves or signal arrival. This implication is shared with models of payoff interactions (e.g., Arthur, 1989).

1.4. WHAT IS COMMUNICATED OR OBSERVED?

We now describe in somewhat more detail alternative sets of assumptions in observational influence models and the implications of these differences.

1.4.1. Observation of Past Actions Only

Here we retain the assumption of the basic cascade model that only past actions are observable, but we consider several model variations.

Discrete, Bounded, or Gapped Actions vs. Continuous Unbounded Actions

If the action space is continuous, unbounded, and without gaps, an individual's action is always at least slightly sensitive to his private signal. Thus, actions always remain informative, and information cascades never form. So, for inefficient information cascades to occur, actions must be discrete, bounded, or gapped. As discrete or bounded action spaces become more extensive, cascades become more informative, approaching full revelation.[10]

The assumption that actions are discrete is often highly plausible. We vote for one candidate or another, not for a weighted average of the two. A worker is hired or not hired, fired or not fired. A takeover bidder either does or does not seek control of a target firm. Often alternative investment projects are mutually exclusive. Although the amount invested is often continuous, if there is a fixed cost the option of not investing at all is discretely different from positive investment.

More broadly, one way in which the action set can be bounded is if there is a minimum and maximum feasible project scale. If so, when the public information pool is sufficiently favorable, a cascade at the maximum scale will form, and when the public information pool is sufficiently adverse, individuals will cascade on the minimum scale. Since there is always an option to reject a new project, investment has a natural extreme action of zero. Thus, a lower bound of zero on a continuous investment choice creates cascades of noninvestment (Chari and Kehoe, 2004). Similarly, gaps can create cascades.[11]

If perceptual discretizing is very finely grained, the outcome will still be very close to full revelation. However, perception and analysis are coarse; consider, for example, the tendency of people to round off numbers in memory and conversation. There is evidence of clustering for retail deposit interest rates around integers and that this is caused by limited recall of investors (Kahn, Pennacchi, and Sopranzetti, 1999).

[10] See Lee (1993); see also Gul and Lundholm (1995) and Vives (1993) for continuous settings without cascades. Early cascade models were based on action discreteness (Bikhchandani, Hirshleifer, and Welch, 1992; Welch, 1992).

[11] Asymmetry between adoption and rejection of projects is often realistic and has been incorporated in several social learning models of investment to generate interesting effects. As for gaps, sometimes either a substantial new investment, no change, or disinvestment is feasible, but fixed costs make a small change clearly unprofitable. If so, then a cascade upon no change is feasible. Similarly, a cascade of securities nontrading can form when there is a fixed cost of taking a long or short position, or when there is a minimum trade size. Even if the true action space is continuous, ungapped and unbounded, to the extent that observers are unable to perceive or recall small fractional differences, the actions of their predecessors effectively become either noisy or discrete.

Observability of Predecessors' Payoffs or Signals

Even if individuals observe a subset of past signals, such as the past *k* signals, since in general uncertainty remains, inefficient cascades can form. With regard to settings with observation of past payoffs, inefficient cascades can form and with positively probability last forever because a cascade can lock into an inferior choice before sufficient trials have been performed on the other alternative to persuade later individuals that this alternative is superior (Cao and Hirshleifer, 2002). We discuss research on the effects of observability of past payoffs and signals in more depth in Subsection 1.4.2.

Costless vs. Costly Private Information Acquisition

Individuals often expend resources to obtain signals, but they also often observe private signals costlessly in the ordinary course of life. Most social learning models take the costless route. Costs of obtaining signals can lead to little accumulation of information in the social pool for reasons similar to cascades or herding models with costless information acquisition. Individuals have less incentive to investigate or observe private signals if the primary benefit of using such signals is the information that such use will confer on later individuals. (Burguet and Vives, 2000, examine the conditions under which complete learning occurs in a continuum model with investigation costs.) Indeed, if the basic information cascade setting is modified to require individuals to pay a cost to obtain their private signals, once a cascade is about to start, an individual has no reason to investigate. The outcome is identical to the basic cascade model: information blockage. But the individual is acting without regard to her signal in only a degenerate sense: she has not acquired any signal to regard.

This suggests an extended definition of cascades that can apply to situations in which private signals are costly to obtain. Following Hirshleifer and Teoh (2003a), we define an *investigative cascade* as a situation in which either:

1. An individual acts without regard to his private signal, or
2. An individual chooses not to acquire a costly signal, but he would have acted without regard to that signal had he been forced to acquire it at the same level of precision that he would have voluntarily acquired if he were unable to observe the actions or payoffs of others

Item 1 implies that all information cascades are also investigative cascades. Item 2 is simplest in the special case of a binary decision of whether or not to acquire an information signal of exogenously given precision. Item 2 then reduces to the statement: The individual chooses not to acquire the signal, but if he were forced to acquire it he would ignore its realization (because of the information he has already gleaned by observing others).[12]

[12] Item 2 further allows for the purchase of different possible levels of precision. The definition focuses on the precision that the individual would select under informational autarky. If, under this precision, the individual's action does not depend on the realization, he is in an investigative cascade.

Investigative cascades may occur in the decisions by individuals to invest in different countries. If investigation of each requires a fixed cost, then with a large number of countries investors may cascade on noninvestment (see the related analysis of Calvo and Mendoza, 2001).

Observation of All Past Actions vs. a Subset or Statistical Summary of Actions

Sometimes people can observe only the recent actions of others, a random sample of actions, or the behavior of neighbors in some geographic or other network.[13] In such settings mistaken cascades can still form. For example, if only the preceding k actions are observed, a cascade may form within the first k individuals and then through chaining extend indefinitely. Alternatively, individuals may only be able to observe a statistical summary of past actions. Information blockage and cascades are possible in such a setting as well (Bikhchandani, Hirshleifer, and Welch, 1992). A possible application is to the purchase of consumer products. Aggregate sales figures for a product matter to future buyers because they reveal how previous buyers viewed desirability of alternative products (Bikhchandani, Hirshleifer, and Welch, 1992; Caminal and Vives, 1999).

Observation of Past Actions, Accurately or with Noise

When past actions are observed with noise, social learning is still imperfect (Vives, 1993), and (depending on the setting) cascades can still form (Cao and Hirshleifer, 1997). In some scenarios a model in which individuals learn from price reduces, in effect, to a basic social learning model with indirect observation of a noisy statistical summary of the past trades of others.

Choice of Timing of Moves vs. Exogenous Moves

Consider a setting in which individuals or firms with private signals about project quality have a choice about whether to invest or delay. In other words, firms decide when to exercise their investment options. Then in equilibrium there is delay (Chamley and Gale, 1994) because a firm that waits can learn from the actions of others. However, if all were to wait, there would be no advantage to delay. Thus, in equilibrium, firms with favorable signals randomize strategies in deciding how long to delay before being the first to invest. If only a few firms invest (by firms that have received favorable signals), other firms infer that the state of the world is bad, and investment activity ends. However, if many firms invest, this conveys favorable information and spurs a sudden rush to invest by the other firms (even firms with adverse signals). Indeed, in the limit a period of little investment is followed by either a sudden surge in investment or a collapse. Thus, the model illustrates simultaneity. In equilibrium, cascades occur and information is aggregated inefficiently.

[13] Bala and Goyal (1998) analyze learning from the actions and payoff experiences of neighbors. They show that this leads to convergence of behavior and, under some conditions, efficient outcomes.

Allowing for uncertainty about signal precision leads to a surprisingly simple outcome (Zhang, 1997). Suppose that investors know the precisions only of their own signals about project quality. In the unique symmetric equilibrium, those investors whose favorable signals are less precise delay longer than those with more precise favorable signals; noisy information encourages waiting for corroboration. In equilibrium there is delay until the critical investment date of the individual who drew the highest precision is reached. Once he or she invests, other investors all immediately follow, though investment may be inefficient. This sudden onset of investment illustrates simultaneity in an extreme form.[14]

In settings with social learning, information blockages, delays in investment, periods of sudden shifts in investment, and overshooting can occur, either with (Caplin and Leahy, 1994; Grenadier, 1999) or without (Caplin and Leahy, 1993; Persons and Warther, 1997) information cascades. These models share the broad intuition that informational externalities cause socially undesirable choices about whether and when to invest. For example, Caplin and Leahy (1994) analyze information cascades in the cancelation of investment projects when timing is endogenous. Individual cancellations can trigger sudden crashes in the investments of many firms.

Financial innovations such as leveraged buyouts often seem to follow a boom-and-bust pattern. Several authors have explained this pattern as resulting from managers adopting the innovation based on observation of the payoffs resulting from the repeated actions of other firms. In the model of Persons and Warther (1997), there is a tendency for innovations to "end in disappointment" even though all participants are fully rational. Participants expect to gain from extending the boom until disappointing news arrives. Related notions of informational overshooting have been applied to real estate and stock markets (Zeira, 1999).

Presence of an Evolving Publicly Observable State Variable

In models of cascades in the exercise of investment options, the trigger for exercising an option is often the exogenous continuous evolution of a publicly observable state variable that affects the profitability of investment. In the model of Grenadier (1999), eventually a small move in the state variable triggers a cascade of option exercise.

Stable vs. Stochastic Hidden Environmental Variables

The attractiveness of market conditions for financial transactions such as raising capital varies greatly over time. When the underlying state of the world is stochastic but unobservable, there can be fads wherein the probability that action changes is much higher than the probability of a change in the state of the world (Bikhchandani, Hirshleifer, and Welch, 1992). Moscarini et al. (1998) examine how long cascades can last as the

[14]Chamley (2004a) finds that when individuals have different prior beliefs, multiple equilibria generate different amounts of public information. Imperfect information aggregation can also occur in a rational expectation (simultaneous trading) modeling approach, when information is costly to acquire and investment is a discrete decision, causing price and investment fluctuations (Beaudry and Gonzalez, 2003).

environment shifts. Hirshleifer and Welch (2002) consider an individual or firm subject to memory loss about past signals but not actions. They describe the determinants (such as environmental volatility) of whether memory loss causes inertia (a higher probability of continuing past actions than if memory were perfect) or impulsiveness (a lower probability).

Homogeneous vs. Heterogeneous Payoffs

Individuals have different preferences, though this is probably more important in nonfinancial settings. Suppose that different individuals value adoption differently. A rather extreme case is opposing preferences or payoffs, such that under full information two individuals would prefer opposite behaviors. If each individual's type is observable, different types may cascade upon opposite actions.

However, if the type of each individual is only privately known and if preferences are negatively correlated, learning may be confounded; individuals do not know what to infer from the mix of preceding actions they observe, so they simply follow their own signals (Smith and Sørensen, 2000).

Endogenous Cost of Action: Models with Markets and Endogenous Prices

We cover this topic separately in Section 1.9.

Single or Repeated Actions and Private Information Arrival

Most models with private information involve a single irreversible action and a single arrival of private information. In Chari and Kehoe (2004), in each period one investor receives a private signal, and investors have a timing choice as to when to commit to an irreversible investment. In equilibrium there are inefficient cascades. If individuals take repeated, similar actions and continue to receive nonnegligible additional information, actions will, of course, become very accurate. However, there can still be short-run inefficiencies (e.g., Hirshleifer and Welch, 2002).

Discrete vs. Continuous Signal Values

Depending on probability distributions, with continuous signal values, limit cascades instead of cascades can occur (Smith and Sørensen, 2000). Of course, signal values are often discrete. For example, the buyer of a consumer product may observe as a signal of quality the number of "stars" or "thumbs-up" the product has received by a reviewing agency.[15] Furthermore, the empirical and policy significance of the two

[15]In practice, signal discreteness is rampant. Often the signal is information about whether something does or does not fall into some discrete category. For example, in voting for a U.S. Presidential candidate, an individual may take into account whether the individual currently is or is not Vice President. In deciding whether or not to bet on a horse, a gambler may use as a signal whether or not the horse won the last race; he may not know its exact time. When people obtain advice about a course of action, the advisor often recommends an alternative, with little elaboration.

predictions is much the same. Information arriving too late to be helpful for most individuals' decisions is similar to information being completely blocked for some period (Gale, 1996).

Exogenous Rules vs. Endogenous Contracts and Institutional Structure

Institutional rules and compensation contracts can be designed to manage herding and information cascades in project choice.[16]

1.4.2. Observation of Consequences of Past Actions

If vicarious learning can be used to aggregate the outcomes of many past trials of alternatives, one might expect that society could overcome information blockages to converge on correct actions. However, as emphasized by Shiller (2000b), imperfect rationality makes conversation a very imperfect aggregator of information. Biases induced by conversation are therefore likely to be important in terms of stock market behavior.

In formal modeling in an imperfectly rational setting (though not one designed specifically to capture Shiller's arguments), Banerjee and Fudenberg (2004) find convergence to efficient outcomes if people sample at least two predecessors. In their model, in each period a continuum of individuals tries choice alternatives. Since each individual observes only a sample from past history, the shadow of history is not overwhelming. Particular individuals fall into cascades, but different individuals make different choices. With a continuum of individuals, society cannot get unanimously stuck on a bad choice. Information about the payoffs from all possible options is continually regenerated, creating a rich inventory of information from which to draw.

A setting that is closer to the basic cascade model allows for observation of payoff outcomes without assuming the infinitely rich inventory of past information. In Cao and Hirshleifer (2002), there are two alternative project choices, each of which has an unknown value state. Payoffs are in general stochastic, each period conditional on the value state. Rational individuals receive private signals and act in sequence, and individuals can observe all past actions and project payoffs. Nevertheless, idiosyncratic cascades still form. For example, a sequence of early individuals may cascade on Project A, and its payoffs may become visible to all, perhaps revealing the value state perfectly. But since the payoffs of Alternative B are still hidden, B may be the superior project. Indeed, the ability to observe past payoffs can sometimes trigger cascades even more quickly, reducing average decision quality and welfare—that is, there is paradoxicality.

Intuitively, comparing the different settings when only a sample of past actions and outcomes is observed, decisions are improved because the shadow of the past becomes less overwhelming. When individuals are discrete, a sampling scenario makes it less likely that society will unanimously fix on a bad behavior, because there is more opportunity for a few individuals who observe an unusual historical sample to choose deviant

[16] See Prendergast (1993), Khanna (1997), and Khanna and Slezak (2000) (discussed later); see also Ottaviani and Sørensen (2001).

actions that generate new corrective information. Bad cascades become less frequent. In a sampling setting, having a greater number of individuals also reduces the likelihood of chance unanimous fixation on a bad action. At the extreme of an infinite number of individuals (as with a continuum), the risk of unanimous bad cascades can be eliminated.

Potential industry entrants can learn indirectly about the actions of previous entrants by observing market price, since this is affected by previous decisions. In the model of Caplin and Leahy (1993), entrants do not possess any private information prior to entry. Information problems slow the adjustment of investment to sectoral economic shocks.

1.4.3. Conversation, Media, and Advertising

A growing recent literature provides evidence suggesting that conversation in social networks conveys valuable information for financial decisions and spreads corporate and individual behaviors.[17] Biases in conversation contribute to the spread of mistaken beliefs. Contributing to this problem is a tendency for people to take at face value statements that they hear from acquaintances and the news media rather than rationally discounting for cheap talk.

News media activity can provide a measure of the extent to which information is being conveyed to investors. Veldkamp (2006) provides a model of "frenzies" in emerging equity markets in which media coverage rises and investors become better informed about asset payoffs and therefore face less risk so that asset prices rise. She provides supporting evidence.

Some individuals are more central than others in the social network that disseminates financial ideas and information. The news media creates nodes of high influence. Recent research has confirmed that the political opinions disseminated by media outlets affect those of viewers (DellaVigna and Kaplan, 2007). There is every reason to believe that media dissemination affects the financial ideologies of receivers as well.

Part of the effect of the media results from the sheer existence of high-influence nodes in the social network, especially since media commentators may have different beliefs from the public at large. Other effects arise from the self-interest of journalists and media firms, which can also influence the viewpoints expressed or the stories selected for reporting. This could bias stories because of a direct financial interest on the part of journalists in the firm they are reporting on, or bias could come from the benefits of

[17] Analysts who have old-school ties to corporate managers at a company make better stock recommendations about the company (Cohen, Frazzini, and Malloy, 2008a). Mutual fund managers who have old-school ties to corporate directors are more willing to take a large position in the firm and achieve better return performance on their holdings (Cohen, Frazzini, and Malloy, 2008b); and investors who have stronger social interaction based on several measures (old college ties, sharing the same profession, and geographical proximity) make more similar portfolio choices (Massa and Simonov, 2005). Gupta-Mukherjee (2007) finds that information relevant to achieving investment performance is transmitted among fund managers (along fund–fund networks) and between fund managers and companies in which they invest (along fund–company networks), where network linkage is identified by geographical proximity.

reporting a story that will grab the attention of the public (possibly at the expense of reporting more important stories).

Financial firms influence investors both by disclosures to the media and through advertising. Mullainathan and Shleifer (2005) argue that audiences like to see news that matches their beliefs and are more likely to be persuaded by advertising messages that fit their predispositions. Mullainathan and Shleifer provide evidence that, over the course of the Internet bubble, in good times (after high market returns) financial firms emphasize in their advertisements how their products create opportunity for investors, whereas in bad times advertisements emphasize safety.

1.5. PSYCHOLOGICAL BIAS

Conformism allows individuals to obtain the benefit of the valuable ideas of others. Several researchers have modeled the circumstances under which a propensity toward conformism is favored by natural selection and how conformism maintains cultural differences between groups (Henrich and Boyd, 1998; Boyd and Richerson, 2005). Kuran (1989) analyzes the effects of external pressures for and preferences for conformity; Bernheim (1994) analyzes the consequences of a preference for conformity.

Even without a direct preference for conformity, psychological bias can promote herding and cascades. Several models of herding or cascades assume either mechanistic or imperfectly rational decision makers, including Ellison and Fudenberg (1993, 1995; rules of thumb), Hirshleifer, Subrahmanyam, and Titman (1994; "hubris" about the ability to obtain information quickly), Bernardo and Welch (2001; overconfidence about the quality of information signals), Hirshleifer and Noah (1999; misfits of several sorts), and Hirshleifer and Welch (2002; memory loss about past signals).

A reasonable imitation strategy for individuals is to base choices on the payoffs that past adopters have received and on the market shares of the choice alternatives, as in the model of Smallwood and Conlisk (1979). An individual may observe a past sample of individuals and take an action based on the actions and payoffs within this sample (Ellison and Fudenberg, 1993, 1995).

If individuals use a diversity of decision rules (whether rational, quasi-rational, or simple rules of thumb), there will be greater diversity of action choices after rational individuals fall into a cascade. Action diversity can be informative and can break mistaken cascades (Bernardo and Welch, 2001; Hirshleifer and Noah, 1999). Consistent with Bernardo and Welch (2001), experiments show that individuals often overweight private signals, breaking cascades (Goeree, Palfrey, Rogers, and McKelvey, 2007).

Evidence of emotional contagion within groups suggests that there may be merit to popular views that there are contagious manias or fads in speculative markets (see also Shiller, 2000b; Lynch, 2000; and Lux, 1995). However, there are rational models of bubbles and crashes that do not involve herding (see, e.g., the review of Brunnermeier, 2001).

In security market settings, the assumption that the variance of aggregate noise trading is large enough to influence prices nonnegligibly (as in DeLong, Shleifer, Summers, and Waldmann, 1990), and subsequent models of exogenous noise implicitly reflects an assumption that individuals are irrationally correlated in their trades. This could be a result of herding (social interaction) or merely a common irrational influence of some noisy variable on individuals' trades. Park and Sgroi (2008) find evidence of irrational herding in an experimental security market.

We and others have argued that limits to investor attention are important for financial disclosure, financial reporting, and capital markets.[18] Such limits to attention may pressure individuals to herd or cascade despite the availability of a rich set of public and private information signals (beyond past actions of other individuals). A related issue is whether the tendency to herd or cascade is greater when the private information that individuals receive is hard to process (cognitive constraints and the use of heuristics for hard decision problems were emphasized by Simon, 1955; in the context of social influence, see Conlisk, 1996). In this regard, there is evidence that apparent herd behavior by analysts is greater for diversified firms, for which the task that analysts face is more difficult (Kim and Pantzalis, 2000).

1.6. REPUTATION, CONTRACTS, AND HERDING

The seminal paper on reputation and herd behavior, Scharfstein and Stein (1990), captures the insight of John Maynard Keynes that "it is better to fail conventionally than to succeed unconventionally." Consider two managers who face identical binary investment choices. Managers may have high or low ability, but neither they nor outside observers know which. Observers infer the ability of managers from whether their investment choices are identical or opposite and then update based on observing investment payoffs. Managers are paid according to observers' assessment of their abilities. It is assumed that high-ability managers will observe identical signals about the investment project, whereas low-ability managers observe independent noise.

There is a herding equilibrium in which the first manager makes the choice that his signal indicates, whereas the second manager always imitates this action regardless of her own signal. If the second manager were to follow her own signal, observers would correctly infer that her signal differed from that of the first manager, and as a result they would infer that both managers are probably of low quality. In contrast, if she takes the same choice as the first manager, even if the outcome is poor, observers conclude that there is a fairly good chance that both managers are high quality and that the bad outcome occurred by chance.

During bad times, the necessity for even a good firm to take actions indicative of poor performance can create an opening for a firm that has a choice to take such actions without severe reputational penalty. Rajan (1994) considers the incentive for banks with private information about borrowers to manage earnings

[18] See the review of Daniel, Hirshleifer, and Teoh (2002) and the models of Hirshleifer and Teoh (2003b, 2004), Peng and Xiong (2006), and Dellavigna and Pollet (2006, 2007).

upward by relaxing their credit standards for loans and by refraining from setting aside loan-loss reserves. In a bad aggregate state, even the loans of high-ability managers do poorly, so observers are more tolerant of a banker who sets aside loan-loss reserves. Thus, a set-aside of reserves triggers by a bank triggers set-asides by other banks. This simultaneity in the actions of banks is somewhat analogous to the delay and sudden onset of information cascades in the models Zhang (1997) and Chamley and Gale (1994).

Furthermore, Rajan shows that banks tighten credit in response to declines in the quality of the borrower pool. Thus banks amplify shocks to fundamentals. Rajan provides evidence from New England banks in the 1990s of such delay in increasing loan-loss reserves, followed by sudden simultaneous action.

It is often argued that stock market analysts have a reputational incentive to herd in their forecasts of future earnings. The classic model along these lines is Trueman (1994), which we cover in the next section. One of his findings is that analysts have an incentive to make forecasts biased toward the market's prior expectation. Brandenburger and Polak (1996) show that a firm or set of firms with superior information can have a reputational incentive to make investment decisions consistent with observers' prior belief about which project choice is more profitable—a sort of herding of managers on outsiders rather than each other. There can also be an incentive for subordinate managers to make recommendations consistent with the prior beliefs of their superiors (Prendergast, 1993).

In contrast with the model of Scharfstein and Stein, in which it is better to fail as part of the herd than to succeed as a deviant, in Zwiebel (1995) it is always best to succeed. Herding (and antiherding) is caused by the fact that a manager's success is measured relative to the success of others. The first premise of the model is that there are common components of uncertainty about managerial ability. As a result, observers exploit relative performance of managers to draw inferences about differences in ability. The second premise is that managers are averse to the risk of being exposed as having low ability (perhaps because the risk of firing is nonlinear). For a manager who follows the standard behavior, the industry benchmark can quite accurately filter out the common uncertainty. This makes following the industry benchmark more attractive for a fairly good manager than a poor one, even if the innovative project stochastically dominates the standard project. The alternative of choosing a deviant or innovative project is highly risky in the sense that it creates a possibility that the manager will do very poorly relative to the benchmark.[19]

However, in Zwiebel's model a very good manager can be highly confident of beating the industry benchmark even if he chooses a risky, innovative project. If this project is superior, it pays for him to deviate. Thus, intermediate-quality managers herd, whereas very good or very poor managers deviate. Zwiebel's approach suggests that under some circumstances portfolio managers may herd by reducing the risk of their portfolios relative to a stock market or other index benchmark, but under other circumstances they may intentionally deviate from the benchmark.

[19]Relative wealth concerns can also induce investment herding (DeMarzo, Kaniel, and Kremer, 2007).

Institutions and compensation schemes can be designed to address or exploit managers' incentives to cascade or to make choices to match an observer's priors (Prendergast, 1993 [discussed above]; Khanna, 1997; Khanna and Slezak, 2000). Khanna (1997) examines the optimal compensation scheme when managers have incentives to cascade in their investment decisions. In his model, a manager who investigates potentially has an incentive to cascade on the action of an earlier manager. Furthermore, a manager may delay investigation about the profitability of investment in the expectation of gleaning information more cheaply by observing the behavior of the competitor. Khanna describes optimal contracts that address the incentives to investigate and to cascade and the implications for compensation and investments across various industries.

Within the firm, the incentive to cascade on the recommendations of other managers makes it hard to motivate managers to make meaningful recommendations. In the model of Khanna and Slezak (2000), cascading among managers reduces the quality of project recommendations and choices. This is a drawback of a regime of "team decisions," in which managers make decisions sequentially and observe each others' recommendations. Incentive contracts that eliminate cascades may be too costly to be desirable for the shareholders. A hub-and-spoke hierarchical structure in which managers independently report recommendations to a superior eliminates cascades but requires superiors to incur costs of monitoring subordinates to prevent communication. Thus, under different conditions the optimal organizational form can be either teams or hierarchy.

1.7. SECURITY ANALYSIS

Most of the literature on information cascades in securities markets has focused on direct cascades in trading (Section 1.8) and elucidates the conditions under which such cascades can or cannot form. However, even in those scenarios in which direct cascades in trading cannot form, cascades of investigation can form before any trading has occurred. Such cascades still affect trading behavior.

1.7.1. Investigative Herding

Consider a sequence of individuals deciding whether to access a costly source of private information, such as an investment newsletter. If individuals can observe the decisions of predecessors (directly, through conversation, or through circulation data), an information cascade of acquisition of this information can form. The cascade may eventually be broken owing to a negative externality: When more investors have access to an information signal, its value goes down.

Such cascades are one case of what we call *investigative herding*. Positive payoff externalities can also create investigative herding (though information still plays an important role). The analysis of Brennan (1990) was seminal in illustrating a source of positive payoff externalities in the analysis of securities and the way this can create investigative herding. In his overlapping generations model, private information about a security is only reflected in market price the next period if a prespecified number of individuals have acquired the signal. Thus, the benefit to an investor of acquiring information about an asset can be low if no other investor acquires the information. However, if a group of investors tacitly coordinates on acquiring information, the investors who obtain information first do well.

This insight raises the question of whether investigative herding can occur in settings with greater resemblance to standard models of security trading and price determination. In the model of Froot, Scharfstein, and Stein (1992), investors with exogenous short horizons find it profitable to herd by investigating the same stock. In so doing they are indirectly able to effect what amounts to a tacit manipulation strategy. When they buy together the price is driven up, and then they sell together at the high price. Thus, herding even on “noise” (a spurious uninformative signal) is profitable.

However, even in the absence of opportunities for herding there is a potential incentive for individuals, acting on their own, to effect such manipulation strategies. If individuals are allowed to trade to “arbitrage” such manipulation opportunities, it is not clear that such opportunities can persist in equilibrium. This raises the question of whether there are incentives for herding per se rather than for herding as tacit manipulation.

Hirshleifer, Subrahmanyam, and Titman (1994) examine the security analysis and trading decisions of risk-averse individuals, where investigation of a security leads some individuals to receive information before others. They find a tendency toward herding. The presence of investigators who receive information late confers an obvious benefit on those who receive information early; the late informed drive the price in a direction favorable to the early informed. But by the same token, the early informed push the price in a direction unfavorable to the late informed.

The key to the herding result is that the presence of the late informed allows the early informed to unwind their positions sooner. This allows the early informed to reduce the extraneous risk they would have to bear if, to profit on their information, they had to hold their positions longer. This risk reduction that the late informed confer on the early informed is a genuine *ex ante* net benefit—it is not purely at the expense of the late informed. Overconfidence about the ability to become informed early further encourages herding in this model; each investor expects to come out the winner in the competition to study the “hot” stocks.

1.7.2. Herd Behavior by Stock Analysts and Other Forecasters

We expect rational forecasts to glean information from the forecasts of others, causing herding. A further question is whether, owing to reputational effects or irrationality, herding causes biased forecasts or recommendations.

Several studies provide evidence of herding in various kinds of forecasts, such as forecasts of the Japanese macroeconomy (Ashiya and Doi, 2001). If herding occurs for reputational reasons, other things equal, we would expect forecasters to tilt their forecasts toward the most accurate among other forecasters. Ehrbeck and Waldmann (1996) show that the pattern of repeated forecasts over time made by accurate (low mean-squared-error) forecasters tend to differ (e.g., smaller forecast revisions) from that of less accurate forecasters. Inconsistent with a rational reputational approach, in their tests economic forecasters bias their forecasts in directions characteristic of *less* accurate forecasters. Nevertheless, most analytical literature on stock market analysts has focused on rational reputational reasons for bias.

Analyst earnings forecasts are biased (see Givoly and Lakonishok, 1984; Brown, Foster, and Noreen, 1985). Forecasts are generally optimistic in the United States and other countries, especially at horizons longer than one year (see, e.g., Capstaff, Paudyal, and Rees, 1998). More recent evidence indicates that analysts' forecasts have become pessimistic at horizons of three months or less before the earnings announcement (see, e.g., Richardson, Teoh, and Wysocki, 2004).

A concern for reputation can pressure analysts to herd. The compensation received by analysts is related to their ranking in a poll by *Institutional Investor* about the best analysts (Stickel, 1992). Furthermore, analysts whose forecasts are less accurate than peers' are more likely to experience job turnover (Mikhail, Walther, and Willis, 1999).[20] These findings suggest that analysts may have an incentive to adjust their forecasts to maintain good reputations for high accuracy. Less experienced analysts are more likely than experienced ones to be terminated for "bold" (deviant) forecasts that deviate from the consensus forecast. This seems to place a higher pressure to herd on those analysts for whom uncertainty about ability is greater (Hong, Kubik, and Solomon, 2000). Clement and Tse (2005) finds that analysts with greater prior accuracy and experience are more likely to make "bold" (deviant) forecasts (see Trueman, 1994, below) and that herding forecast revisions tend to be less accurate than bold forecast revisions.

True herding (issuing forecasts that are biased toward those announced by previous analysts—a form of social interaction between analysts) should be distinguished from issuing forecasts that are biased toward prior earnings expectations. In the reputation model of Trueman (1994), both occur. In his analysis, an analyst has a greater tendency to herd if he is less skillful at predicting earnings—it is less costly to sacrifice a poor signal than a good one.

When analysts herd, a shift in forecast by one analyst stimulates response by others. A challenge for testing this prediction and for evaluating whether the response is appropriate is that it is difficult to show causality. Changes in consensus analyst forecasts are indeed positively related to subsequent revisions in analysts' forecasts (Stickel, 1990), which is potentially consistent with herd behavior. This relationship is weaker for the high-precision analysts who are ranked in an elite category by *Institutional Investor* magazine than for analysts who are not. Thus, it appears that analysts ranked as elite are less prone to herding than those who are not, consistent with the prediction of the Trueman model.

Experimental evidence involving experienced professional stock analysts has also supported the model (Cote and Sanders, 1997). Cote and Sanders report that these

[20]The importance of relative evaluation supports the premise of reputational models of herding. However, Mikhail et al. find no relation between either absolute or relative profitability of an analyst's *recommendations* and probability of turnover.

forecasters exhibited herding behavior. Furthermore, the amount of herding was related to the forecasters' perception of their own abilities and their motivation to preserve or create their reputations.

Two papers that provide methodologies for estimating herding or exaggeration of differences (dispersing, the opposite of herding) by analysts in the field (Zitzewitz, 2001; Bernhardt, Campello, and Kutsoati, 2006) find dispersing, that is, analysts *exaggerate* their differences. Zitzewitz also finds that analysts under-update their forecasts in response to public information, indicating an overweighting of prior private information. This evidence is supportive of overconfidence by analysts in their own private signals or with reputational models in which some individuals intentionally diverge (e.g., Prendergast and Stole, 1996; Ottaviani and Sørensen, 2006).

It is also often alleged that analysts herd in their choice of stocks to follow. There is high variation in analyst coverage of various firms (Bhushan, 1989); this does not prove that herding is the source of this variation. Rao, Greve, and Davis (2001) provide evidence that analysts tend to follow each other in initiating coverage on a stock, and they provide a cascades interpretation.

There are also allegations that analysts herd inappropriately in their stock recommendations. The evidence of Welch (2000) indicates that revisions in the buy and sell stock recommendations of a security analyst are positively related to revisions in the buy and sell recommendations of the next two analysts. Welch traces this influence to short-term information, identified by estimating the ability of the revision to predict subsequent returns.[21]

Welch also finds that analysts' choices are correlated with the prevailing consensus forecast. The "influence" of the consensus on later analysts is not stronger when it is a better predictor of subsequent stock returns. In other words, the evidence is consistent with analysts herding even on consensus forecasts that aggregate information poorly. This is consistent with either agency effects such as reputational herding or imperfect rationality on the part of analysts. Finally, Welch finds an asymmetry—that the tendency to herd is stronger when recent returns have been positive ("good times") and when the consensus is optimistic. He speculates that this could lead to greater fragility during stock market booms and the occurrence of crashes.

A different way to test for the effects of herding in recommendations is to examine the stock price reactions to new recommendations that are close to versus farther from the consensus forecast (Jegadeesh and Kim, 2007). Such tests indicate that stock price reactions are stronger when the new recommendation deviates farther from the consensus. Assuming that the market is efficient, this suggests that analysts herd (and that investors appropriately adjust for this fact in setting prices).

There is mixed evidence of herding in recommendations of investment newsletters (Jaffe and Mahoney, 1999; Graham, 1999). Graham (1999) develops and tests a reputation-based model of the recommendations of investment newsletters, in the spirit of Scharfstein and Stein (1990). He finds that analysts with better private information

[21] This could reflect cascading, or could be a clustering effect wherein the analysts commonly respond to a common information signal.

are less likely to herd on the market leader, the Value Line investment survey. This finding is consistent with either reputational herding or information cascades.

1.8. HERD BEHAVIOR AND CASCADES IN SECURITY TRADING

Some sociologists have emphasized that the "weak ties" of liaison individuals, who connect partly separated social networks, are important for spreading behaviors across networks (Granovetter, 1973). Recent literature in economics has examined the strength of peer-group effects in a number of different contexts (see, e.g., the review of Glaeser and Scheinkman, 2000).

Questionnaire/survey evidence indicates that word-of-mouth communication is important to the trading decisions of both individual and institutional investors (Shiller and Pound, 1989). Employees are influenced by the choices of coworkers in their decisions as to whether to participate in various employer-sponsored retirement plans (Duflo and Saez, 2002, 2003). Furthermore, there is both modern and historical evidence suggesting that social interactions between individuals affect decisions about equity participation and other financial decisions.[22] There is also evidence of clustering in the trading of institutional managers. Mutual fund managers located in a given city tend to be more correlated in their purchases or sales of stocks than managers located in different cities (Hong, Kubik, and Stein, 2005), possibly owing to access to similar information sources.

1.8.1. Evidence on Herding in Securities Trades

Herding on Endorsements

According to the information cascades theory, endorsements can be extremely influential if the endorser has a reputation for accuracy and if the endorsement involves an actual informative action by the expert. This could take the form of knowing that the expert took a similar action (buying a stock), but it could also involve the expert investing his reputation in the stock by recommending it.

Empirically, the choice by a big-five auditor, top-rank investment bank, or venture capitalist to invest its reputation in certifying a firm causes investors to update favorably.[23] There are many examples of influential investors. The publication of news that Warren Buffett has purchased a stock is associated with a positive stock price reaction that is on average greater than 4% (Martin and Puthenpurackal, 2008). Stock prices react to the news of the trades of insiders (e.g., Givoly and Palmaon, 1985). Such trades

[22]See Kelly and O'Grada (2000); Hong, Kubik, and Stein (2004); Ivkovich and Weisbenner (2007); Shive (2008); and Brown, Ivkovich, Smith, and Weisbenner (2008).
[23]See the model of Titman and Trueman (1986), and the evidence of Beatty and Ritter (1986), Beatty (1989), Simunic (1991), and Michaely and Shaw (1995).

provide information, and this price evidences indicates that observers use it to adjust their demand for stock.

Investing in human capital is a form of endorsement; the signing of a famous name to a management team affects the way a startup is perceived by investors. Investors are sometimes irrationally influenced by famous but incompetent analysts—stock market "gurus." This may involve a limited attention/availability effect wherein investors use an analyst's visibility as an indicator of ability. A would-be guru can exploit the flaws of this heuristic by using even outlandish publicity stunts to gain notoriety—see, for example, the description of Joseph Granville's career in Shiller (2000b). There is also evidence that investors are influenced by implicit endorsements, as with default settings for contributions in 401(k) plans—see Samuelson and Zeckhauser (1988) and Madrian and Shea (2001).

Herding on Trades

A key challenge for empirical tests for herding is to show that there is actual social interaction or strategic complementarity rather than clustering based solely on some external causal factor (Manski, 1993). If an external factor shifts the cost or benefit of some action (such as buying a stock) for a group of investors, their trades will shift together even if there is no social interaction.

Several lines of attack have been used to identify herding in financial markets. One approach is to carefully control variables that jointly affect the behavior of different individuals (see, e.g., Grinblatt, Keloharju, and Ikaheimo, 2008 on demand for cars; for a general analysis of econometric issues in measuring social interaction, see Brock and Durlauf, 2000). Using an instrumental variable approach, Brown, Ivkovich, Smith, and Weisbenner (2008) show that the effect of neighbors on stock market participation is stronger in communities with stronger social interactions. To identify the effects of local peers on an individual's stock ownership (as distinguished from the effects of other factors that may affect all local individuals), Brown et al. focus on the effects of stock ownership by the *nonlocal parents* of the local peers. Ng and Wu (2006) provide relatively direct evidence of social influence in trades through word-of-mouth communication in trading rooms in China; see also Ivkovich and Weisbenner (2007).

If higher population density encourages social interaction, density should affect volume of trading (see the tests of Eleswarapu, 2004). Survey evidence indicates that households that are more social or that attend church participate more in the stock market (Hong, Kubik, and Stein, 2004), suggesting that participation is contagious. Shive (2008) measures the opportunity for investors who own a stock to "infect" nonholders in a municipality by the product of the number of owners and the number of nonowners of the stock; models from epidemiology contain such product terms. She finds that this product term is a predictor of trading in the 20 most actively traded Finnish stocks, consistent with social interactions affecting trading.

A few studies examine natural or artificial experiments that rule out the possibility of an omitted influence. There is evidence of the peer effect of roommates on grade-point average and on decisions to join fraternities even when roommates are assigned randomly (Sacerdote, 2001), which avoids the possible bias. Also, a growing literature starting with Anderson and Holt (1996) has confirmed learning by observing actions and the existence of information cascades in the experimental laboratory.[24]

The causation issue is especially tricky in financial market trading tests because of the market clearing condition as mediated by price. Correlation in trades within a group of investors (conditioned on past price movements in some tests) may merely reflect herding (or other reasons for correlated trading) by some other investor group of investors. For example, individual investors buying and selling in tandem could result from some other group of investors such as mutual funds buying and selling in tandem, influencing prices. If individual investors supply liquidity to institutions by trading as contrarians in response to price movements (as found by Kaniel, Saar, and Titman, 2008), they will tend to trade together.

If there are only two groups of traders, by market clearing herding by one group of traders causes correlation in the trades of the other group, even if there are no interactions and no strategic complementarities between members of this other group. Thus, to verify that a group is truly herding, it is crucial to either control for price or find some other way to verify the causality of the behavioral convergence.

Several alternative measures of herding in trading behavior have been developed in papers on the behavior of institutional investors.[25] Bikhchandani and Sharma (2001) critically review alternative empirical measures of herding. Griffiths et al. (1998) find increased similarity of behavior in successive trades for securities that are traded in an open-outcry market rather than a system trading market on the Toronto stock exchange, consistent with the possibility of imitation trading raised by the evidence of Biais, Hillion, and Spatt (1995). Grinblatt and Keloharju (2000) provide evidence consistent with herding by individuals and institutions.

Institutional investors constitute a large fraction of all investors. By market clearing it is impossible for all investors to be buyers or all to be sellers. So, although testing for herding by such a large group is not unreasonable, it is helpful to examine finer subdivisions of investors.

Recent studies find evidence of correlated trading by various categories of institutional investors, especially those trading in small firms. Whether this reflects actual herding by (interaction among) institutions, common responses to common information signals, or correlated trading in response to herding by individual or other institutional investors is unclear. There is evidence that the trades of individual investors as a group are correlated (Kumar and Lee, 2006) and evidence from trading in China of stronger correlation in trades among individual investors who are geographically close (Feng and Seasholes, 2004).[26]

[24]See also Hung and Plott (2001), Anderson (2001), Sgroi (2003) and Celen and Kariv (2005, 2004). Consistent with cascades, female guppies tend to reverse their mate choices in experiments where they observe other females choosing different males (Dugatkin and Godin (1992)).

[25]See Lakonishok, Shleifer, and Vishny (1992), Grinblatt, Titman, and Wermers (1995), and Wermers (1999).

[26]Several papers provide evidence of correlation in the trades of institutional investors (referred to as 'herding' in this literature with no clear implication of interaction between traders), and provide many interesting stylized facts (Lakonishok, Shleifer, and Vishny, 1992; Grinblatt, Titman, and Wermers, 1995; Kodres and Pritsker, 1997; Wermers, 1999; Nofsinger and Sias, 1999; and Sias, 2004).

Fund managers who are doing well tend to lock in their gains toward the end of the year by indexing the market, whereas funds that are doing poorly deviate from the benchmark to try to overtake it (Brown, Harlow, and Starks, 1996, and Chevalier and Ellison, 1997). Chevalier and Ellison (1999) identify possible compensation incentives for younger managers to herd by investing in popular sectors and find empirically that younger managers choose portfolios that are more "conventional" and that have lower nonsystematic risk.

There is evidence suggesting that mutual fund herding affects prices (Brown, Wei, and Wermers, 2008). Mutual funds tend to buy stocks that have experienced consensus analyst upgrades and to sell stocks with consensus downgrades. Brown, Wei, and Wermers find that the upgraded stocks at first achieve superior return performance but subsequently underperform, suggestive of either reputational or imperfectly rational herding.

1.8.2. Financial Market Runs and Contagion

The problem with motivating provision of public goods is that the contributions of one party confer positive externalities on others. When there are strategic complementarities in contributions, there is the possibility of a "run," possibly triggered by information disclosure, in which contributions suddenly shrink to zero or some other low level. As a result, unbiased disclosure can reduce welfare (Teoh, 1997).

The most familiar form of a financial market run is the bank run. There is a negative payoff externality in which withdrawal by one depositor or the refusal of a creditor to renegotiate a loan reduces the expected payoffs of others. Bernardo and Welch (2004) model how financial market runs can arise endogenously among investors in a stock because of illiquidity.

A traditional view is that bank runs are due to "mob psychology" or "mass hysteria" (see the references discussed in Gorton, 1988). At some point economists may revisit the role of emotions in causing bank runs, or "panics," and more generally causing multiple creditors to refuse to finance distressed firms. Such an analysis will require attending to evidence from psychology about the way emotions affect judgments and behavior.

The main existing models of bank runs and financial distress are based on full rationality (for reviews of models and evidence about bank runs, see, e.g., Calomiris and Gorton, 1991; and Bhattacharya and Thakor, 1993, Section 5.2). The externality in withdrawals can lead to multiple equilibria involving runs on the bank or firm or to bank runs triggered by random shocks to withdrawals (see, e.g., Diamond and Dybvig, 1983). This of course does not preclude the possibility that there is also an informational externality.

The informational hypothesis (e.g., Gorton, 1985, 1988) holds that bank runs result from information that depositors receive about the condition of banks' assets. When a distressed firm seeks to renegotiate its debt, the refusal of one creditor may make others more skeptical. Similarly, if some bank depositors withdraw their funds from a troubled bank, others may infer that those who withdrew had adverse information about the value of the bank's illiquid assets, leading to a bank run (see, e.g., Chari and Jagannathan, 1988; Jacklin and Bhattacharya, 1988).

Since the decision to withdraw is bounded (an investor can only withdraw up to the amount of his or her order deposit), bank runs can be modeled as information cascades. There is a payoff interaction as well. However, at the start of the run, when only a few creditors have withdrawn, the main effect may be the information conveyed by the withdrawals rather than the reduction in the bank's liquidity. Furthermore, if asset values are imperfectly correlated, cascades can pass contagiously between banks and cause mistaken runs in banks that could have remained sound (on information and contagious bank runs, see Gorton, 1988; Chen, 1999; and Allen and Gale, 2000).

There is evidence of geographical contagion between bank failures or loan-loss reserve announcements and the returns on other banks (Aharony and Swary, 1996; Docking, Hirschey, and Jones, 1997). This suggests that bank runs could be triggered by information rather than being a purely noninformational (multiple equilibria, or effects of random withdrawal) phenomenon.[27] There is also evidence of contagion effects in a sample of United States bank failures during the period 1930–32 (Saunders and Wilson, 1996), but see also Calomiris and Mason (1997, 2001).

The problem of financial runs can potentially explain regional financial crises as well. Adverse information can cause lenders to be reluctant to extend credit, and owing to the externality in providing capital, this can potentially lead to a collapse. Chari and Kehoe (2003) model international financial crises as informational runs. The model of Teoh (1997) suggests that intransparency can in principle have the desirable effect of preventing crises.

1.8.3. Exploiting Herding and Cascades

Firms often market experience goods by offering low introductory prices. In cascade theory, the low price induces early adoptions, which helps start a positive cascade. Welch (1992) developed this idea to explain why initial public offerings of equity are on average severely underpriced by issuing firms. The pricing decision for the Microsoft IPO seems to have reflected this consideration (Uttal, 1986, p. 32), and later authors have provided supporting evidence (Amihud, Hauser, and Kirsh, 2003).

The advantages of inducing information cascades may apply to auctions of other goods. In the model of Neeman and Orosel (1999), there is a potential winner's curse, and a seller (such as a firm selling assets) can gain from approaching potential buyers sequentially and inducing information cascades rather than conducting an English auction.

[27] There is also evidence of contagion in speculative attacks on national currencies (Eichengreen, Rose, and Wyplosz, 1996).

1.9. MARKETS, EQUILIBRIUM PRICES, AND BUBBLES

In classical models of asset markets such as the Capital Asset Pricing Model (CAPM), investors are rational and markets are perfect and competitive. There is complete agreement about probability distributions of exogenous variables, which are common knowledge. As a result, risk-adjusted returns are unpredictable. Furthermore, in a classical market there is no excess volatility, if by this term we mean some faulty processing of information by the market that creates opportunities for abnormal trading profits. So, fully rational and frictionless models of cascades or herding cannot explain anomalous evidence regarding return predictability or excess volatility based on public information. (For recent surveys of theory and evidence on investor psychology in capital markets, see, e.g., Hirshleifer, 2001; Daniel, Hirshleifer, and Teoh, 2002.) To explain return patterns that are anomalous for classical models, market frictions or imperfect rationality are needed.[28]

However, information blockages and herding can still affect prices. For example, in the model of Abreu and Brunnermeier (2003), the common knowledge assumption is violated, and arbitrageurs who seek to profit from the end of a bubble are not sure when other arbitrageurs will start to sell. Arbitrageurs may use a public news event to synchronize, causing the bubble to burst.

Within a fully rational setting (and with common knowledge about probability distributions), cascades or herding can block information aggregation. Cascades or herding can affect how much information gets into that information set in two ways. First, there can be direct cascades in investor trading, causing some information to remain private that otherwise would be reflected in prices. Here we discuss some models in which this occurs because of market imperfections. Second, even if markets are perfect, cascades or herding can cause individuals to cascade in their investigation behavior, which affects the amount of private and public information that is generated in the first place. For example, if individuals cascade in subscribing or not subscribing to a stock market newsletter, the effect of this on the distribution of private information across investors affects trading and prices.

An intuition similar to the intuition of cascades or herding models with exogenous action cost can be extended to the issue of how quickly investors learn in competitive securities markets. In cascades and other learning models, the past history of informative actions creates an information pool that can crowd out the accumulation of new information because new decision makers have insufficient incentive to take informative actions. Even in a market setting without cascades, an informed trader does not internalize the benefit that other traders have from learning his private information as revealed through trading. As more private information is reflected in price, informed traders have diminishing incentive to trade speculatively against the market price (setting aside any change in the riskiness of speculative trading). If informed traders trade less, their trades tend to be lost amid the uninformative trades, reducing the aggregation of their information

[28]In rational expectations models of information and securities trading, returns are predictable owing to so-called noise or liquidity trading. But limited amounts of noncontingent liquidity trading will not explain major bubbles and crashes.

into market price. Thus, the rate of convergence of price to efficiency can slow over time (Vives, 1995).

In settings without technological externalities, market interactions typically lead to dispersing (antiherding) because the attempt to acquire a resource makes it more costly for others to do so.[29] Consider, for example, competition among demanders in the market for bread. An increase in an individual's demand drives up the price, causing others to reduce their demand quantities. This supports market clearing; prices rise to constrain purchases to the available supply.

Similarly, in the market for a security, the tendency to imitate past trades is limited by the fact that past trades tend to drive prices to be averse to further trades in the same direction. Nevertheless, under asymmetric information there are circumstances under which cascades or other forms of herding occur.

Under asymmetric information, social learning, and a perfectly liquid securities market, the basic argument for why cascades of trading will *not* form is still fairly direct. If at a given price an individual were going to buy (for example) regardless of his information signal, a rational seller who understands this would charge a higher price. For this reason, even in a securities market model where the action space is discrete and there there are transactions costs, such as that of Glosten and Milgrom (1985), the argument implies that there will be no information cascades.

To see this in more detail, consider the market clearing condition. If privately informed traders were buying regardless of their signals, *a fortiori* so would uninformed traders; having no signal encourages buying more than having an adverse signal. But if, foreseeably, both informed and uninformed will try to buy, then (the argument goes) the market maker should set prices higher—a contradiction. Indeed, some experimental tests in a simple Glosten/Milgrom setting do not yield cascades.[30]

There are three main snags with the argument against direct cascades in trading decisions. First is that special constraints may prevent prices from adjusting. Second is that the party on the opposite side of the transaction (call it the "market maker") may not know for sure that the investor is going to buy regardless of his signal. The third is that owing to transaction costs, an individual may *refrain* from trading regardless of her signal, which blocks information aggregation.

With respect to the first snag, there are special circumstances in which nonmarket clearing prices are imposed. In the short run, the expectation that NYSE specialists will maintain an "orderly market" by keeping prices continuous can force temporary deviations of prices from fundamental values, blocking information flow. This suggests that during extreme market periods such as crashes, cascades can form. Sometimes prices are explicitly constrained by price move limits. This permits cascades in which all investors try to buy (or all sell), resulting in nontrading.

[29]There is a growing literature on externalities as a source of herding, as in work on strategic complementarities (Haltiwanger and Waldman, 1989) and agglomeration economies (Krugman and Venables, 1995).

[30]See Cipriani and Guarino (2005) and Drehmann, Oechssler, and Roider (2005). On the other hand, other experimental evidence suggests that market settings do not solve the problem of inefficient information aggregation (Hey and Morone, 2004).

The second problem with the argument against direct trading cascades is that the party on the opposite side of the transaction (call it the “market maker”) may not know for sure that the investor is going to buy, regardless of his signal. We will first discuss how this issue can lead to what we will call quasi-cascades. We then discuss how cascades proper can occur.

It is standard to assume that informed investors know more than the market maker about a single dimension of uncertainty, the expected payoff of the security. Suppose that in addition there is a second type of informational advantage to informed investors over the market maker—knowledge about *whether* informative signals were sent. Then a price rise can encourage an investor with an adverse signal to buy when there is a transaction cost or bid–ask spread (Avery and Zemsky, 1998). The price rise persuades the investor that others possess favorable information, whereas the market maker adjusts prices sluggishly in response to this good news.

This relative sluggishness of the market maker arises as a result of his ignorance over whether an informative signal was sent. Informed traders—even those with adverse signals—at least know that information signals were sent, so that the previous order probably came from a favorably informed trader. In contrast, the market maker places greater weight on the possibility of a liquidity trade.

The behavior described by Avery and Zemsky is very cascade-like in that the individual trades in opposition to her private signal—a rather extreme behavioral coarsening. However, it is not strictly an information cascade, because when no information signal is received, the investor takes a different action than she does when information is received. There are really three possible signal realizations—favorable, unfavorable, and “no signal.” Action does depend on the value of this appropriately redefined signal. Still, the result is a quasi-cascading phenomenon with partial information blockage.[31]

The third snag with the argument against cascades in trading decisions is that transaction costs can easily cause cascades of nontrading. Bid–ask spreads or other transactions costs, by deterring trade, can block information flow. In the model of Cipriani and Guarino (2007), transaction costs cause cascades of nontrading that block information aggregation and cause prices to deviate from full-information fundamental value. Cipriani and Guarino (2007) also provide experimental support for the predictions of their model.

Cipriani and Guarino (2003) provide a modified version of Glosten/Milgrom with multiple securities. Starting with a single security intuition, the trading of informed

[31]Private information about the existence or quality of private signals (i.e., multiple dimensions of uncertainty) can also lead to information blockages in which learning about precision stops (Gervais, 1996). In the model of Gervais (1996), owing to uncertainty about investors' information precision, traders' private information is not fully incorporated into price. Informed investors know the precision of their private signal, but the market maker does not. Initially a high bid–ask spread acts as a filter by deterring trade by informed investors unless they have high precision. By observing whether trade occurs, the market maker updates about signal precision and asset value and narrows the spread over time. Eventually even investors with imprecise signals trade, and the market maker stops learning about precision. The independence of the decision to trade the private information about precision is a behavioral coarsening and causes some information about the insiders' precision to remain forever private.

investors causes information to be partly reflected in price. At some point, as price becomes more informative, having one more conditionally independent private signal causes an investor to update expected fundamental value only modestly. So, an investor who has a nonspeculative reason to purchase the security finds it profitable to buy even if his private information signal is adverse; similarly, an investor who has a nonspeculative motive to sell does so regardless of his signal. In other words, he is in a cascade. With all informed investors in a cascade, further aggregation of information is completely blocked. Thus, in contrast to Avery and Zemsky, information cascades in proper form. In addition, cascades can result from contagion across markets; trading in one asset can trigger a cascade in another market. Furthermore, cascades can occur in both markets at the same time, leading to complete information blockage.

In Lee (1998), quasi-cascades result in temporary information blockage, then information avalanches. This results from transaction costs and discreteness in trades, which can cause informed investors to be sidelined. In sequential trading, hidden information becomes accumulated as the market reaches a point at which, owing to transaction costs, trading temporarily ceases. Eventually a large amount of private information can be revealed by a small triggering event. The triggering event is a rare, low-probability adverse signal realization. An individual who draws this signal value sells. Other individuals who observe this sale are drawn into the market, causing a market crash or information avalanche.[32]

A key issue regarding the occurrence of information blockage in these models is the significance of the assumption of discrete actions. Any model that attempts to explain empirical phenomena such as market crashes as (quasi-)cascades must calibrate with respect to minimum trade size or price movements. Such constraints are most likely to be more significant in markets that are less liquid.

Perhaps the more important role of cascades is likely to be in the decision of whether or not to participate at all rather than in the decision of whether to buy or sell at a particular price. If there is a fixed setup cost (perhaps psychic) of participating, there can be a substantial discreteness to individual decisions that does not rely on limiting the size of trades to a single unit. Cascades of participation versus nonparticipation may have important pricing effects.

Some of the most important puzzles in finance involve failures of investors to participate in asset classes. For example, there is the puzzle of insufficient participation in equity markets and the preference for participating in the markets for local and familiar stocks, which includes the home bias.[33] The phenomenon of underpriced neglected

[32]When investors are risk averse and action choices are discrete, even with endogenous price, cascades proper form (Décamps and Lovo, 2006).

[33]Huberman (2001) provides evidence and insightful discussion indicating that individuals prefer to invest in familiar stocks; Cao, Han, Hirshleifer, and Zhang (2008) model the effects of familiarity on economic decisions and capital markets. There is also evidence of local preferences in investment for both institutional investors (Coval and Moskowitz, 1999) who achieve better performance on local investments (Coval and Moskowitz, 2001), and individual investors (Zhu, 2003; Ivkovich and Weisbenner, 2005) who in some studies do (Ivkovich and Weisbenner, 2005) and in others do not (Zhu, 2003; Seasholes and Zhu, 2005). On the home bias puzzle of international finance, see Tesar and Werner (1995) and Lewis (1999).

stocks (which is exogenous in the model of Merton, 1987) can also be viewed as a puzzle of nonparticipation.

Cascades within a purely rational setting offer a partial explanation. Suppose that investors are more likely to interact and observe the behavior of other investors in the same income class than those in different income classes. If a group of low-income investors do not invest in the stock market (possibly for historical reasons or because some are deterred by fixed costs of participation), their choices can trigger a cascade of nonparticipation. Similarly, if investors in a locality observe (or discuss) each other's investment choices, and if some individuals invest locally (for historical or informational reasons), this can trigger a cascade of local investment.

Such arguments may require greater fleshing out to provide a complete explanation for the puzzles, since rationally an investor should draw inferences from what investors in other groups are doing. For example, even if one's peers are not investing in the stock market, there is also information to be gleaned from the fairly obvious fact that someone else is. Of course, an individual may believe that the benefits of stock investing are heterogeneous and that her own benefits are more similar to those of her peers than the nonpeers who are making different choices.

Psychological biases can reinforce the effects of rational social learning, which may help explain cascades of nonparticipation. For example, suppose that some individuals irrationally fear and avoid unfamiliar stocks. Then other investors who observe low participation in such stocks may draw adverse inferences about the benefits of participation, causing a cascade. This occurs if investors draw such inferences in a quasi-rational fashion that fails to adjust for the irrationality of others. More generally, cascades in market participation offer a rich avenue for theoretical analysis. For example, fragility in cascades of participation may help explain bubbles and crashes in stocks or portfolios.

There is starting to be some exploration of how the decisions of individuals over time as to whether or not to participate in trading cause information blockages to form and clear.[34] In settings with limited participation, large crashes can be triggered by minimal information, and the sidelining and entry of investors can cause skewness and volatility to vary, conditional on past price moves.

The use of the *availability heuristic* (Tversky and Kahneman, 1973) and the *mere exposure effect* (Moreland and Beach, 1992) should also affect participation in markets and the amount of buzz about or neglect of a stock. Using the availability heuristic, people judge how common something is by how easy it is to retrieve or imagine examples of it. Among other things, this causes vivid case examples to be too persuasive of the truth of a proposition. The mere exposure effect is the tendency for people to like things that they have been exposed to more than things they have not.

In the context of risk regulation, Kuran and Sunstein (1999) develop the notion of *availability cascades*, in which social processes together with availability bias make a belief or behavior self-reinforcing. Kuran and Sunstein define an availability cascade as "a self-reinforcing process of collective belief formation by which an expressed

[34]See Romer (1993); Lee (1998); Cao, Coval, and Hirshleifer (2002); and Hong and Stein (2003).

perception triggers a chain reaction that gives the perception of increasing plausibility through its rising availability in public discourse" (1999, 283). Availability cascades can result either from information cascades or reputational effects or a combination of the two.

Availability cascades offer a possible explanation for security market bubbles and waves of corporate events. Highly favorable publicity about a firm or market theory makes supportive positive arguments more salient and "available" to investors. Furthermore, mere exposure should also make the firm or transaction more familiar and therefore more appealing. Such effects make enthusiasm for investment self-reinforcing. The phenomenon of hot versus cold IPOs and of sudden excitement at different times about types of transactions (hostile takeovers, leveraged buyours, asset-backed securities, and so forth) seem to be availability cascades.

Popular allegations that securities markets are irrational often emphasize the contagiousness of emotions such as panic or frenzy. Critics often go on to argue that this causes excess volatility, destabilizes markets, and makes financial systems fragile (see, e.g., the critical review of Bikhchandani and Sharma, 2001, and references therein). There is indeed evidence that the contagious spread of emotions affects perceptions and behavior (see, e.g., Hatfield, Cacioppo, and Rapson, 1993; Barsade, 2002).

Prevailing models of capital market trading and equilibrium are quite limited in the forms of social influence and information transmission that they accommodate. This applies both to classical models of information and securities markets such as Grossman and Stiglitz (1976), Kyle (1985), and most of the work reviewed here on cascades and herding in capital markets.

The first premise is that the beliefs and behavior of an investor affect others only through market price.[35] This is perhaps most easily seen in the "rational expectations" approach of Grossman and Stiglitz (1976), wherein uninformed investors observe the market price, draw an inference from that price about the beliefs of informed traders, and trade based on their resulting beliefs.

The second premise is there is no localization in the influence of an investor on others. This rules out conversation with neighbors within a social network (which could be organized geographically or in other ways) and in which media may have influential nodes.

The evidence discussed in Section 1.8 suggests that social interactions between individuals affect financial decisions, which suggests that the social or geographical localization of information may be an important part of the process by which trading

[35]In the Capital Asset Pricing Model (CAPM) and its generalizations, beliefs are exogenous and investors form demands for securities, taking prices as given. In consequence, an investor's characteristics affect other investors only through the effect of his behavior on prices. In models of information and securities market, interactions between investors are still mediated by price. In Kyle (1985), a competitive market maker quotes a pricing function that fixes the price of the security as a function of aggregate demand. Investors submit demand quantities, and their orders are fulfilled at a price determined by the pricing function. So, one investor affects another solely through the effect of his order on price (via the pricing function). In Glosten and Milgrom (1985), the market maker sets a bid–ask spread, and each investor (who can be either informed or uninformed) submits a buy or sell order that is executed at the prespecified price. An investor is affected by the previous actions of other investors solely through the effect of those actions on bid–ask prices.

behaviors spread. Furthermore, there are often strategic complementarities or threshold effects in social processes, wherein the adoption of a belief or behavior by a critical number of individuals leads to a tipping in favor of one behavior versus another (Granovetter, 1978; Schelling, 1978; Kuran 1989).

Thus, an important direction for further empirical research is to examine how and whether a localized process of contagion of beliefs and attitudes affects stock markets (see, e.g., Shiller, 2000a) and whether securities market price patterns are consistent with rational models of contagion. Shive (2008) finds that "socially motivated trades" (as measured based on the predictive power of an epidemiological model) predict future returns and that this effect does not reverse out; this suggests that such trading helps aggregate meaningful information.

An important direction for future theoretical research is to examine the implications for securities market trading, participation, and prices of conversation between individuals (see, e.g., DeMarzo, Vayanos, and Zwiebel, 2001; Ozsoylev, 2005). DeMarzo, Vayanos, and Zwiebel (2003) provide a general analysis of the effects of *persuasion bias* in social networks. Individuals do not adjust appropriately for the fact that the information they receive from others may have come from a single common source. Individuals who have more connections in the social network have a stronger influence on beliefs in the network (even if their information is not more accurate).

Ozsoylev (2005) provides a model in which investors learn by observing the signals of other investors to which they are linked within a social network. Investors who belong to a relatively tightly integrated social cluster have relatively high correlation in their investment positions. Furthermore, although investors are rational, those who have many social network connections have a stronger influence on asset prices than those with fewer connections.

Agency-induced herding by investment managers, such as reputational effects, can also create mispricing (Brennan, 1993; Dasgupta and Prat, 2005). Brennan (1993) analyzes the asset-pricing implications of index-herding behavior. Any price effects of herding caused by agency problems should be driven by the behavior of institutional investors. Dasgupta, Prat, and Verardo (2005) test their model prediction that the reputation return premium increases with the past trades of career-driven traders and find that buying by such traders predicts return underperformance, and selling predicts overperformance. There is evidence suggesting that institutional investors are a source of positive portfolio return serial correlations (both own and cross correlations of the securities held by institutions); see Sias and Starks (1997). There is also evidence that the autocorrelation of the returns of emerging stock markets increased sharply at the time that institutional investors were expanding their positions in emerging markets (Aitken, 1998). Aitken argues that this indicates the effect of fluctuating sentiment by institutional investors.

There is a substantial literature on contagion between the debt or equity markets of various nations (see, e.g., Bikhchandani and Sharma, 2001). With regard to price effects of herding, there are some large correlations in returns, but it is hard to measure whether this is an effect of herding. Many studies have examined how the occurrence of a crisis in one country affects the probability of crisis in another country (Berg and Pattillo, 1999, review this research).

Experimental asset markets have been found to be capable of aggregating a great deal of the private information of participants; however, in complex environments the literature has found that blockages form so that information aggregation is imperfect.[36] Experimental laboratory research (see, e.g., Cipriani and Guarino, 2007) and field experiments (see Alevy, Haigh, and List, 2007, who test for cascades among financial market professionals on the Chicago Board of Trade) are promising ways of testing the way cascades and herding affect information aggregation in markets and will offer important directions for improving theories of financial herding.

Correlated sentiment for certain kinds of real assets may move prices and resources for nonfundamental reasons. Gompers and Lerner (2000) provide evidence of "money-chasing deals" in venture capital. Inflows into venture capital funds are associated with higher valuations of the new investments made by these funds, but not with the ultimate success of the firms. However, Froot, O'Connell, and Seasholes (2001) find that portfolio flows in and out of 44 countries during 1994 to 1998 were *positive* forecasters of future equity returns, with statistical significance in emerging markets.

1.10. CASCADES AND HERDING IN FIRM BEHAVIOR

It is often alleged in the popular press that managers are foolishly prone to fads in management methods (for examples and formal analysis, see Strang and Macy, 2001), investment choices, and disclosure or reporting practices. We first consider investment and financing choices and then disclosure and reporting.

1.10.1. Investment and Financing Decisions

Managers learn by observing the actions and performance of other managers, both within and across firms. This suggests that firms will engage in herding and be subject to information cascades, leading to management fads in accounting, financing and investment decisions. We have already discussed models of social learning and investment decisions.[37]

As for stylized facts, the popularity of various investment valuation methods, securities to issue, and so on have certainly waxed and waned. There are booms and quiet periods in new issues of equity that are related to past stock market returns and to the past average initial returns from buying an IPO (see, e.g., Lowry and Schwert, 2002; Ritter and Welch, 2002; Rau and Stouraitis, 2008). However, it is not easy to test whether fluctuations in investments and strategies are the result of irrationality, rational but imperfect aggregation of private information signals, or direct responses to fluctuations in public observables.

[36]See Noeth et al. 2002; Bloomfield and Libby (1996); and the survey of Libby, Bloomfield, and Nelson (2002).

[37]In addition, a growing literature analyzes how rates of learning vary during macroeconomic fluctuations and how the learning process causes booms and crashes in levels of investment (see, e.g., Gonzalez, 1997; Chamley, 1999; and Veldkamp, 2000).

Takeover markets have been subject to seemingly idiosyncratic booms and busts of enormous scale, such as the wave of conglomerate mergers in the 1960s and 1970s, in which firms diversified across various industries, the subsequent refocusing of firms through restructuring and bustup takeovers in the 1980s, followed by the merger boom of the 1990s. Targets of a takeover bid are "put into play" and often quickly receive competing offers, despite the negative cost externality of having a competitor. A distinct kind of evidence suggesting that the decision to engage in a takeover is contagious is that within a 1981–90 sample; a firm was more likely to merge if one of its top managers was a director of another firm that had engaged in a merger during the preceding three years (Haunschild, 1993).

Several studies test for herd behavior in corporate investment decisions. This raises the question of whether there is a general tendency toward strategic imitation.[38] Survey evidence on Japanese firms indicates that a factor that encourages firms to engage in direct investment in an emerging economy in Asia is whether other firms are investing in that country. This is consistent with possible cascading based on a manager's perception that rival firms possess useful private information about the desirability of such investment (Kinoshita and Mody, 2001). Gilbert and Lieberman (1987) examined the relationships among the investments of 24 chemical producers over two decades. They found that larger firms in an industry tend to invest when their rivals do not. In contrast, smaller firms tend to be followers in investment. This behavior is consistent with the "fashion leader" version of the cascades model (Bikhchandani, Hirshleifer, and Welch, 1992) in which the less precise free-ride informationally on the more precise (if large firms have greater absolute benefit from acquiring precise information or if there are scale economies in information acquisition).

Analogous to herding on endorsements in individual investor trading, there are endorsement effects in real investments. Often investments are clustered in time, but it is important to evaluate carefully whether investors are really imitating each other. Real estate investment is a natural field of application for cascades/endorsement effects because the investment decisions are discrete and conspicuous. (Caplin and Leahy, 1998, analyze real estate herding/cascading.)

Economists have studied agglomeration economies as an explanation for geographical concentration of investment and economic activity. Such effects are surely important, but geographical concentration can occur without agglomeration economies owing to learning by observation of others—"spurious agglomeration" (DeCoster and Strange, 1993). Barry, Görg, and Strobl (2004) test between agglomeration economies and what they call "demonstration effects," whereby a firm locates in a host country because the presence of other firms there provides information about the attractiveness of the host country. They conclude that both effects are important.

[38]D'Arcy and Oh (1997) study cascades in the decisions of insurers to underwrite risks and the pricing of insurance. Foresi, Hamao, and Mei (1998) provide evidence consistent with imitation in the investment decisions of Japanese firms. Greve (1998) provides evidence of firm imitation in the choice of new radio formats in the United States. Kedia and Rajgopal (2006) find geographical clustering in the decision by firms to use option compensation for rank and file workers.

1.10.2. Disclosure and Reporting Decisions

Since investors make comparisons in evaluating firms, the disclosure and reporting practices of one firm impose externalities on others. This point is reinforced by the likelihood that the voluntary disclosure/reporting practices chosen by firms help establish informal standards for other firms (though such practices are also, of course, subject to regulation).

Disclosure and reporting practices vary over time. For example, in recent years it has been popular for U.S. firms to disclose *pro forma* earnings in ways that differ from the GAAP-permitted definitions on firms' financial reports. Regulators have expressed concern about this practice, but firms argue that this allows them to reflect better long-term profitability by adjusting for nonrecurring items. It is also possible that firms are just herding or exploiting herd behavior by investors.

Surprisingly, Pincus and Wasley (1994) report that voluntary accounting changes by firms do not appear to be clustered in time and industry, suggesting no herding behavior in accounting changes. This result further suggests that firms do not switch accounting methods in response to changes in macroeconomic investment conditions that are experienced at about the same time by similar firms within an industry. Rather, the voluntary accounting changes would appear to be made in response to firm-specific needs, such as a firm-specific need to manage earnings.

However, it is not obvious why firms would manage earnings in response to firm-specific but not common factor shocks. One speculative possibility is that there is a concern for relative performance, as reflected in the model of Zwiebel (1995), combined with some deviation from perfect rationality that causes investors to adjust imperfectly for accounting method in evaluating firms' earnings.[39] The concern for relative performance may create a stronger incentive for managers to manage earnings upward when the firm is doing poorly relative to peers than when the entire industry is doing poorly.

Geographical proximity does affect reporting practices. Kedia and Rajgopal (2008) find that the accounting practices of neighboring firms are correlated with the likelihood that a firm misreports accounting items, resulting in accounting restatements. Presumably this is because managers of firms in close proximity interact and share information about their accounting practices. One would expect such an effect to be strongly in gray areas that could result in eventual restatement.

Furthermore, Kedia and Rajgopal (2008) find that counties that are farther from SEC regional offices have a higher frequency of income-decreasing restatements, and Defond, Francis, and Hu (2008) find that distance from Securities and Exchange Commission (SEC) regional offices affect the likelihood of severe problems, SEC enforcement actions, and consequently the quality of the auditor and the financial statements. To the extent that word-of-mouth communication among auditors, managers, and

[39]For example, Daniel, Hirshleifer, and Teoh (2002) suggest that owing to limited attention, investors are too credulous in the sense of failing to adjust for the interested motives of firms and that this explains actions such as issuing equity or failing to disclose information. Hirshleifer, Lim, and Teoh (2004) provide a model of how informed parties adjust their disclosure decisions to exploit the limited attention of observers.

regulators is localized, we expect a greater chance that information about questionable reporting practices will leak to regulators when enforcement offices are close.

Option backdating affects the way that compensation levels are reported to investors. Bizjak, Lemmon, and Whitby (2007) find that option backdating seems to spread contagiously through interlocked boards of directors and auditor links and are also affected by geographical proximity. This suggests that word-of-mouth discussion between managers and directors spreads corporate behaviors.

1.11. CONTAGION OF FINANCIAL MEMES

Rational and behavioral models of information in capital markets usually focus on very simple signals and signal processing. Such modeling does not capture the way that people accept or rule out arguments based on chains of reasoning, triggered associations, and emotional reactions. So, most existing information models say little about how thoughts and beliefs about firms or capital markets spread from person to person.

In reality, investors use verbal "reasons" to decide how to trade, and these reasons are often not cogent. Barber, Heath, and Odean (2003) find that individual investors and their investment clubs tend to buy stocks for which plausible reasons are available (such as a firm being on a media list of most admired companies). The reasons examined do not predict superior trading performance. Reasons seem to be spread by social interaction; stock clubs favor such stocks more strongly than individuals.

Reasons, or financial *memes* (units of cultural replication, as defined in the introduction), can be simple ("buy on the dips") or elaborate (e.g., portfolio theory). The contagion of such memes, their effects on markets, and (more ambitiously) the way combinations of memes evolve as they move from person to person are the subjects of a missing chapter in financial theory.

Leading the charge, Robert Shiller and coauthors have discussed and provided survey evidence about "popular models" during bubbles in securities and real estate markets (see Footnote 1).[40] For example, Shiller, Kon-Ya, and Tsutsui (1996) find that during the late 1980s to early 1990s U.S. investors were much more prone to viewing the Japanese stock market as overpriced than were Japanese investors and that the popularity of this meme among Japanese investors followed market movements. They also report that a substantial fraction of investors viewed the market as overpriced but recommended staying in the market because it was going to go up before it fell—a bubble-promoting viewpoint. More research is needed about what affects the popularity of the "The market is overpriced (or underpriced)" and the "The market is rising (or falling)" memes.

Case and Shiller (1989; 2003) find that real estate investors in "glamour cities" have unrealistic beliefs about the risk and expected appreciation of residential housing and

[40]See also Lynch (2000) for intriguing discussion. The phrase *popular model* implicitly connotes some kind of analysis (even if erroneous). But a meme can become popular without any supporting analysis.

that these beliefs seem to come from "simplistic theories" that are more prevalent in glamour cities than in a nonglamour city (Milwaukee).[41]

Apart from Shiller's discussions and pioneering surveys, behavioral economics and finance have focused mainly on the effects of individual psychological biases and the way individuals interact through buying and selling as mediated by price. The memetic approach focuses specifically on the social transmission mechanism (e.g., conversation or news media communication); a meme can replicate only through information that is conveyed from one person to another. Furthermore, memetics focuses on the idea's ability to hijack the psychological mechanisms of its carriers, much the way a virus hijacks cellular mechanisms within the bodies of carriers to replicate itself.

Thus, the memetic approach is not primarily about importing ideas from social psychology into behavioral finance. It is about the way memes exploit the process by which ideas are stored and replicated. This suggests a different kind of question from those typically asked in behavioral economics or in social psychology. A standard question is, "What are the kinds of biases that individuals are prone to?" Somewhat closer to memetics is the question. "How do individual biases affect whether people will be tempted to succumb to some exogenously specified popular idea about financial markets?" Some less conventional questions: "What are the characteristics of an *idea* (meme) in relation to its environment that help it replicate and predominate within the population of ideas?"[42] "How do financial ideologies evolve?" Thus, the memetic perspective suggests different hypotheses, approaches to modeling, and tests.[43,44]

In terms of modeling, a memetic approach considers shifts in population frequencies, as in population genetics models and models of disease contagion. The memetic stance further suggests developing models in which the meme is the (as-if) optimizer, selecting characteristics that will promote its own reproduction. This is analogous to models of adaptiveness in evolutionary biology in which genes have maximal reproductive

[41]One such meme is the idea that more desirable real estate tends to appreciate more rapidly. Case and Shiller (2003, p. 325) propose and provide survey evidence that this is because people "confuse the level of prices with the rate of change." Another fallacy is the idea that when housing is scarce, price becomes irrelevant.

[42]There is a population of people and a population of ideas that people hold. If an idea spreads from one person to another, it has reproduced, and its frequency within the population of ideas has risen. Of course, individual psychology (and the answers to the first two questions) is still crucial for answering memetic questions.

[43]Of course, a correct memetic understanding must be consistent with a correct psychological understanding, just as correct biochemistry must be consistent with correct genetics or physics. But the memetic stance suggests different insights and research questions.

The memetic (i.e., a Darwinian, or adaptationist) stance focuses on how selection results in the evolution of well-adapted cultural variants. This puts the spotlight on the meme metaphorically striving to be fit. A mechanical stance puts the spotlight on forces that push meme frequencies up or down. So long as it makes sense to think about adaptiveness, these perspectives are two sides of the same coin and must be consistent.

There is also subtle discussion among scientists and philosophers as to how to define *replicator*. For our purposes the terminological discussion of whether to call cultural variants *memes* is not crucial, so long as it is accepted that natural selection on cultural variants leads to cumulative evolution and adaptation.

[44]Several objections have been raised about the memetic approach to cultural evolution. This is not the place to review these issues, but we will mention that even some authors who deemphasize the need for replicators (Henrich and Boyd, 2002) still champion the propositions that natural selection operates on cultural variants and that this results in cumulative evolution of culture.

success. As discussed earlier, Shive (2008) models and tests the contagion of investment behaviors through a population, analogous to the spread of a disease. Similarly, Shiller (2000b, Chapter 8) likens the spread of ideas about the stock market to a contagion spread through conversation. Epidemiological models have also been applied to the spread of ideas (see Bartholomew, 1982), an approach that could easily be applied to the spread of an investment idea (e.g., "Sell your losers, ride your winners") as an infection.

Similarly, the rate at which financial ideas replicate in populations can be measured. Shiller 2007, discussed further later) measures the rise and fall in popularity of the concept of the real rate of interest. Natural further memetic questions are, In what ways are this meme adaptive to its environment? How well does it exploit the mechanisms that transmit ideas from person to person? The answer to this will depend on the characteristics of this meme and its environment (most notably the existing population of memes and the population of people who carry those memes). For example, the concept of the real rate requires explanation to be understood. Furthermore, during periods of relatively low and stable inflation, the real/nominal distinction becomes less useful in daily life, which hinders the meme's spread.[45]

Some characteristics that help a meme reproduce include logical cogency, ease of cognitive processing (portfolio theory would be more popular if it were easy to learn), and emotional vividness (hence the popularity of dysfunctional investment memes that promise ways to "get rich quick"). The characteristics of the environment that matter include the effectiveness of unrelated memes that struggle for attention in people's minds and the cogency or vividness of directly competing memes that oppose the given meme. For example, the meme "Sell your losers, ride your winners" is opposed by the meme "Buy or sell based on expectations, not history."

A great puzzle in finance is why individual investors trade actively or place money with actively trading money managers. This effort to beat the market leads to wasted transaction cost and (if correlated investor sentiment affects price) to subnormal gross returns. Information cascades and memetics can potentially help explain the seductiveness of active investing. If friends and acquaintances are investing in an active mutual fund, an individual who learns this but does not observe their return performance can be in a cascade and imitate.

Furthermore, a conversational bias helps the survival of the meme that "Active trading is the road to exceptional profits," even if it harms its carriers. People like to talk more about their gains than their losses. If listeners do not discount for this bias, they will overestimate the benefits of active trading. Such effects should be exacerbated when an individual spends more time discussing investments, as occurs when an individual participates in an investment chat room, message board, or investment club.

Biological evolution has led to the development of coadapted teams of genes that produce organisms. Distin (2004) calls coadapted sets of memes *assemblies*,

[45] Berger and Heath (2005) perform several field and experimental laboratory tests of the effects of "idea habitat," the range of situations in which people have the opportunity to use a meme, on its success in reproduction.

examples being religions, scientific theories, and philosophies. We will refer to *financial ideologies*, which could be good or bad. The meme "dead cat bounce" is too simple and disjointed from other financial memes to be an ideology or part of an ideology. But the efficient markets hypothesis is an elaborate financial ideology (and a largely valid one).

It is intuitive that *valid* financial ideologies would develop through cumulative evolution and prosper, but invalid ideologies also do so. Although the surveys of Shiller and his coauthors do not focus on distinguishing simple financial memes from assemblies, their research is a first step toward identifying financial ideologies and devising theories of how they develop.

To give some hints of the memetic approach, we discuss some speculative examples of how popular financial ideologies seem to be adaptive (that is, good at spreading themselves). Shiller (2000b) provides insightful historical review of popular theories that developed and prospered during past market booms. It seems that periods of rising stock prices, which investors observe and seek to explain, provide a salubrious habitat for "new era theories" of the stock market that seem to evolve repeatedly and independently.[46] Such repeated evolution of similar ideologies is reminiscent of the repeated independent biological evolution of behaviors (flight) or organs (the eye). It raises the hope that the adaptiveness concept (for memes, not people) can help explain cumulative memetic evolution.

When an environmental shift creates an exceptional opportunity for a meme to spread, an availability cascade is often triggered (see Section 1.9). The availability cascade concept applies to spasms of activity in public discourse (as with bubbles or hot IPOs) and to the psychological effects of availability. The field of memetics concerns all aspects of the evolution of populations of ideas, including both periods of dynamic shifts and stable evolutionary equilibria, and including all psychological forces that affect the reproduction memes.

The marketing of an IPO can be viewed as a way of trying to trigger an availability cascade wherein the idea that the company is a good investment suddenly becomes popular.[47] In the dot-com period an availability cascade supported the meme that the Internet was a uniquely spectacular investment opportunity. This meme was mutually reinforcing with more specific memes (and availability cascades) about particular dot-com startups, such as Netscape.

As discussed in Section 1.4.3, in bad times people crave safety and in good times opportunity (Mullainathan and Shleifer, 2005). Mullainathan and Shleifer find that advertisers adjust their persuasion activities accordingly. More generally, even apart from the exploitive efforts of sellers, during downturns we expect memes about danger versus safety to spread in popular discussions, and during upturns we expect memes about opportunity. The memes that investment clubs or day-trading are ways for individual investors to achieve exceptional performance became popular during the 1990s,

[46]Shiller (Chapter 5) discusses remarkable parallels between the turn-of-the-century optimism of 1901, the 1920s optimism, the new era thinking of the 1950s and 1960s, and new era thinking in the 1990s.
[47]Pollock, Rindova, and Maggitti (2008) discuss information and availability cascades in IPOs.

a period of rising stock prices. This was probably due in part to a self-feeding effect in which individuals learned from advertising, the media, and word of mouth that many others were engaged in these activities and supposedly profiting thereby. The rise and fall of these practices therefore seem to be availability cascades.

These examples are a kind of microevolution of memes in response to changing market conditions. Perhaps the most basic microevolution in financial markets is the perpetual seesawing contest between bullish and bearish memes. These fluctuate in response to fundamental news, the rise through mutation of salient new memes ("The Internet changes everything"), and emotional factors that influence the public mood.

The value ideology versus growth ideologies also swing widely, the tech boom and bust being the most spectacular recent example. In good times, speculative stocks (growth options) tend do well, providing evidence in favor of growth investing. Furthermore, optimistic mood encourages the pursuit of opportunity. In bad times the evidence from recent returns favors the value ideology. Emotionally, investors seek greater safety. Furthermore, a pessimistic mood causes investors to place a high premium on frugality. The emphasis of the value-investing approach on measures of "cheapness" of stocks plays to this trend.

Furthermore, the value ideology provides an ego-boosting narrative of folly (by others, the greedy speculators) and wisdom (by the self, the wise and prudent value investor). The moralistic tone of value-investing expositions in the popular media is striking. This reveals a successful strategy used by the value meme assembly.

Beneath the microevolutionary tug of war between the value and growth ideologies are deep wells of appeal that allow both to coexist in the long run. We conjecture that frequency-dependent natural selection supports coexistence between conflicting investment ideologies in general. When there are many growth investors, growth stocks become overvalued, increasing the subsequent profits of value investors; greater profits generate more favorable stories that can be passed on, spreading the value ideology. The reverse occurs when there are few growth investors. A similar reasoning suggests that bullish versus bearish ideologies are self-limiting. In the language of genetics, we expect a balanced polymorphism.

Shiller (2007) argues that changes in popular economic models are central to understanding asset overvaluation. Following Modigliani and Cohn (1979), he proposes that owing to money illusion on the part of investors, low nominal interest rates raise asset valuations. Investors who are unaware of the real-versus-nominal interest rate distinction discount at nominal rates without adjusting forecasts of future dividends for forecasts of inflation. He documents the historical rise in public references to "real interest rates" and finds that recently the use of the phrase has dropped precipitously. He also documents the recent rise in popularity of the meme that global markets are "awash with liquidity," by performing a Lexis-Nexis search of this phrase.

There are also managerial fads and stable popular theories about management methods (e.g., EVA), capital budgeting techniques (e.g., payback versus NPV), and other aspects of corporate investment and financing policy. As discussed in Section 1.4.3, there is evidence that some corporate behaviors spread across firms through the interlocking social networks of managers and boards. Goldfarb, Kirsch, and Miller (2007) model the influence of the "get big fast" ideology about the best strategy for dot-com startups to achieve success; they provide evidence suggesting that it resulted in overinvestment by existing firms and too little entry. (This ideology was motivated in large part by the economic theory literature on network externalities and increasing returns.) They model the adoption of this meme as an information cascade.

In summary, to develop a memetic understanding of the micro- and macroevolution of these ideologies, research is needed to understand how they exploit investors' psychological needs. Financial ideologies also influence public policy. Hirshleifer (2008) argues that an ideology of anti-short-termism combines a number of ideas that are conceptually unrelated but emotionally linked and that this meme assembly is an important driver of financial and accounting regulation.

1.12. CONCLUSION

Observers have long wondered whether sudden large shifts in prices, actions, and resource allocation in capital markets are caused by social contagion. Such phenomena include asset and security market booms and crashes; real investment booms and busts; takeover and financing waves (including boom/bust patterns in the design of securities, such as asset-backed securities); shifts in disclosure practices, such as the use of *pro forma* earnings disclosures; and financial market runs. Often the behaviors of relevant participants converge despite negative payoff externalities, suggesting that such patterns may be caused by contagious effects of psychological bias or by the failure of rational social learning to aggregate information efficiently.

We review here the causes and effects of contagion of thoughts and behaviors in capital markets among participants such as investors, managers, security analysts, advisors, and media commentators. We examine theory and evidence about how rational observational learning, agency problems, and psychological forces affect contagion in firms' investment, financing, and disclosure choices and in investors' trading decisions and the consequences of contagion for the pricing of real assets and securities. In addition to herding by managers or other agents, we consider how such agents exploit the readiness of investors to herd.

An externality problem is central to the theory of rational observational learning. Each individual maximizes his own payoff without regard to the effect of his choice on the information obtained by later decision makers by observing the individual's action choice and/or the payoff consequences of that choice. Over time, individuals act mainly based on the accumulated inventory of public information (perhaps including payoff information) generated by past actions. This delays or blocks the generation and revelation of further information.

Thus, in a range of economic settings, even if payoffs are independent and people are rational, decisions tend to quickly converge on mistaken action choices. In other words, the resulting cascade or herd is idiosyncratic. Owing to idiosyncrasy, rational individuals place only modest faith in the conventional action; the cascade or herd is easily dislodged (fragility). Furthermore, rational observational learning tends to cause simultaneity (delay followed by a sudden spasm of joint action), paradoxicality (increasing the availability of public information by various means can decrease welfare and decision accuracy), and path dependence (chance early events have big effects on ultimate outcomes).

Depending on the exact assumptions, information may be completely suppressed for a period (until a cascade is dislodged). Under other assumptions, no cascades form and asymptotically all uncertainty is resolved, but too slowly (relative to full aggregation of private signals). Information cascades require discrete, bounded, or gapped action space (or cognitive constraints); these conditions are highly plausible in many investment settings. Even when these conditions fail owing to noise, the growth in accuracy of the public information pool tends to be self-limiting because an individual who places heavy weight on the information derived from past actions puts little weight on his own signal, making his own action less informative to those who follow.

Markets have been praised as marvels of spontaneous information aggregation (Hayek, 1945). Indeed, we see that rational price setting in perfect markets encourages investors to use their private signals rather than imitating the trades of others, discouraging direct information cascades of trading. However, in settings with multiple dimensions of uncertainty, quasi-cascading behavior can occur in which individuals trade in opposition to their information signals.

Furthermore, transaction costs, minimum trade sizes, or psychological biases can prevent price from fully aggregating information. As a result, proper trading cascades can form, including cascades in participation versus nonparticipation in markets. Even without frictions or biases, information cascades (and other sources of herd behavior) in investigation indirectly affect trading and the amount of information aggregated into market price. So, despite some arguments to the contrary, in several economic settings information cascades affect securities trading and prices.

Although inefficient behaviors can be locked in by noninformational factors (such as positive payoff externalities), the theories of rational observational learning, especially the information cascades theory, differ in their implication of fragility. Therefore, payoff externality models are helpful in explaining herds that seem stable and robust.[48] Reputational models, for example, generate stable patterns of herding or dispersing through endogenously generated payoff externalities. Reputational models help explain when herding versus dispersing will occur and offer implications about the effect on the pressure to herd on the career status of managers.

We need new models of price setting in financial markets in which beliefs are not transmitted solely through price, in which ideas spread between neighbors in a social network and by means of electronic media. Furthermore, we need analysis that reflects the fact that the information conveyed is not necessarily a simple normally distributed signal, processed in some rational or quasi-rational fashion. Often what are conveyed are investment ideas or memes, and we need to understand how both isolated financial memes (e.g., "This stock is going to rise") and full-fledged financial ideologies (such as

[48]However, informational effects including cascades can still occur in settings with payoff externalities, leading to idiosyncratic (though not fragile) outcomes.

"new era" theories and the value and growth ideologies) spread from person to person as a sort of social epidemic. Empirical testing will of course be crucial to the success of a research program based on thought contagion.

Such a program involves understanding the spread of particular financial memes viewed as an epidemic, studying the characteristics of financial memes that tend to promote their own replication, and assessing the characteristics of the environment that favor different kinds of memes. In other words, it involves studying what makes some financial memes better adapted than others. Even more ambitiously, it is important to understand why certain financial memes complement each other and how this leads to the cumulative evolution of financial ideologies that are well adapted in the sense that they are good at spreading themselves.

We have suggested that many of the puzzles and anomalies of capital markets can be understood by going beyond a focus on individual psychological biases and the way biased individuals interact through trading and market price. Instead, as Robert Shiller has proposed, we need to understand the social processes that lead to the spread of popular ideas. For dynamic phenomena such as bubbles and crashes, the concept of availability cascades is especially promising. We have offered some tentative speculative memetic explanations for such phenomena as the survival of invalid capital budgeting methods and money-losing active trading strategies, bubbles, hot IPOs, short-run fluctuations in the profitability of value versus growth strategies and frequencies of the value and growth ideologies, and long-run coexistence of these ideologies.

References

Abreu, D., and M. K. Brunnermeier, Bubbles and crashes, *Econometrica*, 2003, 71, 173–204.

Admati, A., and P. Pfleiderer, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies*, 1988, 1, 3–40.

Aharony, J., and I. Swary, Additional evidence on the information-based contagion effects of bank failures, *Journal of Banking and Finance*, 1996, 20, 57–69.

Aitken, B., Have institutional investors destabilized emerging markets? *Contemporary Economic Policy*, 1998, 16, 173–184.

Alevy, J. E., M. S. Haigh, and J. A. List, Information cascades: Evidence from a field experiment with financial market professionals, *Journal of Finance*, 2007, 62, 151–180.

Allen, F., and D. Gale, Financial contagion, *Journal of Political Economy*, 2000, 108, 1–33.

Amihud, Y., S. Hauser, and A. Kirsh, Allocations, adverse selection, and cascades in IPOs: Evidence from the Tel Aviv Stock Exchange, *Journal of Financial Economics*, 2003, 68, 137–158.

Anderson, L. R., Payoff effects in information cascade experiments, *Economic Inquiry*, 2001, 39, 609–615.

Anderson, L. R., and C. A. Holt, Information cascades in the laboratory, *American Economic Review*, 1996, 87, 847–862.

Arthur, B., Competing technologies, increasing returns, and lock-in by historical events, *Economic Journal*, 1989, 99, 116–131.

Ashiya, M., and T. Doi, Herd behavior of Japanese economists, *Journal of Economic Behavior and Organization*, 2001, 46, 343–346.

Avery, C., and P. Zemsky, Multi-dimensional uncertainty and herd behavior in financial markets, *American Economic Review*, 1998, 88, 724–48.

Bala, V., and S. Goyal, Learning from neighbors, *Review of Economic Studies*, 1998, 65, 595–621.

Banerjee, A., A simple model of herd behavior, *Quarterly Journal of Economics*, 1992, 107, 797–817.

Banerjee, A., and D. Fudenberg, Word-of-mouth learning, *Games and Economic Behavior*, 2004, 46, 1–22.

Barber, B., C. Heath, and T. Odean, Good reasons sell: Reason-based choice among individual investors in the stock market, *Management Science*, 2003, 49, 1636–1652.

Barry, F., H. Görg, and E. Strobl, Foreign direct investment, agglomerations, and demonstration effects: An empirical investigation, *Review of World Economics*, 2004, 127, 583–600.

Barsade, S. G., The ripple effect: Emotional contagion within groups, *Administrative Science Quarterly*, 2002, 47, 644–675.

Bartholomew, D. J., *Stochastic Models for Social Processes* (3rd ed.), Wiley, 1982.

Beatty, R., Auditor reputation and the pricing of initial public offerings, *Accounting Review*, 1989, 64, 693–709.

Beatty, R. P., and J. R. Ritter, Investment banking, reputation, and the underpricing of initial public offerings, *Journal of Financial Economics*, 1986, 15, 213–232.

Beaudry, P., and F. M. Gonzalez, An equilibrium analysis of information aggregation in investment markets with discrete decisions, *Journal of Economic Theory*, 2003, 113, 76–103.

Berg, A., and C. Pattillo, Are currency crises predictable? A test, *IMF Staff Papers*, 1999, 46, 107–138.

Berger, J. A., and C. Heath, Idea habitats: How the prevalence of environmental cues influences the success of ideas, *Cognitive Science*, 2005, 29, 195–221.

Bernardo, A. E., and I. Welch, On the evolution of overconfidence and entrepreneurs, *Journal of Economics and Management Strategy*, 2001, 10, 301–330.

Bernardo, A. E., and I. Welch, Liquidity and financial market runs, *Quarterly Journal of Economics*, 2004, 119, 135–158.

Bernhardt, D., M. Campello, and E. Kutsoati, Who herds? *Journal of Financial Economics*, 2006, 80, 657–675.

Bernheim, B. D., A theory of conformity, *Journal of Political Economy*, 1994, 102, 841–877.

Bhattacharya, S., K. Chatterjee, and L. Samuelson, Sequential research and the adoption of innovations, *Oxford Economic Review*, 1986, 38, 219–243.

Bhattacharya, S., and A. Thakor, Contemporary banking theory, *Journal of Financial Intermediation*, 1993, 3, 2–50.

Bhushan, R., Firm characteristics and analyst following, *Journal of Accounting and Economics*, 1989, 11, 255–274.

Biais, B., P. Hillion, and C. Spatt, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance*, 1995, 50, 1655–89.

Bikhchandani, S., D. Hirshleifer, and I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy*, 1992, 100, 992–1026.

Bikhchandani, S., D. Hirshleifer, and I. Welch, Learning from the behavior of others: Conformity, fads, and informational cascades, *Journal of Economic Perspectives*, 1998, 12, 151–170.

Bikhchandani, S., D. Hirshleifer, and I. Welch, Information cascades, in S. N. Durlauf and L. E. Blume (eds.), *The New Palgrave Dictionary of Economics Online* (2nd ed.), Palgrave Macmillan, 2008.

Bikhchandani, S., D. Hirshleifer, and I. Welch, Informational cascades and rational herding: An annotated bibliography and resource reference, working paper, Anderson School of Management, UCLA, available at www.info-cascades.info, 2008.

Bikhchandani, S., and S. Sharma, Herd behavior in financial markets, *IMF Staff Papers*, 2001, 47, 279–310.

Bizjak, J. M., M. L. Lemmon, and R. J. Whitby, Option backdating and board interlocks, working paper, Portland State University, 2007.

Bloomfield, R., and R. Libby, Market reactions to differentially available information, *Journal of Accounting Research*, 1996, 34, 183–207.

Boyd, R., and P. J. Richerson, *The Origin and Evolution of Cultures*, Oxford University Press, 2005.

Brandenburger, A. A., and B. Polak, When managers cover their posteriors: Making the decisions the market wants to see, *Rand Journal of Economics*, 1996, 27, 523–541.

Brennan, M. J., Latent assets, *Journal of Finance*, 1990, 45, 709–730.

Brennan, M. J., Agency and asset pricing, working paper, Anderson School of Management, UCLA #6-93, 1993.

Brock, W. A., and S. N. Durlauf, Interactions-based models, NBER working paper no. T0258, 2000.

Brown, J. R., Z. Ivkovich, P. A. Smith, and S. Weisbenner, Neighbors matter: Causal community effects and stock market participation, *Journal of Finance*, 2008, 63, 1509–1531.

Brown, K. C., W. V. Harlow, and L. T. Starks, Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *Journal of Finance*, 1996, 51, 85–110.

Brown, N. C., K. D. Wei, and R. R. Wermers, Analyst recommendations, mutual fund herding, and overreaction in stock prices, working paper, Leventhal School of Accounting, University of Southern California, 2008.

Brown, P., G. Foster, and E. Noreen, Security analyst multi-year earnings forecasts and the capital market, in *Studies in Accounting Research, no. 21*, 1985, American Accounting Association.

Brunnermeier, M., *Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis and Herding*, Oxford University Press, 2001.

Burguet, R., and X. Vives, Social learning and costly information acquisition, *Journal of Economic Theory*, 2000, 15, 185–205.

Calomiris, C., and J. Mason, Contagion and bank failures during the Great Depression: The June 1932 Chicago banking panic, *American Economic Review*, 1997, 87, 863–883.

Calomiris, C. W., and G. Gorton, The origins of banking panics: Models, facts, and bank regulation, in R. G. Hubbard (ed.), *Financial Markets and Financial Crises*, pp. 109–173, University of Chicago Press, 1991.

Calomiris, C. W., and J. R. Mason, Causes of US bank distress during the depression, working paper, LeBow College of Business, Drexel University, 2001.

Calvo, G., and E. Mendoza, Rational contagion and globalization of securities markets, *Journal of International Economics*, 2001, 51, 79–113.

Caminal, R., and X. Vives, Price dynamics and consumer learning, *Journal of Economics and Management Strategy*, 1999, 8, 95–131.

Cao, H. H., J. D. Coval, and D. Hirshleifer, Sidelined investors, trading-generated news, and security returns, *Review of Financial Studies*, 2002, 15, 615–648.

Cao, H. H., B. Han, D. Hirshleifer, and H. H. Zhang, Fear of the unknown: Familiarity and economic decisions, working paper, Merage School of Business, UC Irvine, 2008.

Cao, H. H., and D. Hirshleifer, Limited observability, reporting biases, and informational cascades, working paper, Fisher College of Business, The Ohio State University, 1997.

Cao, H. H., and D. Hirshleifer, Taking the road less traveled: Does conversation eradicate pernicious cascades? Dice Center working paper Series 2002-8, The Ohio State University, 2002.

Caplin, A., and J. Leahy, Sectoral shocks, learning and aggregate fluctuations, *Review of Economic Studies*, 1993, 60, 777–794.

Caplin, A., and J. Leahy, Business as usual, market crashes, and wisdom after the fact, *American Economic Review*, 1994, 84, 548–565.

Caplin, A., and J. Leahy, Miracle on sixth avenue: Information externalities and search. *Economic Journal*, 1998, 108, 60–74.

Capstaff, J., K. Paudyal, and W. Rees, Analysts' forecasts of German firms' earnings: A comparative analysis. *Journal of International Financial Management and Accounting*, 1998, 9, 83–116.

Case, K. E., and R. J. Shiller, The efficiency of the market for single-family homes, *American Economic Review*, 1989, 79, 125–137.

Case, K. E., and R. J. Shiller, Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, 2003, 34, 299–362.

Celen, B., and S. Kariv, Distinguishing informational cascades from herd behavior in the laboratory, *American Economic Review*, 2004, 94, 484–497.

Celen, B., and S. Kariv, An experimental test of observational learning under imperfect information, *Economic Theory*, 2005, 26, 677–699.

Chamley, C., Coordinating regime switches, *Quarterly Journal of Economics*, 1999, 114, 869–905.

Chamley, C., and D. Gale, Information revelation and strategic delay in irreversible decisions, *Econometrica*, 1994, 62, 1065–1085.

Chamley, C. P., Delays and equilibria with large and small information in social learning, *European Economic Review*, 2004, 48, 477–501.

Chamley, C. P., *Rational Herds: Economics Models of Social Learning*, Cambridge University Press, 2004.

Chari, V., and P. J. Kehoe, Hot money, *Journal of Political Economy*, 2003, 11, 1262–1292.

Chari, V., and P. J. Kehoe, Financial crises as herds: Overturning the critiques, *Journal of Economic Theory*, 2004, 119, 128–150.

Chari, V. V., and R. Jagannathan, Banking panics, information and rational expectation equilibrium, *Journal of Finance*, 1988, 43, 749–761.

Chen, Y., Banking panics: The role of the first-come, first-served rule and information externalities, *Journal of Political Economy*, 1999, 107, 946–968.

Chevalier, J., and G. Ellison, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy*, 1997, 105, 1167–1200.

Chevalier, J., and G. Ellison, Career concerns of mutual fund managers, *Quarterly Journal of Economics*, 1999, 114, 389–432.

Chowdhry, B., and V. Nanda, Multimarket trading and market liquidity. *Review of Financial Studies*, 1991, 4, 483–511.

Cipriani, M., and A. Guarino, Herd behavior and contagion in financial markets, working paper, New York University, 2003.

Cipriani, M., and A. Guarino, Herd behavior in a laboratory financial market, *American Economic Review*, 2005, 95, 1427–1443.

Cipriani, M., and A. Guarino, Transaction costs and informational cascades in financial markets: Theory and experimental evidence. European Central Bank working paper no. 736, 2007.

Clement, M. B., and S. Y. Tse, Financial analyst characteristics and herding behavior in forecasting, *Journal of Finance*, 2005, 60, 307–341.

Cohen, L., A. Frazzini, and C. J. Malloy, Sell-side school ties, Harvard Business School Finance working paper no. 08-074, 2008.

Cohen, L., A. Frazzini, and C. J. Malloy, Valuing reciprocity, working paper, Harvard Business School, 2008.

Conlisk, J., Why bounded rationality? *Journal of Economic Literature*, 1996, 34, 669–700.

Cote, I., and D. Sanders, Herding behavior: Explanations and implications, *Behavioral Research in Accounting*, 1997, 9, 20–45.

Coval, J. D., and T. J. Moskowitz, Home bias at home: Local equity preference in domestic portfolios, *Journal of Finance*, 1999, 54, 145–166.

Coval, J. D., and T. J. Moskowitz, The geography of investment: Informed trading and asset prices, *Journal of Political Economy*, 2001, 109, 811–841.

Daniel, K. D., D. Hirshleifer, and S. H. Teoh, Investor psychology in capital markets: Evidence and policy implications, *Journal of Monetary Economics*, 2002, 49, 139–209.

D'Arcy, S. P., and P. Oh. The cascade effect in insurance pricing, *Journal of Risk and Insurance*, 1997, 64, 465–480.

Dasgupta, A., and A. Prat, Reputation and asset prices: A theory of information cascades and systematic mispricing, working paper, London School of Economics, 2005.

Dasgupta, A., A. Prat, and M. Verardo, The price of conformism, London School of Economics working paper, 2005.

Dawkins, R., *The Selfish Gene*, Oxford University Press, 1976.

Décamps, J.-P., and S. Lovo, Informational cascades with endogenous prices: The role of risk aversion, *Journal of Mathematical Economics*, 2006, 42, 109–120.

DeCoster, G. P., and W. C. Strange, Spurious agglomeration, *Journal of Urban Economics*, 1993, 33, 273–304.

Defond, M. L., J. R. Francis, and X. Hu, The geography of auditor independence and SEC enforcement, working paper, Leventhal School of Accounting, University of Southern California, 2008.

DellaVigna, S., and E. Kaplan, The Fox News effect: Media bias and voting, *The Quarterly Journal of Economics*, 2007, 122, 1187–1234.

DellaVigna, S., and J. Pollet, Investor inattention, firm reaction, and Friday earnings announcements, working paper, U.C. Berkeley, 2006.

DellaVigna, S., and J. M. Pollet, Demographics and industry returns, *American Economic Review*, 2007, 97, 1667–1702.

DeLong, J. B., A. Shleifer, L. Summers, and R. J. Waldmann, Noise trader risk in financial markets, *Journal of Political Economy*, 1990, 98, 703–738.

DeMarzo, P., R. Kaniel, and I. Kremer, Technological innovation and real investment booms and busts, *Journal of Financial Economics*, 2007, 85, 735–754.

DeMarzo, P., D. Vayanos, and J. Zwiebel, Social networks and financial markets, working paper, MIT and Stanford University, 2001.

DeMarzo, P., D. Vayanos, and J. Zwiebel, Persuasion bias, social influence, and uni-dimensional opinions, *Quarterly Journal of Economics*, 2003, 118, 909–968.

Devenow, A., and I. Welch, Rational herding in financial economics, *European Economic Review*, 1996, 40, 603–615.

Diamond, D. W., and P. H. Dybvig, Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 1983, 91, 410–419.

Distin, K., *The Selfish Meme: A Critical Reassessment*, Cambridge University Press, 2004.

Docking, D., M. Hirschey, and E. Jones, Information and contagion effects of bank loan-loss reserve announcements, *Journal of Financial Economics*, 1997, 43, 219–239.

Drehmann, M., J. Oechssler, and A. Roider, Herding and contrarian behavior in financial markets: An Internet experiment. *American Economic Review*, 2005, 95, 1403–1426.

Duflo, E., and E. Saez, Participation and investment decisions in a retirement plan: The influence of colleagues' choices, *Journal of Public Economics*, 2002, 85, 121–148.

Duflo, E., and E. Saez, The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment, *Quarterly Journal of Economics*, 2003, 118, 815–842.

Dugatkin, L. A., Sexual selection and imitation: Females copy the mate choice of others, *American Naturalist*, 1992, 139, 138–149.

Dugatkin, L. A., and J. J. Godin, Reversal of female mate choice by copying in the guppy (*poecilia reticulata*), *Proceedings of the Royal Society of London, Series B*, 1992, 249, 179–184.

Ehrbeck, T., and R. Waldmann, Why are professional forecasters biased? Agency versus behavioral explanations, *Quarterly Journal of Economics*, 1996, 111, 21–40.

Eichengreen, B., A. Rose, and C. Wyplosz, Contagious currency crises: First tests, *Scandinavian Journal of Economics*, 1996, 98, 463–484.

Eleswarapu, V., Demography and intensity of stock trading: The effect of social interaction, working paper, Cox School of Business, Southern Methodist University, 2004.

Ellison, G., and D. Fudenberg, Rules of thumb for social learning, *Journal of Political Economy*, 1993, 101, 612–643.

Ellison, G., and D. Fudenberg, Word of mouth communication and social learning, *Quarterly Journal of Economics*, 1995, 110, 93–126.

Feng, L., and M. Seasholes, Correlated trading and location, *Journal of Finance*, 2004, 59, 2117–2144.

Foresi, S., Y. Hamao, and J. Mei, Interaction in investment among rival Japanese firms, *Japan and the World Economy*, 1998, 10, 393–407.

Froot, K., D. Scharfstein, and J. C. Stein, Herd on the street: Informational inefficiencies in a market with short-term speculation, *Journal of Finance*, 1992, 47, 1461–1484.

Froot, K. A., P. O'Connell, and M. Seasholes, The portfolio flows of international investors, *Journal of Financial Economics*, 2001, 59, 151–193.

Gale, D., What have we learned from social learning? *European Economic Review*, 1996, 40, 617–628.

Gervais, S., Market microstructure with uncertain information precision: A multiperiod analysis, working paper, Wharton School, University of Pennsylvania, 1996.

Gibson, R. M., and J. Hoglund, Copying and sexual selection, *Trends in Research in Evolution and Ecology*, 1992, 7, 229–232.

Gilbert, R. J., and M. Lieberman, Investment and coordination in oligopolistic industries, *RAND Journal of Economics*, 1987, 18, 17–33.

Giraldeau, L.-A., The ecology of information use, in J. Krebs and N. B. Davies (eds.), *Behavioral Ecology: An Evolutionary Approach* (4th ed.), Blackwell, 1997.

Givoly, D., and J. Lakonishok, The quality of analysts' forecast of earnings, *Financial Analysts Journal*, 1984, 40, 40–47.

Givoly, D., and D. Palmaon, Insider trading and the exploitation of inside information: Some empirical evidence, *Journal of Business*, 1985, 58, 69–87.

Glaeser, E. L., and J. Scheinkman, Non-market interactions, NBER working paper no. W8053, 2000.

Glosten, L. R., and P. R. Milgrom, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics*, 1985, 14, 71–100.

Goeree, J., T. Palfrey, B. Rogers, and R. McKelvey, Self-correcting information cascades, *Review of Economic Studies*, 2007, 74, 733–762.

Goldfarb, B., D. Kirsch, and D. A. Miller, Was there too little entry during the dot-com era? *Journal of Financial Economics*, 2007, 86, 100–144.

Gompers, P., and J. Lerner, Money chasing deals? The impact of fund inflows on private equity valuations, *Journal of Financial Economics*, 2000, 55, 281–325.

Gonzalez, F., Individual experimentation and aggregate fluctuations. working paper, University of British Columbia Discussion Paper No. 97–24, 1997.

Gorton, G., Bank suspension of convertibility, *Journal of Monetary Economics*, 1985, 15, 177–193.

Gorton, G., Banking panics and business cycles, *Oxford Economic Papers*, 1988, 40, 751–781.

Graham, J. R., Herding among investment newsletters: Theory and evidence, *Journal of Finance*, 1999, 54, 237–268.

Granovetter, M., The strength of weak ties, *American Journal of Sociology*, 1973, 78, 1360–1380.

Granovetter, M., Threshold models of collective behavior, *American Journal of Sociology*, 1978, 83, 1420–1443.

Grenadier, S. R., Information revelation through option exercise, *Review of Financial Studies*, 1999, 12, 95–129.

Greve, H. R., Managerial cognition and the mimetic adoption of market positions: What you see is what you do, *Strategic Management Journal*, 1998, 19, 967–988.

Griffiths, M., B. Smith, A. Turnbull, and R. White, Information flows and open outcry: Evidence of imitation trading, *Journal of International Financial Markets, Institutions and Money*, 1998, 8, 101–116.

Grinblatt, M., and M. Keloharju, The investment behavior and performance of various investor types: A study of Finland's unique data set, *Journal of Financial Economics*, 2000, 55, 43–67.

Grinblatt, M., M. Keloharju, and S. Ikaheimo, Social influence and consumption: Evidence from the automobile purchases of neighbors, *Review of Economics and Statistics*, 90, 2008.

Grinblatt, M., S. Titman, and R. Wermers, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review*, 1995, 85, 1088–1105.

Grossman, S. J., and J. E. Stiglitz, Information and competitive price systems, *American Economic Review*, 1976, 66, 246–253.

Gul, F., and R. Lundholm, Endogenous timing and clustering of agents' decisions, *Journal of Political Economy*, 1995, 103, 1039–1066.

Gupta-Mukherjee, S., Informal information networks: The impact on performance of mutual fund portfolios, working paper, Loyola University Chicago, 2007.

Haltiwanger, J., and M. Waldman, Limited rationality and strategic complements: The implications for rationality, *Quarterly Journal of Economics*, 1989, 104, 463–483.

Hatfield, E., J. L. Cacioppo, and R. L. Rapson, Emotional contagion, *Current Directions in Psychological Sciences*, 1993, 2, 96–99.

Haunschild, P. R., Interorganizational imitation: The impact of interlocks on corporate acquisition activity, *Administrative Science Quarterly*, 1993, 38, 564–592.

Hayek, F. V., The use of knowledge in society, *American Economic Review*, 1945, 35, 519–530.

Hendricks, K., and D. Kovenock, Asymmetric information, information externalities, and efficiency: The case of oil exploration, *RAND Journal of Economics*, 1989, 20, 164–182.

Henrich, J., and R. Boyd, The evolution of conformist transmission and the emergence of between-group differences, *Evolution and Human Behavior*, 1998, 19, 215–242.

Henrich, J., and R. Boyd, On modeling cognition and culture: Why cultural evolution does not require replication of representations, *Journal of Cognition and Culture*, 2002, 2, 87–112.

Hey, J. D., and A. Morone, Do markets drive out lemmings—or vice versa? *Economica*, 2004, 71, 637–659.

Hirshleifer, D., Investor psychology and asset pricing, *Journal of Finance*, 2001, 64, 1533–1597.

Hirshleifer, D., Psychological bias as a driver of financial regulation, *European Financial Management*, forthcoming, 2008.

Hirshleifer, D., S. S. Lim, and S. H. Teoh, *Disclosure to an audience with limited attention*, Fisher College of Business, Ohio State University, 2004.

Hirshleifer, D., and R. Noah, Misfits and social progress, working paper, Ohio State University, 1999.

Hirshleifer, D., A. Subrahmanyam, and S. Titman, Security analysis and trading patterns when some investors receive information before others, *Journal of Finance*, 1994, 49, 1665–1698.

Hirshleifer, D., and S. H. Teoh, Herd behavior and cascading in capital markets: A review and synthesis, *European Financial Management*, 2003a, 9, 25–66.

Hirshleifer, D., and S. H. Teoh, Limited attention, information disclosure, and financial reporting, *Journal of Accounting & Economics*, 2003b, 36, 337–386.

Hirshleifer, D., and I. Welch, An economic approach to the psychology of change: Amnesia, inertia, and impulsiveness, *Journal of Economics and Management Strategy*, 2002, 11, 379–421.

Hong, H., J. Kubik, and J. C. Stein, Social interactions and stock market participation, *Journal of Finance*, 2004, 59, 137–163.

Hong, H., J. Kubik, and J. C. Stein, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *Journal of Finance*, 2005, 60, 2801–2824.

Hong, H., and J. C. Stein, Differences of opinion, short-sales constraints and market crashes, *Review of Financial Studies*, 2003, 16, 487–525.

Hong, H. G., J. Kubik, and A. Solomon, Security analysts' career concerns and herding of earning forecasts, *RAND Journal of Economics*, 2000, 31, 121–144.

Huberman, G., Familiarity breeds investment, *Review of Financial Studies*, 2001, 14, 659–680.

Hung, A., and C. Plott, Information cascades: Replication and an extension to majority rule and conformity rewarding institutions, *American Economic Review*, 2001, 91, 1508–1520.

Ivkovich, Z., and S. Weisbenner, Local does as local is: Information content of the geography of individual investors' common stock investments, *Journal of Finance*, 2005, 60, 267–306.

Ivkovich, Z., and S. Weisbenner, Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices, *Review of Financial Studies*, 2007, 20, 1327–1357.

Jacklin, C., and S. Bhattacharya, Distinguishing panics and information-based bank runs: Welfare and policy implications, *Journal of Political Economy*, 1988, 96, 568–592.

Jaffe, J. F., and J. M. Mahoney. The performance of investment newsletters, *Journal of Financial Economics*, 1999, 53, 289–307.

Jegadeesh, N., and W. Kim, Do analysts herd? An analysis of recommendations and market reactions, NBER working papers, no. 12866, 2007.

Kahn, C. M., G. G. Pennacchi, and B. Sopranzetti, Bank deposit rate clustering: Theory and empirical evidence, *Journal of Finance*, 1999, 54, 2185–2214.

Kaniel, R., G. Saar, and S. Titman, Individual investor trading and stock returns, *Journal of Finance*, 2008, 63, 273–310.

Kedia, S., and S. Rajgopal, Neighborhood matters: The impact of location on broad-based stock option plans, working paper, Rutgers Business School, 2006.

Kedia, S., and S. Rajgopal, Geography and the incidence of financial misreporting, working paper, 2008.

Kelly, M., and C. O. O'Grada, Market contagion: Evidence from the panics of 1854 and 1857, *American Economic Review*, 2000, 90, 1110–1124.

Khanna, N., Optimal contracting with moral hazard and cascading, *Review of Financial Studies*, 1997, 11, 559–596.

Khanna, N., and S. L. Slezak, The effect of organizational form on information aggregation and project choice: The problem of informational cascades, *Journal of Economics and Management Strategy*, 2000, 9, 115–156.

Kim, C., and C. Pantzalis, Analyst herding behavior and firm diversification: Evidence from panel data, working paper, City University of New York – Queens College, 2000.

Kinoshita, Y., and A. Mody, Private information for foreign investment in emerging economies, *Canadian Journal of Economics*, 2001, 34, 448–464.

Kodres, L. E., and M. Pritsker, Directionally similar position taking and herding by large futures market participants, working paper, IMF and Board of Governors of the Federal Reserve System, 1997.

Krugman, P. R., and A. J. Venables, Globalisation and the inequality of nations, *Quarterly Journal of Economics*, 1995, 110, 857–880.

Kumar, A., and C. M. Lee, Retail investor sentiment and return comovements, *Journal of Finance*, 2006, 61, 2451–2486.

Kuran, T., Sparks and prairie fires: A theory of unanticipated political revolution, *Public Choice*, 1989, 61, 41–74.

Kuran, T., and C. Sunstein, Availability cascades and risk regulation, *Stanford Law Review*, 1999, 51, 683–768.

Kyle, A. S., Continuous auctions and insider trading, *Econometrica*, 1985, 53, 1315–1335.

Lakonishok, J., A. Shleifer, and R. W. Vishny, The impact of institutional trading on stock prices, *Journal of Financial Economics*, 1992, 32, 23–43.

Lee, I. H., On the convergence of informational cascades, *Journal of Economic Theory*, 1993, 61, 395–411.

Lee, I. H., Market crashes and informational avalanches, *Review of Economic Studies*, 1998, 65, 741–759.

Lewis, K. K., Trying to explain home bias in equities and consumption, *Journal of Economic Literature*, 1999, 37, 571–608.

Libby, R., R. Bloomfield, and M. W. Nelson, Experimental research in financial accounting, *Accounting, Organizations and Society*, 2002, 27, 775–810.

Lowry, M., and G. W. Schwert, IPO market cycles: Bubbles or sequential learning? *Journal of Finance*, 2002, 57, 1171–1200.

Lux, T., Herd behavior, bubbles and crashes, *Economic Journal*, 1995, 105, 881–896.

Lynch, A., Thought contagions in the stock market, *Journal of Psychology and Financial Markets*, 2000, 1, 10–23.

Madrian, B., and D. Shea, The power of suggestion: Inertia in 401(k) participation and savings behavior, *Quarterly Journal of Economics*, 2001, 116, 1149–1187.

Manski, C. F., Identification of endogenous social effects: The reflection problem, *Review of Economic Studies*, 1993, 60, 531–42.

Martin, G. S., and J. Puthenpurackal, Imitation is the sincerest form of flattery: Warren Buffett and Berkshire Hathaway, working paper, American University, April 2008.

Massa, M., and A. Simonov, The role of college interaction in portfolio choice and stock market prices, CEPR working paper 4815, 2005.

Merton, R. C., A simple model of capital market equilibrium with incomplete information, *Journal of Finance*, 1987, 42, 483–510.

Michaely, R., and W. H. Shaw, Does the choice of auditor convey quality in an initial public offering? *Financial Management*, 1995, 24, 15–30.

Mikhail, M. B., B. R. Walther, and R. H. Willis, Does forecast accuracy matter to security analysts? *Accounting Review*, 1999, 74, 185–200.

Modigliani, F., and R. Cohn, Inflation, rational valuation, and the market, *Financial Analyst Journal*, 1979, 35, 24–44.

Moreland, R. L., and R. Beach, Exposure effects in the classroom: The development of affinity among students, *Journal of Experimental Social Psychology*, 1992, 28, 255–276.

Moscarini, M., M. Ottaviani, and L. Smith, Social learning in a changing world, *Economic Theory*, 1998, 13, 657–665.

Mullainathan, S., and A. Shleifer, Persuasion in finance, NBER working paper no. 11838, 2005.

Neeman, Z., and G. Orosel, Herding and the winner's curse in markets with sequential bids, *Journal of Economic Theory*, 1999, 85, 91–121.

Ng, L. K., and F. Wu, Peer effects in the trading decisions of individual investors, working paper, University of Wisconsin-Milwaukee, 2006.

Noeth, M., C. F. Camerer, C. R. Plott, and M. Weber, Information aggregation in experimental asset markets: Traps and misaligned beliefs, *Review of Financial Studies*, 2002, 15, 927–957.

Nofsinger, J. R., and R. W. Sias, Herding and feedback trading by institutional and individual investors, *Journal of Finance*, 1999, 54, 2263–2295.

Ottaviani, M., and P. N. Sørensen, Herd behavior and investment: Comment, *American Economic Review*, 2000, 90, 695–704.

Ottaviani, M., and P. N. Sørensen, Information aggregation in debate: Who should speak first? *Journal of Public Economy*, 2001, 81, 393–421.

Ottaviani, M., and P. N. Sørensen, The strategy of professional forecasting, *Journal of Financial Economics*, 2006, 81, 441–466.

Ozsoylev, H. N., Asset pricing implications of social networks, working paper, Saïd Business School and Linacre College, University of Oxford, 2005.

Park, A., and D. Sgroi, When herding and contrarianism foster market efficiency: A financial trading experiment, University of Toronto working paper 316, 2008.

Peng, L., and W. Xiong, Investor attention, overconfidence and category learning, *Journal of Financial Economics*, 2006, 80, 563–602.

Persons, J. C., and V. A. Warther, Boom and bust patterns in the adoption of financial innovations. *Review of Financial Studies*, 1997, 10, 939–967.

Pincus, M., and C. Wasley, The incidence of accounting changes and characteristics of firms making accounting changes, *Accounting Horizons*, 1994, 8, 1–24.

Pollock, T. G., V. Rindova, and P. G. Maggitti, Market watch: Information and availability cascades among the media and investors in the U.S. IPO market, *Academy of Management Journal*, 2008, 51, 335–358.

Prendergast, C., A theory of yes men, *American Economic Review*, 1993, 83, 757–770.

Prendergast, C., and L. Stole, Impetuous youngsters and jaded oldtimers: An analysis of behavioral decision-making rules, *Journal of Political Economy*, 1996, 104, 1105–1134.

Rajan, R. G., Why credit policies fluctuate: A theory and some evidence, *Quarterly Journal of Economics*, 1994, 109, 399–442.

Rao, H., H. R. Greve, and G. F. Davis, Fool's gold: Social proof in the initiation and abandonment of coverage by Wall Street analysts, *Administrative Science Quarterly*, 2001, 46, 502–526.

Rau, R., and A. Stouraitis, The timing of corporate event waves, working paper, City University of Hong Kong, 2008.

Richardson, S., S. H. Teoh, and P. D. Wysocki, The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives, *Contemporary Accounting Research*, 2004, 21, 885–924.

Ritter, J. R., and I. Welch, A review of IPO activity, pricing, and allocations, *Journal of Finance*, 2002, 57, 1795–1828.

Romer, D., Rational asset-price movements without news, *American Economic Review*, 1993, 83, 1112–1130.

Sacerdote, B., Peer effects with random assignment: Results for Dartmouth roommates, *Quarterly Journal of Economics*, 2001, 116, 681–704.

Samuelson, W., and R. Zeckhauser, Status-quo bias in decision making, *Journal of Risk and Uncertainty*, 1988, 1, 7–59.

Saunders, A., and B. Wilson, Contagious bank runs: Evidence from the 1929–1933 period, *Journal of Financial Intermediation*, 1996, 5, 409–423.

Scharfstein, D. S., and J. C. Stein, Herd behavior and investment, *American Economic Review*, 1990, 80, 465–479.

Schelling, T. C., *Micromotives and Macrobehavior*, Norton, 1978.

Seasholes, M. S., and N. Zhu, Is there information in the local portfolio choices of individuals?, working paper, U. C. Berkeley, September 2005.

Sgroi, D., The right choice at the right time: A herding experiment in endogenous time, *Experimental Economics*, 2003, 6, 159–180.

Shiller, R. J., Conversation, information, and herd behavior, *American Economic Review*, 2000, 85, 181–185.

Shiller, R. J., *Irrational Exuberance*, Princeton University Press, 2000.

Shiller, R. J., Low interest rates and high asset prices: An interpretation in terms of changing popular economic models, NBER working paper no. W13558, 2007.

Shiller, R. J., F. Kon-Ya, and Y. Tsutsui, Why did the Nikkei crash? Expanding the scope of expectations data collection, *Review of Economics and Statistics*, 1996, 78, 156–164.

Shiller, R. J., and J. Pound, Survey evidence on the diffusion of interest and information among investors, *Journal of Economic Behavior and Organization*, 1989, 12, 46–66.

Shive, S., An epidemic model of investor behavior, *Journal of Financial and Quantitative Analysis*, forthcoming, 2008.

Sias, R. W., Institutional herding, *Review of Financial Studies*, 2004, 17, 165–206.

Sias, R. W., and L. Starks, Return autocorrelation and institutional investors, *Journal of Financial Economics*, 1997, 46, 103–131.

Simon, H. A., A behavioral model of rational choice, *Quarterly Journal of Economics*, 1955, 69, 99–118.

Simunic, D., Empirical assessment of the impact of auditor quality on the valuation of new issues, *Journal of Accounting and Economics*, 1991, 14, 375–399.

Smallwood, D., and J. Conlisk, Product quality in markets where consumers are imperfectly informed, *Quarterly Journal of Economics*, 1979, 93, 1–23.

Smith, L., and P. N. Sørensen, Pathological outcomes of observational learning, *Econometrica*, 2000, 68, 371–398.

Stickel, S. E., Predicting individual analyst earnings forecasts, *Journal of Accounting Research*, 1990, 28, 409–417.

Stickel, S. E., Reputation and performance among security analysts, *Journal of Finance*, 1992, 48, 1811–1836.

Strang, D., and M. Macy, In search of excellence: Fads, success stories, and adaptive emulation, *American Journal of Sociology*, 2001, 107, 147–182.

Teoh, S. H., Information disclosure and voluntary contributions to public goods, *Rand Journal of Economics*, 1997, 29, 385–406.

Tesar, L., and I. M. Werner, Home bias and high turnover, *Journal of International Money and Finance*, 1995, 14, 467–492.

Titman, S., and B. Trueman, Information quality and the valuation of new issues, *Journal of Accounting and Economics*, 1986, 8, 159–172.

Trueman, B., Analyst forecasts and herding behavior, *Review of Financial Studies*, 1994, 7, 97–124.

Tversky, A., and D. Kahneman, Availability: A heuristic for judging frequency and probability, *Cognitive Psychology*, 1973, 5, 207–232.

Uttal, B., Inside the deal that made Bill Gates $300,000,000, *Fortune*, July 21, 1986, 23–31.

Veldkamp, L. L., Slow boom, big crash, Stanford University, SIEPR policy paper no. 00-16, 2000.

Veldkamp, L. L., Media frenzies in markets for financial information, *American Economic Review*, 2006, 96, 577–601.

Vives, X., How fast do rational agents learn? *Review of Economic Studies*, 1993, 60, 329–347.

Vives, X., The speed of information revelation in a financial market mechanism, *Journal of Economic Theory*, 1995, 67, 178–204.

Welch, I., Sequential sales, leaning, and cascades, *Journal of Finance*, 1992, 47, 695–732.

Welch, I., Herding among security analysts, *Journal of Financial Economics*, 2000, 58, 369–396.

Wermers, R., Mutual fund herding and the impact on stock prices, *Journal of Finance*, 1999, 54, 581–622.

Zeira, J., Informational overshooting, booms and crashes, *Journal of Monetary Economics*, 1999, 43, 237–257.

Zhang, J., Strategic delay and the onset of investment cascades, *RAND Journal of Economics*, 1997, 28, 188–205.

Zhu, N., The local bias of individual investors, Yale ICF working paper, 2003.

Zitzewitz, E., Measuring herding and exaggeration by equity analysts and other opinion sellers, Stanford GSB working paper no. 1802, 2001.

Zwiebel, J., Corporate conservatism and relative compensation, *Journal of Political Economy*, 1995, 103, 1–25.

CHAPTER 2

How Markets Slowly Digest Changes in Supply and Demand

Jean-Philippe Bouchaud
Science & Finance, Capital Fund Management, Paris

J. Doyne Farmer
Santa Fe Institute and LUISS Guido Carli, Rome

Fabrizio Lillo
University of Palermo and Santa Fe Institute

2.1. Introduction 59
2.1.1. Overview 59
2.1.2. Organization 60
2.1.3. Motivation and Scope 61
2.1.4. Approach to Model Building 63
2.2. Market Structure 64
2.3. Information, Liquidity, and Efficiency 65
2.3.1. Information and Fundamental Values 66
2.3.2. Market Efficiency 67
2.3.3. Trading and Information 68
2.3.4. Different Explanations for Market Impact 69
2.3.5. Noise Trader Models and Informed vs. Uninformed Trading 70
2.3.6. A Critique of the Noise Trader Explanation of Market Impact 71
2.3.7. The Liquidity Paradox: Prices Are Not in Equilibrium 72
2.3.8. Time Scales and Market Ecology 73

Note: We want to thank Klaus Reiner Schenk-Hoppé and Thorsten Hens for giving us the opportunity to write this chapter and for their patience. We also want to thank our collaborators who helped us to shape our understanding of the subtle world of market microstructure and impact: R. Almgren, S. Ciliberti, C. Deremble, Z. Eisler, A. Gerig, L. Gillemot, E. Henry, A. Joulin, J. Kockelkoren, E. Moro, G. La Spada, Y. Lemperiere, R.N. Mantegna, S. Mike, A. Ponzi, M. Potters, G. Vaglica, H. Waelbroeck, J. Wieslinger, and M. Wyart. JDF gratefully acknowledges support from Bill Miller, Barclays Bank, and NSF grant HSD-0624351.

2.3.9. The Volatility Puzzle 75
2.3.10. The Kyle Model 76
2.4. Large Fluctuations and Long Memory of Order Flow 77
2.4.1. Empirical Evidence for Long Memory of Order Flow 77
2.4.2. On the Origin of Long Memory of Order Flow 79
2.4.3. Theory for Long Memory in Order Flow Based on Strategic Order Splitting 80
2.4.4. Evidence Based on Exchange Membership Codes 82
2.4.5. Evidence for Heavy Tails in Volume 83
2.5. Summary of Empirical Results for Diverse Types of Market Impact 84
2.5.1. Impact of Individual Transactions 86
2.5.2. Impact of Aggregate Transactions 86
2.5.3. Hidden Order Impact 88
2.5.4. Upstairs Market Impact 90
2.6. Theory of Market Impact 90
2.6.1. Why Is Individual Transaction Impact Concave? 91
2.6.2. A Fixed Permanent Impact Model 93
2.6.3. The MRR Model 94
2.6.4. A Transient Impact Framework 95
2.6.5. History Dependent, Permanent Impact 99
2.6.6. Empirical Results 103
2.6.7. Impact of a Large Hidden Order 106
2.6.8. Aggregated Impact 108
2.7. The Determinants of the Bid–Ask Spread 111
2.7.1. The Basic Economics of Spread and Impact 111
2.7.2. Models for the Bid–Ask Spread 114
2.7.3. Limit vs. Market Orders: The Microstructure Phase Diagram 117
2.7.4. Spread Dynamics After a Temporary Liquidity Crisis 123
2.8. Liquidity and Volatility 125
2.8.1. Liquidity and Large Price Changes 125
2.8.2. Volume vs. Liquidity Fluctuations as Proximate Causes of Volatility 127
2.8.3. Spread vs. Volatility 129
2.8.4. Market Cap Effects 132
2.9. Order Book Dynamics 133
2.9.1. Heavy Tails in Order Placement and the Shape of the Order Book 133
2.9.2. Volume at Best Prices: The Glosten-Sandas Model 135
2.9.3. Statistical Models of Order Flow and Order Books 137
2.10. Impact and Optimized Execution Strategies 142
2.11. Toward an Empirical Characterization of a Market Ecology 144
2.11.1. Identifying Hidden Orders 145
2.11.2. Specialization of Strategies 146
2.12. Conclusion 148
Appendix 2.1: Mechanical vs. Nonmechanical Impact 150
A2.1.1. Definition of Mechanical Impact for Order Books 150
A2.1.2. Empirical Results 152
Appendix 2.2: Volume Fluctuations 153
Appendix 2.3: The Bid–Ask Spread in the MRR Model 155
References 156

Abstract

In this chapter we revisit the classic problem of *tâtonnement* in price formation from a microstructure point of view, reviewing a recent body of theoretical and empirical work explaining how fluctuations in supply and demand are slowly incorporated into prices. Because revealed market liquidity is extremely low, large orders to buy or sell can only be traded incrementally, over periods of time as long as months. As a result order flow is a highly persistent long-memory process. Maintaining compatibility with market efficiency has profound consequences on price formation, on the dynamics of liquidity, and on the nature of impact. We review a body of theory that makes detailed quantitative predictions about the volume and time dependence of market impact, the bid–ask spread, order book dynamics, and volatility. Comparisons to data yield some encouraging successes. This framework suggests a novel interpretation of financial information, in which agents are at best only weakly informed and all have a similar and extremely noisy impact on prices. Most of the processed information appears to come from supply and demand itself, rather than from external news. The ideas reviewed here are relevant to market microstructure regulation, agent-based models, cost-optimal execution strategies, and understanding market ecologies.

Keywords: financial markets, market microstructure, price impact, market ecology

2.1. INTRODUCTION

In this chapter we discuss the slow process by which markets "digest" fluctuations in supply and demand, reviewing a body of work that suggests a new approach to the classic problem of *tâtonnement*—the dynamic process through which markets seek to reach equilibrium.

2.1.1. Overview

The foundation of this approach is based on several empirical observations about financial markets, the most important of which is long memory in the fluctuations of supply and demand. This is exhibited in the placement of trading orders, and corresponds to long-term, slowly decaying positive correlations in the initiation of buying vs. selling. It is observed in all the stock markets studied so far at very high levels of statistical significance. It appears that the primary cause of this long memory is the incremental execution of large hidden trading orders. The fact that the long memory of order flow must coexist with market efficiency (at least in a statistical sense) has a profound influence on price formation, causing dynamic adjustments of liquidity that are strongly asymmetric between buyers and sellers.

This has important consequences for market impact. (By market impact we mean the average response of prices to trades; liquidity refers to the scale of the market impact.[1]) We discuss theoretical work predicting the average market impact as a function of both volume and time. The asymmetric liquidity adjustments needed to maintain compatibility between the long memory of order flow and market efficiency can equivalently be interpreted in terms of the temporal response of market impact, leading to a slow decay of market impact with time.

This work also has important consequences about the interpretation and effect of information in financial markets. In particular, the explanation for market impact that we develop here differs from the standard view in the finance literature, which holds that the shape of the impact function is determined by differences in the information content of trades. The body of work reviewed here instead assumes that the impact of trades depends only on their predictability; for example, that highly predictable trades have little impact, as originally postulated by Hasbrouck (1988). We argue that this is a much simpler explanation that produces stronger predictions, is more plausible from a theoretical point of view, and is more in line with what is observed in the data.

The implications of this work range upward from microstructure—that is, at the level of individual price changes—to patterns of price formation on time scales that can be measured in months. At the microstructure level this work makes several predictions, such as the relationship between market impact, the bid–ask spread, and volatility. It also make predictions about the impact of large trades executed over long periods of time as well as the effect such trades may have in causing clustered volatility.

2.1.2. Organization

In the remainder of the introduction we discuss the motivation and scope of the work described here and discuss our approach to creating a theory for market microstructure, which is somewhat unusual within economics. In Section 2.2 we discuss the institutional aspects of the markets that form the basis of our empirical studies and define some of the terms that will be used throughout the paper. In Section 2.3 we lay out some of the main conceptual issues, discussing the concept of information in finance and its relationship to market efficiency and the important role that liquidity (or more accurately, the lack of liquidity) plays in forming markets. We critique so-called "noise trader" models and present an alternative point of view. In Section 2.4 we present the empirical evidence for long memory in order flow, develop a theory for its explanation based on strategic order splitting, and present evidence that this theory is correct. In Section 2.5 we describe the various types of impact and review the empirical evidence. In Section 2.6 we develop a theory for market impact for each type of impact. In Section 2.7 we discuss the problem of explaining the behavior of the bid–ask spread and compare theory and empirical observations. Section 2.8 discusses the close relationship between liquidity and volatility. In Section 2.9 we discuss models for the order book,

[1]Market impact is closely related to the demand elasticity of price and is typically measured as the return associated with a transaction as a function of volume. Liquidity (as we will use it here) measures the size of the price response to a trade of a fixed size and is inversely proportional to the scale of the impact. If trading a given quantity produces only a small price change, the market is liquid, and if it produces a large price change, it is illiquid.

which can be regarded as models for liquidity. Section 2.10 discusses the problem of trading in an optimal manner to minimize execution costs. Section 2.11 describes recent attempts to characterize trading ecologies of market behavior in short time scales, and Section 2.12 presents our conclusions.

2.1.3. Motivation and Scope

Markets are places where buyers meet sellers and the prices of exchanged goods are fixed. As originally observed by Adam Smith, during the course of this apparently simple process, remarkable things happen. The information of diverse buyers and sellers, which may be too complex and textured for any of them to fully articulate, is somehow incorporated into a single number—the price. One of the powerful achievements of economics has been the formulation of simple and elegant equilibrium models that attempt to explain the end results of this process without going into the details of the mechanisms through which prices are actually set.

There has always been a nagging worry, however, that there are many situations in which broad-brush equilibrium models that do not delve sufficiently deeply into the process of trading and the strategic nature of its dynamics may not be good enough to tell us what we need to know; to do better we will ultimately have to roll up our sleeves and properly understand how prices change from a more microscopic point of view. Walras himself worried about the process of *tâtonnement*, the way in which prices settle into equilibrium. While there are many proofs for the existence of equilibria, it is quite another matter to determine whether or not a particular equilibrium is stable under perturbations—that is, whether prices initially out of equilibrium will be attracted to an equilibrium. This necessarily requires a more detailed model of the way prices are actually formed. There is a long history of work in economics seeking to create models of this type (see e.g., Fisher, 1983), but many would argue that this line of work was ultimately not very productive, and in any case it has had little influence on modern mainstream economics.

A renewed interest in dynamical models that incorporate market microstructure is driven by many factors. In finance, one important factor is growing evidence suggesting that there are many situations where equilibrium models, at least in their current state, do not explain the data very well. Under the standard model prices should change only when there is news, but there is growing evidence that news is only one of several determinants of prices and that prices can stray far from fundamental values (Campbell and Shiller, 1989; Roll, 1984; Cutler et al., 1989; Joulin et al., 2008).[2] Doubts are further fueled by a host of studies in behavioral economics demonstrating the strong boundaries of rationality. Taken together this body of work calls into question the view that prices always remain in equilibrium and respond instantly and correctly to new information.

The work reviewed here argues that trading is inherently an incremental process and that for this reason, prices often respond slowly to new information. The reviewed body of theory springs from the recent empirical discovery that changes in supply and

[2]See Engle and Rangel (2005) for a dissenting view.

demand constitute a long-memory process; that is, that its autocorrelation function is a slowly decaying power law (Bouchaud et al., 2004; Lillo and Farmer, 2004). This means that supply and demand flow in and out of the market only very gradually, with a persistence that is observed on time scales of weeks or even months. We argue that this is primarily caused by the practice of order splitting, in which large institutional funds split their trading orders into many small pieces. Because of the heavy tails in trading size, there are long periods where buying pressure dominates and long periods where selling pressure dominates. The market only slowly and with some difficulty "digests" these swings in supply and demand. To keep prices efficient in the sense that they are unpredictable and there are not easy profit-making opportunities, the market has to make significant adjustments in liquidity. Understanding how this happens leads to a deeper understanding of many properties of market microstructure, such as volatility, the bid–ask spread, and the market impact of individual incremental trades. It also leads to an understanding of important economic issues that go beyond market microstructure, such as how large institutional orders impact the price and in particular how this depends on both the quantity traded and on time. It implies that the liquidity of markets is a dynamic process with a strong history dependence.

The work reviewed here by no means denies that information plays a role in forming prices, but it suggests that for many purposes this role is secondary. In the last half of the twentieth century, finance has increasingly emphasized information and deemphasized supply and demand. The work we review here brings forward the role of fluctuations and correlations in supply and demand, which may or may not be exogenous. As we view it, it is useful to begin the story with a quantitative description of the properties of fluctuations in supply and demand. Where such fluctuations come from doesn't really matter; they could be driven by rational responses to information or they could simply be driven by a demand for liquidity. In either case, they imply that there are situations in which order arrival can be very predictable. Orders contain a variable amount of information about the hidden background of supply and demand. This affects how much prices move and therefore modulates the way in which information is incorporated into prices. This notion of information is internal to the market. In contrast to the prevailing view in market microstructure theory, there is no need to distinguish between "informed" and "uninformed" trading to explain important properties of markets, such as the shape of market impact functions or the bid–ask spread.[3]

We believe that the work here should have repercussions on a wide gamut of questions:

- At a very fundamental level, how do we understand why prices move, how information is reflected in prices, and what fixes the value of the volatility?
- At the level of price statistics, what are the mechanisms leading to price jumps and volatility clustering?

[3] Since modern continuous double auction markets are typically anonymous, it is hard to see how the identity of traders could play an important role in the size of the price response to trades. See the discussion in Section 2.3.

- At the level of market organization, what are the optimal trading rules to ensure immediate liquidity and orderly flow to investors?
- At the level of agent-based models, what are the microstructural ingredients necessary to build a realistic agent-based model of price changes?
- At the level of trading strategies and execution costs, what are the consequence of empirical microstructure regularities on transaction costs and implementation shortfall?

We do not wish to imply that these questions will be answered here—only that the work described here bears on all of them. We will return to discussing the implications in our conclusions.

2.1.4. Approach to Model Building

Because this work reflects an approach to model building that many economists will find unfamiliar, we first make a few remarks to help the reader understand the philosophy behind this approach. Put succinctly, our view is that the enormous quantities of data that are now available fundamentally change the approach one should take to building economic theories about financial markets.

In recent years the computer has made it possible to automate markets, has enabled an explosion in the amount of recorded data, and has made it possible to analyze unprecedented quantities of information. Financial instruments are now typically standardized, stable entities that are traded day after day by many thousands of market participants. Modern electronic markets offer an open and transparent environment that allows traders across the world to get real-time access to prices, and most important for science, makes it possible to save detailed records of human decision making. The past decades have seen an explosion in the volume of stored data. For example, the total volume of data related to U.S. large caps on, say, October 2, 2007, was 57 million lines, approximately a gigabyte of stored data. The complete record of world financial activity is more than a terabyte per day. Each market has slightly different rules of operation, making it possible to compare market structures and the way they affect price formation and, most important of all, to look for patterns of behavior that are common across all market structures. The system of world financial markets can be viewed as a huge social science experiment in which profit seekers spend large quantities of their own money to collect enormous quantities of data for the pleasure of scientists.

With so much data it becomes possible to change the style in which economics is done. When one has only a small amount of noisy data, statistical testing must be done with great care, and it is difficult to test and reject competing models unless the differences in their predictions are very large. Data snooping is a constant worry. In contrast, with billions of data points, if an effect is not strong enough to leap out of the noise, it is unlikely to be of any economic importance. Even more important is the effect this has on developing and testing theories. With a small data set inference requires strong priors. This fosters an approach in which one begins with pure theory and tests the resulting models only after they are fully formulated; there is less opportunity to let the data speak for itself. Without great quantities of data it is difficult to test a theory in a fully quantitative manner, and so predictions of theories are typically qualitative.

The work reviewed in this chapter takes advantage of the size of financial data sets by strongly coupling the processes of model formation and data analysis. This begins with a search for empirical regularities, that is, behaviors that under certain circumstances follow consistent quantitative laws. Even though such effects do not have the consistency of the laws of physics, one can nonetheless be somewhat more ambitious than simply trying to establish a set of "stylized facts." An attempt is made to describe regularities in terms that are sufficiently quantitative so that theories have a clear target and can thus sensibly make strongly falsifiable predictions. A key goal of such theories is, of course, to understand the necessary and sufficient conditions for regularities.

The approach for building theory described here is phenomenological. That is, it does not attempt to derive everything from a set of first principles but rather simply tries to connect diverse phenomena to each other to simplify our description of the world. Many economists will be uncomfortable with this approach because it often lacks "economic content,"—that is, the theories developed do not invoke utility maximization. In this sense this body of work lies somewhere between pure econometrics and what is usually called a theory in microeconomics. Even though the models infer properties of agent behavior and connect them to market properties such as prices, there is no attempt to derive the results from theories that maximize preferences. Instead we content ourselves with weaker assumptions, such as market efficiency. Given all the empirical problems surrounding the concept of utility, we view this as a strength rather than a weakness.

The work described here is still in an early stage and is very much in flux; many of these results are quite new, and indeed our own view is still changing as new results appear.

2.2. MARKET STRUCTURE

All of the work described here is based on results from studying stocks from the London, Paris, New York (NYSE and NASDAQ), and Spanish stock markets. These markets differ in their details, but they all do at least half of their trading (and in some cases all their trading) through a continuous double auction. "Auction" indicates that participants may place quotes (also called *orders*) stating the quantities and prices at which they are willing to trade; "continuous" indicates that they can update, cancel, or place new quotes at any time, and "double" indicates that the market is symmetric between buyers and sellers.[4]

[4]There are some small exceptions to symmetry between buying and selling, such as the uptick rule in the NYSE, but these are relatively small effects.

There are some important differences in the way these markets are organized. The NYSE was unusual (until the end of 2007) in that each stock has a designated specialist who maintains and clears the limit order book. The specialist can see the identity of all the quotes and can selectively show them to others. The specialist can also trade for his own account but has regulatory obligations to "maintain an orderly market." The London Stock Exchange, in contrast, has no specialists. It is completely transparent in the sense that all orders are visible to everyone, but it is completely anonymous in the sense that there is no information about the identity of the participants, and such information is not disclosed even to the counterparties of transactions. The Spanish Stock Market is unusual in that membership codes for quotes are publicly displayed. Thus these exchanges are generically similar but have their own peculiar characteristics.

Markets also differ in the details of the types of orders that can be placed. For example, the types of orders in the London Stock Exchange are called "limit orders," "market orders with limiting price," "fill-or-kill," and "execute and eliminate." To treat these different types simply and in a unified manner, we simply classify them based on whether an order results in an immediate transaction, in which case we call it an *effective market order*, or whether it leaves a limit order sitting in the book, in which case we call it an *effective limit order*. Marketable limit orders (also called *crossing limit orders*) are limit orders that cross the opposing best price and so result in at least a partial transaction. The portion of the order that results in an immediate transaction is counted as an effective market order, whereas the nontransacted part (if any) is counted as an effective limit order; thus in this case a single action by the participant gets counted as two separate orders. Note that we typically drop the term *effective* so that, for example, *market order* means *effective market order*. Similarly, a limit order can be removed from the book for many reasons; for example, because the agent changes her mind, because a time specified when the order was placed has been reached, or because of the institutionally mandated 30-day limit on order duration. We will lump all these together and simply refer to them as *cancellations*.

In addition to continuous double auctions, the London Stock Exchange has what is called the *off-book market* and the New York Stock Exchange has what is called the *upstairs market*. These are both bilateral exchanges in which members can interact in person or via telephone to arrange transactions. Such transactions are then reported publicly at a later time. With exceptions noted in the text, all the results obtained are from the continuous markets.

2.3. INFORMATION, LIQUIDITY, AND EFFICIENCY

The aim of this section is to motivate the empirical study of microstructure in a broader economic context—that of the information content of prices and the mechanisms that can lead to market efficiency. We discuss several fundamental questions concerning how markets operate. The discussion here sets the stage for the detailed quantitative investigations that we report in the following sections. Since one of our main subjects here is market impact, we review and critique the standard model for market impact, which is based on informed vs. uninformed trading.

2.3.1. Information and Fundamental Values

It is often argued that there is a fundamental value for stocks, correctly known to at least some informed traders who buy underpriced stocks and sell overpriced stocks. By doing so they make a profit and, through the very impact of their trades, drive back the price toward its fundamental value. This mechanism is the cornerstone of the theory of efficient markets and is often used to justify unpredictable prices. In such a framework, the fundamental value of a stock can only change with unanticipated news. The scenario is then the following: A piece of news becomes available, and market participants work out how this changes the fundamental price of the stock and trade accordingly. After a (supposedly fast) phase of *tâtonnement*, the price converges to its new equilibrium value, and the process repeats itself. To explain deviations from this picture, one can add a suitable fraction of uninformed trades to add some high-frequency noise.

Is this picture fundamentally correct to explain the reason that prices move and to account for the observed value of the volatility? Judging from the literature, it looks as if a majority of academics still believe that this story is at least a good starting point (but see, for example, Lyons, 2001). Recent empirical microstructure studies open the way to testing in detail the basic tenets and the overall plausibility of the standard equilibrium picture. We hope to convince the reader that the story is in fact significantly different. That is, we argue that an alternative way of looking at events provides superior explanatory power based on a simpler set of hypotheses. Before discussing at length the microstructural evidence for a change of paradigm, we would like at this stage to make several general comments that will be relevant—first on the very notion of fundamental value and information and second on various orders of magnitude and time scales involved in the problem.

Is the fundamental value of a stock or a currency a valid concept in the sense that it can be computed, at least as a matter of principle, with arbitrary accuracy with all information known at time t? The number of factors influencing the fundamental value of a company or of a currency is so large that there should be, at the very least, an irreducible intrinsic error. All predictive tools used by traders, based on economic ratios, earning forecasts, or the like, are based on statistical models detecting trends or mean reversion and are obviously noisy and sometimes even biased. For example, financial experts are known to be on the whole rather bad at forecasting the next earning of a company (see, e.g., Guedj and Bouchaud, 2005). News is often ambiguous and not easy to interpret. But if we accept the idea of an intrinsically noisy fundamental value with some band of width Δ within which the price can almost freely wander, the immediate question is, how large is the uncertainty Δ? Is it very small, say, 10^{-3} in relative terms, or quite a bit larger, say, 100%, as suggested by Black (1986)? If Black is right (which we tend to believe) and the uncertainty in the fundamental value is large, then the information contained in a trade is noisy and the amount of information contained in any given trade is necessarily small. Analysis of price impact makes it clear that the standard deviation of impacts is very large compared to their mean, suggesting that this is indeed the case.

2.3.2. Market Efficiency

Market efficiency is one of the central ideas in finance and appears in many guises. A standard definition of market efficiency (in the informational sense) is that the current price should be the best predictor of future prices; that is, that prices should be a martingale. Another closely related notion is arbitrage efficiency, which in its weakest form states that it should not be possible to make a profit without taking risks; in a stronger form it says that two strategies with the same risk should make the same profits, at least once their usefulness for inclusion in a portfolio is taken into account. Steve Ross, among others, has advocated that market efficiency (rather than equilibrium) should be the core postulate for financial theory (Ross, 2004).

We agree with this point of view, at least in so far as it does not imply believing in allocative efficiency; that is, that prices correctly reflect the underying value of the assets. Strictly speaking a market is allocatively efficient if it is Pareto optimal, in the sense that there is no alternative allocation of prices and holdings that makes someone better off without making someone worse off. This is related to whether or not prices are set at their "proper" values. It is entirely possible to imagine a market in which prices are unpredictable and yet in which there is no sense that prices are set correctly. That is, once we depart from neoclassical equilibrium, a market might be informationally efficient yet allocatively inefficient.

A closely related point is that there are two very different possible explanations for market efficiency:

1. The standard view in economics is that perfect efficiency reflects perfect information processing. Traders process each new bit of information as it arrives, and prices immediately go to their new equilibrium values.
2. An obvious alternative is the standard one that explains randomness in many other fields, such as fluid turbulence. Markets are too complicated to be predictable. Under this explanation prices move randomly because investor behavior is complicated, based on many hidden factors, so to an external observer it is "as if" individual investors are just flipping coins.

The correct explanation is likely to be a mixture of both effects. On one hand markets are inherently complicated, but on the other hand, whatever predictability is left over is substantially removed by arbitrageurs. Under this synthetic view, which we take here, one can simply associate an impact with trades, treat all investors as more or less the same, and adjust the expected impact as needed to preserve efficiency based on factors that derive from the predictability of trades.

Finally we want to emphasize that though we believe that market efficiency is a very useful concept and provides an excellent starting point for developing theories, it is inherently contradictory and is at best an approximation. Markets can only be informationally efficient at first order but must necessarily be inefficient at second order. This was originally pointed out by Milton Friedman, who noted that without informed traders to push prices in the right direction, there is no reason that markets should ever be efficient. If markets were truly efficient, informed traders should make the same profits as anyone else, and there would be no motivation for them to remain in the market. Thus markets cannot be fully efficient.

Even if for many purposes it can be a good approximation to assume that markets are efficient, there are other situations in which deviations from efficiency can be quite important. Understanding how markets evolve from inefficient to efficient states, predicting the necessary level of deviations from efficiency that must persist in steady state, and understanding their role in the way markets function remain areas of investigation that are still largely not understood. This is relevant for our discussions on incorporating information into prices because when we speak about information we must have traders to process that information and trade based on it. It is precisely the market impact of these traders that moves prices. Thus while on one hand market impact is a friction, it can also be viewed as the factor that maintains efficiency, and so it is essential that we properly understand it.

2.3.3. Trading and Information

Informational efficiency means that information must be properly incorporated into prices. Under assumptions of rationality, when all traders have the same information, prices should move more or less automatically, with very little trading (Milgrom and Stokey, 1982; Sebenius and Geanakoplos, 1983). But of course that's not true—people don't have the same information, and even if they did, real people are likely to take different views on what the information means. The empirical fact that there is so much trading supports this idea (Shiller, 1981). Grossman and Stiglitz (1980) developed an equilibrium model in which traders have different information that shows that in this situation, trading and price movements are informative (see also Grossman, 1989). If I know that you are rational, and I know that you have different information than I have, when I see you trade and the price rises I can infer the importance of your information and thus I should change my own valuation.

Intuitively the problem with this view is that even small deviations from rationality and perfect information can lead to incorrect prices and instabilities in the price process. Suppose, for example, that you and I both overestimate how much information the other has. Then when I see you trade I change my valuation too much. When I see you buy, I also buy, but I buy more than I should. To make this slightly more quantitative, let the initial price be p_0 and suppose that after Agent A observes new fundamental information the price rises by f, which might or might not be the correct fundamental level. After Agent A trades, the new price becomes $p_1 = p_0 + f$. Agent B sees the price rise by f, and assuming that Agent A has more information than he really does, he buys and causes the price to rise to $p_2 = p_0 + af$. Then B sees the price rise more than f, so he buys, driving it to $p_3 = p_0 + a^2 f$, and so on. This process is clearly unstable if $a > 1$. The agents either need to know the value of a exactly or they need to be able to adapt a based on information that is not contained in the price. It is difficult to understand how they can do this since by definition if they are not rational, not only do they not have full information, they do not know how much information they have, and they thus cannot know *a priori* the proper value of *a*. Under deviations from rationality, deviations from fundamentals are inevitable. For a beautiful model where copycats lead to such instabilities, see P. Curty and M. Marsili (2006).

In its extreme version, this is just the kind of scenario that occurs during a bubble (see Bouchaud and Cont, 1998, for an explicit model). Any reasonable investor who lived through the millennium technology bubble experienced this problem. Even though high prices seemed difficult to rationalize based on values, prices kept going up. This led many sanguine investors to lose confidence in their own valuations and to hang onto their shares much longer than they thought was reasonable. If they didn't do this they experienced losses as measured relative to their peers. Under this view, bubbles stem from the problem of not knowing how much information price movements really contain and the feedback effects that occur when most people think they contain more information than they really do. This point of view differs from that in the standard literature on rational bubbles. As we argue here, though not entirely different, there are important contrasts between this view and the standard rational expectations/noise trader models.

2.3.4. Different Explanations for Market Impact

Why is there market impact? We will distinguish three possibilities:

1. *Trades convey a signal about private information.* This idea, discussed in the previous section, was developed by Grossman and Stiglitz (1980). The arrival of new private information causes trades, which cause other agents to update their valuations, which changes prices. In this case it is fair to say that trades cause price changes, since even if there happens to be no information, unless this is common knowledge the observation of a trade is still interpreted as information, which causes the price to change.
2. *Agents successfully forecast short-term price movements and trade accordingly.* This can result in measurable market impact even if these agents have absolutely no effect on prices at all. If an agent correctly forecasts price movements and trades based on this forecast, when this agent buys there will be a tendency for the price to subsequently rise. In this case causality runs backward, that is, because the price is about to rise, agents are more likely to trade in anticipation of it, but a trade based on no information will have no effect.
3. *Random fluctuations in supply and demand.* Even in the standard market-clearing framework, if a given agent increases her demand while other agents keep theirs constant, when the market clears that agent buys and the price rises. Fluctuations in supply and demand can be completely random, unrelated to information, and the net effect regarding market impact is the same. In this sense impact is a completely mechanical—or better, statistical—phenomenon. As we will see in Appendix 2.1, the meaning of this can be subtle and may depend on the market framework.

All three of these possibilities result in identical short-term market impact—that is, a positive correlation of trading volume and price movement—but they are conceptually very different. If some traders really know the "true" price at some time in the future (say, the end of the day, after the market closes), the observation of an excess of buy trades allows the market to guess that the price will move up and to change the quotes accordingly (see Section 2.7.2 on the Glosten-Milgrom model). In this sense, information has progressively included in prices as a function of the observed order flow. In this picture, as emphasized in Hasbrouck (2007), "orders do not *impact* prices. It is more accurate to say that orders *forecast* prices." But if the mechanical interpretation is correct, correlation between price changes and order flow is a tautology. If prices move only because of trades, "information revelation" may merely be a self-fulfilling prophecy that would occur even if the fraction of informed traders is zero. The only possible differences between these pictures come about in the temporal behavior of impact, which we discuss in Section 2.6.

2.3.5. Noise Trader Models and Informed vs. Uninformed Trading

In behavioral finance, the problem of irrational investors is typically coped with by introducing "noise trader" models, in which some agents (the noise traders) are stupid while others are completely rational (Kyle, 1985; DeLong et al., 1990; Shleifer, 2000). Noise trading could be driven by the need for liquidity (here meaning the need to raise capital for other reasons), it could be driven by the desire to reduce risk, or it could be "irrational behavior," such as trend following. The assumption is made that such investors lack the skill or information-processing ability to collect and/or make full use of information. The rational investors, in contrast, are assumed to correspond to skilled professionals. Their trading is perfect in the sense that they know everything. Examples of what they must know include the strategies of all the noise traders and the fraction of capital traded by noise traders as opposed to rational investors. In such models, prices can deviate from fundamental values due to the action of the noise traders and the desire of the rational agents to exploit them as much as possible, but the rational agents always keep them from deviating too much. The rational traders make "informed" trades while the noise traders make "uninformed" trades.

There are several conceptual problems with noise trader models that are clear *a priori*. No one can seriously dispute that traders must have different levels of skill, but is the noise trader approach the right way to model this? Though it might be fine to model a continuum of skill levels as "low" and "high," the idea of identifying the "high" level with perfect rationality postulates a level of skill at the top end that is difficult to imagine. The panoply of strategies used by real traders is large, and financial professionals (and even private investors) are sufficiently secretive about what they do that it is difficult to imagine that even the most skilled traders could fully understand everyone else's behavior.

Another problematic issue is the operational problem of measuring information. For example, under the theory that urgency is a proxy for informativeness, empirical work on the subject has often defined an informed trade as one that is executed by a market order and defined an uninformed trade as one that is represented by a limit order. This goes against the fact that many of the most successful hedge funds make extensive use of limit orders.[5] The only alternative is to use data that contains information about the identity of the agents making the trades. Such data does indeed confirm that professionals perform better than amateurs (Barber et al., 2004), but as mentioned, there is no demonstration that this means they are rational, and other than stating that professionals make larger profits it is impossible to determine whether or not professionals are good enough to be considered rational. (On this point, see also Odean, 1999).

2.3.6. A Critique of the Noise Trader Explanation of Market Impact

One of the most important questions to ask about any theory is what it explains that is not explained by a simpler alternative. Noise trader models have been proposed to explain why market impact is a concave function of trading volume. The empirical evidence for this concept is discussed in detail in Sections 2.5 and 2.6; in any case, it is a well-established empirical fact that the market impact as a function of trade size has a decreasing derivative. This can be alternatively stated as saying that the price impact per share decreases with the total size of the trade. The standard explanation for this is that it is due to a mixture of informed and uninformed trading. If more informed traders use small trade sizes and less informed traders use large trade sizes, small trades will cause larger price movement per share than large trades.

There are several problems with this theory:

- A concave market impact function is observed in all markets that have been studied, including many such as the London and Paris markets, where the identities of orders are kept completely anonymous. This rules out any explanation that depends on trades made by some agents communicating more information about prices than others and leaves only the possibility that some traders are able to anticipate short-term price movements better than others; see the discussion in Section 2.3.4.
- The model is unparsimonious in the sense that it requires the specification of a function that states the information that traders have as a function of the size of the trades that they use.
- The model is difficult to test because it requires finding a way to specify the information that various groups of traders have *a priori*. One proposal is to do this based of the average profits of different groups of traders. This proposal suffers from the problem that the time horizon for market impact is typically very different than the time horizon on which traders attempt to make profits. A fund manager who intends to a buy a stock and hold it for three years may make the trade to take up that position in a single day. Though this manager might have great skill in predicting stock price movements on a three-year time horizon, she may have no skill at

[5]We are basing this on personal conversations with market practitioners and so can only place a lower bound: We know many people working in many sophisticated trading operations and all of them at least partially use limit orders. We suspect the correct statement is that "most" or even "nearly all" successful hedge funds use limit orders for at least a substantial part of their trading.

all on a daily horizon. Thus in a large fraction of cases, even under large variations in trader skill, impact may have little correlation with profits.

- If it is indeed advantageous to use small trades, then since this is a trivial strategy, one would think that everyone would quickly adopt it and the effect would disappear. In fact, in the past five years or so there has been a huge increase in algorithmic trading, in which brokers automatically execute large trades for clients by cutting up the trades into small pieces. One would therefore think that in modern times the concavity should have diminished or even eliminated entirely. There is little evidence for this; the impact continues to be highly concave.

Thus we have argued in the preceding that the theory is implausible, but even more important, that it makes weak and untestable predictions. The prediction of concavity requires a set of assumptions that are complicated to specify and impossible to measure. The predictions are purely qualitative, and it is not obvious how they might be extended to other properties of impact, such as temporal behavior.

2.3.7. The Liquidity Paradox: Prices Are Not in Equilibrium

We will argue here that liquidity is an important intermediary that modulates the effect of information. We are defining liquidity in terms of the size of the price response to a trade of a given size. High liquidity implies a small price response. Since trades carry information, if the size of trades in response to a given level of information remains constant, as the liquidity varies the price response to information varies with it.

Under the assumption that trading is an intermediate step in the response of prices to information, one can conceptually decompose it into two terms:

$$\Delta p = \mathcal{T}(I)/\lambda \qquad \textbf{(2.1)}$$

where λ is the liquidity and $\mathcal{T}(I)$ is the response of trades to information I. Variations in the liquidity do not tell the full story about the response of prices to information; to do that one would also need to understand $\mathcal{T}(I)$. Nonetheless, as we argue here, the effects of varying liquidity are substantial, and they have the huge advantage of being easily measurable.[6] In contrast, since information is difficult to measure, $\mathcal{T}(I)$ is difficult to measure. Furthermore, the preceding equation should be interpreted rather loosely; as we shall see, impact is in fact neither linear nor permanent.

A very important empirical fact that is crucial to understanding how markets operate is that even "highly liquid" markets are in fact not that liquid. Take, for example, a U.S. large-cap stock. Trading is extremely frequent: a few thousand trades per day, adding up to a daily volume of roughly 0.1 to 1% of total market capitalization. Trading is even more frantic on futures and Forex markets. However, the volume of buy or sell limit orders typically available in the order book at a given instant in time is quite small: only the order of 1% of the traded daily volume—that is, $10^{-4} - 10^{-5}$ of the market cap for

[6]Of course, liquidity may also depend on information, and indeed in Section 2.6 we will develop this connection.

stocks. Of course, this number has an intraday pattern and fluctuates in time, and it can reach much smaller values in liquidity crises.

The fact that the outstanding liquidity is so small has an immediate consequence: Trades must be fragmented. The theoretical motivations for this were originally discussed by Kyle (1985). It is not uncommon that investment funds want to buy large fractions of a company, often exceeding several percent. One possibility is to arrange upstairs block trades, but this lacks transparency and can be costly. If trading occurs through the continuous double auction market, these numbers suggest that to buy 1% of a company requires at least the order of 100–1000 individual trades. This is under the unrealistic assumption that each individual trade completely empties the order book; more realistically, each trade consumes only a fraction of the order book, and the number of trades is even larger. But because 1000 trades corresponds to roughly the whole daily liquidity, it is clear that these trades have to be diluted over several days, since otherwise the market would be completely destabilized. Thus an informed trader cannot use her information immediately and has to trade into the market little by little.

But why is liquidity, as measured by the number of standing limit orders, so low? Both for similar and for opposite reasons. Too large a buy limit order from an "informed" trader would give her away and raise the price of the sellers. Too large a limit order from a liquidity provider would put him at risk of being "picked-off" by an informed trader. There is a kind of hide-and-seek liquidity game taking place in organized markets, where buyers and sellers face a paradoxical situation: Both want to have their trading done as quickly as possible, but both try not to show their hands and reveal their intentions. As a result, markets operate in a regime of vanishing *revealed liquidity* but large *latent liquidity*; this leads to a series of empirical regularities that we will present here.

From a conceptual point of view, however, the most important conclusion of this qualitative discussion is that prices are typically not in equilibrium in the traditional Marshall sense. That is, the true price is very different than it would be if it were set so that supply and demand were equal as measured by the honest intent of the participants, as opposed to what they actually expose. As emphasized previously, the volume of individual trades is much smaller than the total demand or supply at the origin of the trades. This means that there is no reason to believe that instantaneous prices are equilibrium, efficient prices that reflect all known information. Much of the information is necessarily latent, withheld due to the small liquidity of the market, and only slowly revealing itself (see Lyons, 2001, for similar ideas). At best, the notion of equilibrium prices can only make sense over a long time scale; high-frequency prices are necessarily soiled by a significant amount of noise.

2.3.8. Time Scales and Market Ecology

Consider again the case of a typical U.S. large-cap stock—say, Apple, which (as of November 2007) had a daily turnover of around $8 billion. There are on average six transactions per second and on the order of 100 events per second affecting the order book. These are extremely small time scales compared to the typical time for public news events, in which a hot stock like Apple might be mentioned by name every few hours during a period of fast information arrival. Perhaps surprisingly, the number of large jumps in price is much higher. For example, if we define a jump as a one-minute return exceeding three standard deviations, there are on the order of ten such jumps per day, reflecting the very heavy-tailed distribution of high-frequency returns (Joulin et al., 2008). More often than not such jumps occur in the absence of any identified news. It is obviously a particularly important question to understand the origin and the mechanisms leading to these jumps. The difference between the frequency of news and the frequency of jumps already suggests that something else must be at work, such as fluctuations in liquidity, that may have little or nothing to do with external news entering the market.

What is the typical time scale of the round-trip trades of investors? This depends very much on the style of trading; traditional long-only funds have investment horizons on the scale of years, whereas more aggressive long-short statarbs have time scales of weeks or days, sometimes even shorter. Some empirical results support the existence of a broad spectrum of investment horizons (see Sections 2.4.3 and 2.11.1). The optimal frequency of a trading strategy is a trade-off between the expected profit and the friction and transaction costs. Since the fraction of costs grows with the trading volume, large investment funds cannot trade too quickly. This, again, is directly related to the small prevailing liquidity. So it is reasonable to think that information-based trading decisions have intrinsic frequencies ranging from a few days to years. As we have already emphasized, for large investors a single decision may generate many more trades: A decision to buy or sell may persist for days to months, generating a series of small trades. Again, the important message is that low-frequency, large-volume investment decisions imply high-frequency, small-volume trades and that high-frequency prices cannot be equilibrium prices.

There is, however, a potentially viable high-frequency strategy called *market making* that consists of providing instantaneous liquidity to buyers and sellers and trying to eke out a profit from the bid–ask spread. As originally shown by Glosten and Milgrom (1985), the difficulty is to avoid losses due to adverse price moves. Since market makers are offering either to buy or to sell, they are giving a free option to others who might have better information. The profitability of market-making strategies depends both on the spread, which is beneficial, and on the long-term impact of trades, which is detrimental. This intuition will be made more precise and discussed in detail in Section 2.7. On some exchanges market making is institutionalized, with certain obligations and advantages bestowed to those who take the burden of providing liquidity. However, markets have become more and more electronic, with an open order book allowing each investor to behave either as a liquidity provider by posting limit orders or as a liquidity taker by issuing market orders. Depending on market conditions (for example, the instantaneous value of the spread), investors can choose either type of order. There is both empirical and anecdotal evidence that some players implement high-frequency, market-making strategies. This contribution to order flow is often described as "uninformed." Although this flow differs from longer horizon trades, which are supposed to be economically informed, these market-making strategies routinely use sophisticated short-term prediction tools and exploit any profitable high-frequency signals. The preceding simplified separation of market participants into two broad classes—speculators/liquidity hunters that trade at medium to low frequencies and market makers/liquidity providers at high frequencies—is both realistic and useful to understand the *ecology* of financial markets (Handa and Schwartz, 1996; Farmer, 2002; Wyart et al., 2008; Lillo et al., 2008b). The competition between these two categories of traders allows one to make sense of a number of empirical facts, we believe much more usefully than noise trader models. In Section 2.11 we present some recent empirical results on the characterization of a market ecology.

2.3.9. The Volatility Puzzle

Given that markets are ecological systems in which participants have a broad distribution of time horizons from seconds to years, it is perhaps not surprising to see long-memory effects in financial markets—for example, in trading volume, volatility, and order flow. What is *a priori* surprising, however, is that despite the fact that high-frequency prices cannot possibly be in equilibrium because of lack of liquidity, and despite the fact that it should take time for the market to interpret a piece of news and agree on a new price, the average volatility is remarkably constant on a wide range of different time scales. As measured by autocorrelation, prices are remarkably efficient down to the fastest time scales. We have argued that news arrival happens on much longer time scales. Given that this is true, how can prices remain so efficient, at least with respect to linear models, even on very fast time scales?

One possible explanation might be that public information as evidenced on news feeds is only a small part of the available information. Instead, suppose there are many sources of private information, which agents are continually processing. As they make their decisions, they trade. Given that heavily traded stocks average many trades per second, this would suggest that a truly staggering amount of information is being processed. We find this explanation implausible.

The alternative is that there is an information-processing cascade from fundamental information on slow time scales to technical information on fast time scales. As we have argued, fundamental information enters at a relatively slow rate and then is processed and incorporated into prices. Under this view, high-frequency strategies play an important role. Such strategies do not directly process external information but rather serve the role of digesting that information and keeping the price stream unpredictable. Such strategies are not processing fundamental information but rather are acting as technical trading strategies, processing information contained in the time history of prices, trading volume, and other information that is completely internal to the market. The ability to substitute information in a time history for state information is well supported in dynamical systems theory (Packard et al., 1980; Takens, 1981; Casdagli et al., 1991). Thus we argue that in the ecology of financial markets, high-frequency strategies are fed by lower-frequency strategies through an information cascade from longer to shorter time scales and from fundamental to technical information, finally resulting in white noise on all scales.

This also suggests that microstructural effects may influence the value of the volatility, as suggested by Lyons (2001); "microstructure implications may be long-lived" and "are relevant to macroeconomics." We will comment on the relation between microstructure and volatility in Section 2.8. This relation is also relevant for the regulator who might attempt to alter the microstructural organisation of markets to reduce the volatility.

2.3.10. The Kyle Model

A classic noise trader model for market impact, which is a natural point of comparison, is due to Kyle (1985). This model assumes that there are three types of traders: noise traders who make random trades, market makers who set prices to guarantee efficiency, and an insider who has access to superior information. Under the most general version of the model the noise traders and insider trade continuously from a starting time until a final liquidation time, at which point everyone is paid the liquidation price for their holdings. The insider has superior information about the final liquidation price p_∞ and an infinite bank, which she uses to maximize profits at the expense of the noise traders.

The optimal amount that the investor should trade is easily found to be proportional to the difference $p_\infty - p_t$, where p_t is the current price. With the assumption of a linear and permanent impact, in Kyle's notation the price evolution is given by:

$$p_{t+1} - p_t = \lambda\,[\Phi_t + \xi_t] + \eta_t \quad \Phi_t = \beta[p_\infty - p_t] \tag{2.2}$$

where Φ_t is the signed demand of the investor, λ, β are coefficients, ξ_t is the noise trader demand coming from all other market participants, and η_t is a noise term accounting for possible changes of prices not induced by trading (news, etc.). This equation can easily be solved and leads to an exponential relaxation of the initial price toward p_∞ plus a bounded noise term.

The impact in this model can be regarded as essentially mechanical. There is an apparently permanent change in price that is linearly proportional to the total amount that the noise traders and insider trade. We say "apparently permanent" because, since there is a final liquidation time, what happens past this point is undefined. Note that in this model the price will move toward p_∞ regardless of whether it is the correct price; all that is necessary is that insiders *believe* it is the correct price. A random assignment of beliefs about p_∞ will result in a corresponding random set of impacts. Thus, referring to our discussion of the various explanations for market impact in Section 2.3.4, while the Kyle model is built in the spirit of explanation (1), that trades convey a signal about private information, it is equally consistent with (3), random fluctuations in supply and demand.

The assumption of a final liquidation price can naïvely lead to erroneous conclusions. For example, this model suggests that one can easily manipulate the price. However, in the absence of a liquidation price where a transaction with a counterparty can be realized without impact, things are not so trivial: As soon as the investor wants to close his position, he will again mechanically revert the price back to its initial value and take losses. (To see this, note that in a single round trip the investor will buy at a high price and sell at the original price.) The preceding impact model, Eq. (2.2), although used very often in agent-based models of price fluctuations (two of us have also developed similar ideas, i.e., Bouchaud and Cont, 1998; Farmer, 2002), is far too naïve to represent the way real markets operate, at least at the tick-by-tick level.

Thus we see that while the Kyle model provides a good starting point for understanding why there should be market impact and why it is useful to trade into a position incrementally, it falls short of making realistic predictions about impact. We feel that the key elements that need to be extended are (1) removing the final liquidation price, (2) eliminating the infinite bank of the insider and replacing it with the more realistic assumption of a finite, predetermined trading size, and (3) eliminating the distinction between the insider and the noise trader. The aim of the following sections is to explain in detail how to construct a model generalizing Eq. (2.2), using an approach based on robust facts observed in empirical data and consistency arguments. We will find that impact is in general *nonlinear* and *transient*—or equivalently, as explained in Section 2.6.5, *history dependent*. It is only after a properly defined "coarse-graining" procedure that such an impact model can possibly make sense.

2.4. LARGE FLUCTUATIONS AND LONG MEMORY OF ORDER FLOW

From a mechanical point of view, price formation process is the outcome of (i) the flow of orders arriving in the market and (ii) the response of prices to individual orders. Since price dynamics are reasonably well described by a Brownian motion, one might naïvely assume that this would be true for order flow as well. In fact, this is far from the truth. As we explain in detail in this section, order flow is a highly autocorrelated long-memory process. As a consequence, to maintain market efficiency the price response to orders must strongly depend on the past history of order flow. This has profound conseqences on the way in which markets incorporate information.

2.4.1. Empirical Evidence for Long Memory of Order Flow

We discuss here the statistical properties of order flow by considering the time series of signs of orders. Specifically, consider the symbolic time series obtained in event time by replacing buy orders with $+1$ and sell orders with -1, irrespective of the volume of the order. We reduce these series to ± 1 rather than directly analyzing the signed series of order sizes to avoid problems created by the large fluctuations in order size.[7] This reduction can be done for market orders, limit orders, or cancellations, all of which show very

[7]Fluctuations in order size are heavy tailed and have long memory themselves, so statistical averages based on them converge only slowly. The essential behavior is captured by the series of signs.

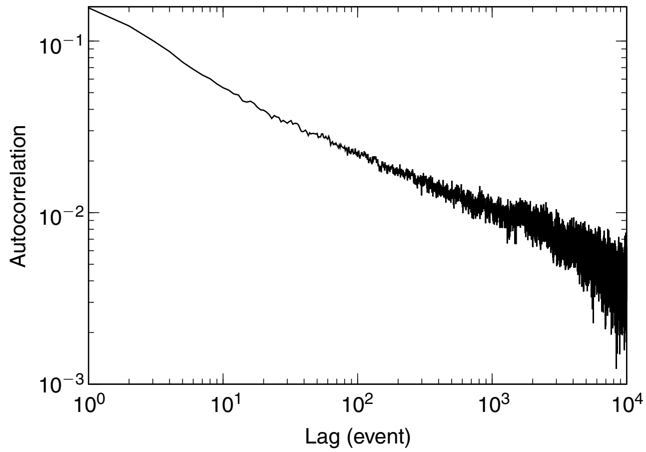

FIGURE 2.1 Autocorrelation function of the time series of signs of orders that resulted in immediate trades (effective market orders) for the stock Vodafone traded on the London Stock Exchange from May 2000 to December 2002, a total of 5.8×10^5 events.

similar behavior.[8] We denote with ϵ_i the sign of the i^{th} market order. Figure 2.1 shows the sample autocorrelation function of the market order sign time series for Vodafone (VOD) in the period 1999–2002 in double logarithmic scale. The figure shows that the autocorrelation function for market order signs decays very slowly. The autocorrelation function is still above the statistical noise level even after 10^4 transactions, which for this stock corresponds to roughly 10 days. This result indicates that if one observes a buy market order now, based on this information alone there is some nonvanishing predictability of the market order signs two weeks from now.

We also note that the autocorrelation function shown in Figure 2.1 is roughly linear in a double logarithmic scale over more than four decades.[9] This suggests that a power-law relation $C_\tau \sim \tau^{-\gamma}$ might be a reasonable description for the sample autocorrelation function.[10]

Stochastic processes for which the autocorrelation function decays asymptotically as a power law with an exponent smaller than one are called *long-memory processes* (Beran, 1994). A precise definition of long memory processes can be given in terms of the autocovariance function Γ_τ. We define a process as long memory if in the limit $\tau \to \infty$

$$\Gamma(\tau) \sim \tau^{-\gamma} L(\tau), \tag{2.3}$$

[8]Long memory is also observed if the side of all activity (bid or ask), including both limit and market orders, is taken together. In contrast, if one assigns to limit orders to sell and cancellations of buy orders a negative sign, corresponding to the fact that the only nonzero price movements it can produce are downward, the combined sequence of signs for market orders, limit orders, and cancellations does not show long memory.
[9]The noisy behavior for large τ comes from the fact that for large lags the statistical errors are remaining roughly constant while the signal decreases, so the relative size of the fluctuations becomes larger.
[10]$f(y) \sim g(y)$ means that there exists a constant $K \neq 0$ such that $\lim_{y\to\infty} f(y)/g(y) = K$.

where $0 < \gamma < 1$ and $L(\tau)$ is a slowly varying function[11] at infinity. The degree of long memory is given by the exponent γ; the smaller γ, the longer the memory. The integral of the autocovariance (or autocorrelation) function of a long-memory process diverges. Long memory can also be discussed in terms of the Hurst exponent H, which is simply related to γ. For a long-memory process, $H = 1 - \gamma/2$ or $\gamma = 2 - 2H$. Short-memory processes have $H = 1/2$, and the autocorrelation function decays faster than $1/\tau$. A positively correlated long-memory process is characterized by a Hurst exponent in the interval $(0.5, 1)$. The use of the Hurst exponent is motivated by the relationship to diffusion properties of the integrated process. For normal diffusion, where by definition the increments do not display long memory, the standard deviation asymptotically increases as $t^{1/2}$, whereas for diffusion processes with long-memory increments, the standard deviation asymptotically increases as $t^H L(t)$, with $1/2 < H < 1$, and $L(t)$ a slow-varying function. In econometrics of financial time series, many variables have the long-memory property. For example, it is widely accepted that the volatility of prices (Ding et al., 1993) and stock market trading volume (Lobato and Velasco, 2000) are long-memory processes. Models of long-memory processes include fractional Brownian noise (Mandelbrot and van Ness, 1968) and the ARFIMA process introduced by Granger and Joyeux (1980) and Hosking (1981).

As Figure 2.1 suggests and as discovered by Bouchaud et al. (2004) and Lillo and Farmer (2004), order flow is also described by a long-memory process. The long-memory of order flow is very robust and is consistently observed for every stock that has so far been examined. Lillo and Farmer tested for long memory in a panel of 20 highly capitalized stocks traded at the London Stock Exchange using Lo's modified R/S test (Lo, 1991), which is known to be a strict test for long memory. They found that even on short samples, in most cases the hypothesis of long memory could not be rejected. The value of H observed in the London Stock Exchange was generally about $H \approx 0.7$, which corresponds to $\gamma = 0.6$. Bouchaud et al. (2004) measured a larger interval of γ values in the Paris Stock Exchange, ranging from 0.2 to 0.7. Long memory has also measured an assortment of stocks in the NYSE; these results are mentioned in Lillo and Farmer (2004) but have not been published in detail.

2.4.2. On the Origin of Long Memory of Order Flow

What causes long memory in order flow? The presence of persistent time correlations in the order flow suggests two possible classes of explanations. The first type of explanation is that this is a property of the order flow of each investor, independent of the behavior of other investors, as proposed by Lillo et al. (2005). The second type of explanation is that investors herd in their trading though an imitation process that involves

[11] $L(x)$ is a slowly varying function (see Embrechts et al., 1997) if $\lim_{x\to\infty} L(tx)/L(x) = 1\ \forall t$. In the preceding definition, and for the purposes of this chapter, we are considering only positively correlated long-memory processes. Negatively correlated long-memory processes also exist, but the long-memory processes we consider in the rest of the chapter are all positively correlated.

an interaction between them, as proposed by LeBaron and Yamamoto (2007). It is of course possible that both effects operate at once, but in any case one would like to know their relative magnitude.

We believe that the evidence gathered so far strongly favors the first explanation. More explicitly, we believe that the dominant cause is the strategic behavior of large investors who split their orders into many small pieces and execute them incrementally. The evidence for this stance comes from two sources. One is the agreement of the properties of the order flow with theory, and the other is additional evidence based on data that gives information about the identity of participants. We summarize both of these here.

2.4.3. Theory for Long Memory in Order Flow Based on Strategic Order Splitting

Lillo et al. (2005) have hypothesized that the cause of the long memory of order flow is a delay in market clearing. To make this clearer, imagine that a large investor decides to buy ten million shares of a company. It is unrealistic for her to simply state her demand to the world and let the market do its job. It is unlikely that enough sellers will be present, and even if there were, revealing the intention to buy a large quantity of shares will very likely push the price up substantially. Instead the large investor keeps her intentions as secret as possible and trades incrementally over an extended period of time, possibly through intermediaries. As already discussed, the strategic reasons for doing this were made clear by Kyle (1985), who investigated a model in which an insider with information about a final liquidation price tries to maximize profits. In simple terms, the motivation is that by splitting the hidden order into small pieces, the investor is able to execute much of the hidden order at prices that do not reflect the full price movement that it will eventually cause.

Our perspective differs from Kyle's in that we assume that the size of the order, which we call the *hidden order*, is given at the outset when the initial trading decision is made. We believe that the size of such orders is largely determined by the fund manager *a priori* and is influenced by a combination of the funds under management and the time scale of the strategy, which is typically much longer than the time scale for completing the trade. The other notable differences are that we do not assume a final liquidation price and we do not make a distinction between informed traders and noise traders. When taken together these differing assumptions create key differences in the predictions of the model in comparison with Kyle.

In several studies based on data giving the identity of hidden orders, about a third of the dollar value of such institutional trades took more than a week to complete (Chan and Lakonishok, 1993, 1995; Vaglica et al., 2008). This conflicts with the standard model of market clearing presented in textbooks, which assumes that agents fully state their supply and demand and that prices are set so that supply and demand are evenly matched. The fact that large orders are kept secret and executed incrementally implies that at any given time there may be a substantial imbalance of buyers and sellers. Effective market clearing is delayed by variable amounts that depend on fluctuations in the size and signs of the unrevealed hidden orders.

We now describe a recently proposed simple model of order flow that postulates the independence of trading activity of investors and which is able to reproduce the long-memory properties of order flow (Lillo et al., 2005). In the simplest version of the model, assume that at any time there are K hidden orders present in the market. Initially the size V of these hidden orders is drawn from a distribution $P(V)$ and the sign ϵ_i is randomly chosen. For simplicity we assume that V is an integer number. We indicate with $V_i(t)$ the volume of hidden order i that has not yet been traded at time t. At each time step t an existing hidden order i is chosen at random with uniform probability, and a unit volume of that order is traded, so that $V_i(t+1) = V_i(t) - 1$. This generates a revealed order of unit volume and sign ϵ_i. A hidden order i is removed if $V_i(t+1) = 0$; that is, when the hidden order is completely traded. When this happens a new hidden order is created with a random sign and a new size.

It is possible to find a closed expression for the autocorrelation function of the trade sign C_τ as a function of the hidden order size distribution $P(V)$. The asymptotic behavior of C_τ can be obtained through a saddle point approximation. If the hidden order size is asymptotically Pareto distributed, that is,

$$P(V) \sim \frac{\alpha}{V^{\alpha+1}} \qquad \textbf{(2.4)}$$

the autocorrelation function of order sign behaves asymptotically as (Lillo et al., 2005)

$$C_\tau \sim \frac{K^{\alpha-2}}{\alpha} \frac{1}{\tau^{\alpha-1}} \qquad \textbf{(2.5)}$$

Thus the model makes the falsifiable prediction that the exponent γ of the power-law asymptotic behavior of the autocorrelation of order sign is determined by the exponent α of the power-law asymptotic behavior of the hidden order size distribution through

$$\alpha = \gamma + 1 \qquad \textbf{(2.6)}$$

Since we observe that $\gamma \simeq 0.5$, this model predicts that $\alpha = 1.5$.

It is worth noting that Lillo et al. (2005) also introduced a more general model in which the number of hidden orders is not constant in time. Specifically, at each time t a new hidden order is generated with probability $0 < \lambda < 1$ if $K(t) > 0$, or probability 1 if $K(t) = 0$. Although this model is not solved analytically, numerical simulation shows that the relation between the exponent of the autocorrelation of order sign and the exponent of order size distribution is the same as in the simpler model where the number of hidden orders is fixed.

This model for the origin of correlation in order flow is in principle empirically testable. The main difficulty arises from the lack of large and comprehensive databases of the hidden orders of investors. There are two ways to check the consistency of the theory. The first one is to compare the distribution of trade sizes in block markets to the autocorrelation function of order signs in order book markets. In block markets, trades are made bilaterally and the identity of counterparties is known. Brokers do not like order splitting and strongly discourage it. Thus block markets can be considered a crude proxy for observing the distributional properties of hidden orders.[12] In the next section we discuss evidence that suggests that block trade volume is indeed asymptotically power-law distributed with an exponent $\alpha \simeq 1.5$. For comparison[13] the average measured values of γ for LSE stocks is $\gamma = 0.57$, close to $\hat{\gamma} = 0.59$ as predicted by $\hat{\gamma} = \alpha - 1$. The second supporting evidence comes from a study of the Spanish Stock Exchange by Vaglica et al. (2008), who have inferred hidden orders using data with membership codes. This study is discussed in Section 2.11.1.

2.4.4. Evidence Based on Exchange Membership Codes

Empirical testing is difficult due to the fact that it is not easy to collect data on the behavior of individual investors. Nonetheless, partial information about the identity of participants can be obtained using data that identifies the broker or the member of the exchange who executes the trade, which we will simply call the *membership code*. In many stock markets, such as the LSE, the Spanish Stock Exchange, the Australian Stock Exchange, and the NYSE, it is possible to obtain data containing this information. It is important to stress that knowing the membership code is not the same as knowing the individual participant, since the member may either trade on its own account or act as a broker for other trades, or do both at once. Nonetheless, several recent papers have demonstrated that it is possible to extract useful information about the identity of individual traders using such information; for example showing that there are consistent behaviors that are persistent in time associated with particular membership codes, that such behaviors can be organized into a taxonomic tree, and that it is possible to detect the presence of large institutional trades (Lillo et al., 2008b; Zovko and Farmer, 2007; Vaglica et al., 2008).

Gerig et al. have used membership codes of the London Stock Exchange to test the hypothesis of the theory presented in Lillo et al. (2005). The autocorrelation function of market order signs is computed by considering realized orders placed by the same membership code or by different membership codes separately. Figure 2.2 shows the autocorrelation function of market order signs with the same membership code, different membership codes, and all transactions irrespective of membership code. The circles in the figure represent the autocorrelation function irrespective of the membership code and, as anticipated, it is well fitted by a power law. When only transactions with the same membership code are considered (the triangles), the autocorrelation is still power law with a slightly smaller exponent. Moreover, for a fixed lag τ, the autocorrelation function with the same membership code is one order of magnitude larger than the autocorrelation function irrespective of the membership code. Finally, when only transactions with different membership codes are considered, the autocorrelation function decays very rapidly to zero, and it is clearly not consistent with power-law

[12]The exception is that it is possible to split an order and trade with multiple brokers.

[13]The error bars in computing both γ and α are substantial, as can be seen by computing them for subsamples of the data, and the close agreement observed by Lillo et al. (2005) between γ and $\alpha - 1$ is probably fortuitious. Unfortunately, there still are not good statistical methods for assigning confidence intervals for exponents of power laws, particularly when the observations have long memory, but the errors can be roughly assessed by examining subsamples.

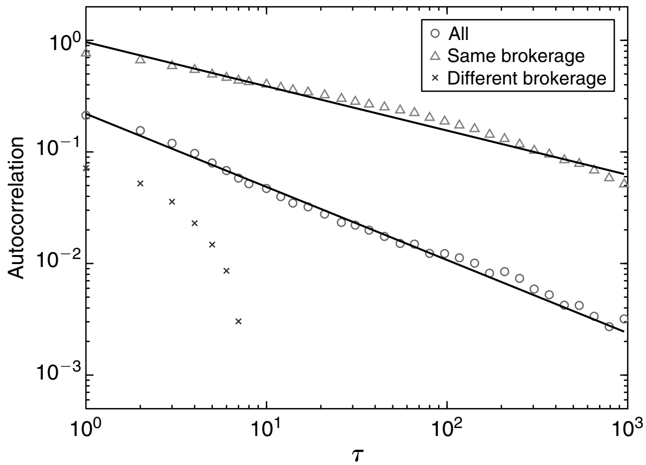

FIGURE 2.2 Autocorrelation of signs-versus-transaction lag for transactions with same membership code, different membership code, and all transactions irrespective of membership code, plotted on double logarithmic scale. The investigated stock is AstraZeneca (AZN) traded the LSE from 2000 to 2002.

behavior. Under the assumption that most investors use only a few brokers to execute a given hidden order, this plot strongly supports the hypothesis that the long memory of signs is due to the presence of investors that place many revealed orders of the same sign and that there is no clear sign of herding behavior among different investors. It is in principle possible that herding happens between investors using the same broker but not between investors with different brokers; however, the reasons why this would occur are unclear, and it seems implausible that it could explain such a dramatic difference.

2.4.5. Evidence for Heavy Tails in Volume

The theory we have developed makes it clear that the distribution of trading volume plays a key role in shaping many properties of the market, including the long memory of order flow, which, as we will show, in turn has important consequences for market impact. In recent years there has been a debate about the statistical properties of trading volume. This is partly due to the fact that markets have different structures and one should be careful in specifying which volume is considered in the analysis. Gopikrishnan et al. (2000) originally observed that volume of trades at the NYSE are asymptotically power-law distributed. Specifically, they claimed that for large volumes the probability distribution scales as

$$P(V > x) \sim x^{-3/2} \qquad \textbf{(2.7)}$$

This law has been termed the “half cubic” law. The NYSE, like many other financial markets, employs two parallel markets that provide alternative methods of trading, called the on-book, or “downstairs,” market and the off-book, or “upstairs,” market. Orders in the on-book market are placed publicly but anonymously and execution is

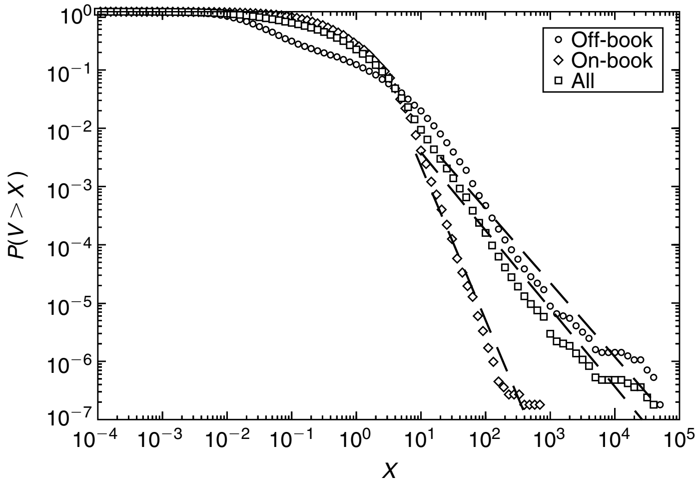

FIGURE 2.3 Volume distributions of off-book trades (*circles*), on-book trades (*diamonds*), and the aggregate of both (*squares*) for a collection of 20 different stocks, normalizing the volume of each by the mean volume before combining. The dashed black lines are for the slope found by the Hill estimator and are shown for the largest 1% of the data. (*Source:* Adapted from Lillo et al., 2005.)

completely automated. The off-book market, in contrast, operates through a bilateral exchange mechanism, via telephone calls or direct contact of the trading parties. The anonymous nature of the on-book market facilitates order splitting—that is, large orders are split into smaller pieces and traded incrementally. On the other hand, the off-book market is a block market, where large orders can be traded in a single transaction. The NYSE data used by Gopikrishnan et al. (2000) includes a mixture of order book trades and block trades. Since the typical size of block trades is much larger than the size of orders traded in the order book, the size of block trades dominates the tail of volume distribution.

This can be seen more clearly in a market (or database) where it is possible to separate block trades from order book trades. In Figure 2.3 (from Lillo et al., 2005) we show the cumulative distribution function of trading volume for off-book trades, on-book trades, and the aggregate of both for a collection of 20 LSE stocks. The distribution of block trades is consistent with the power-law hypothesis of Eq. 2.7 with an exponent close to 1.5, whereas the distribution of order book trades is not consistent with the half-cubic law and instead has a much thinner tail (see also Farmer and Lillo, 2004, and Plerou et al., 2004).

2.5. SUMMARY OF EMPIRICAL RESULTS FOR DIVERSE TYPES OF MARKET IMPACT

The relation between the transacted volume and the consequent expected price shift is called the *price impact*, or alternatively, the *market impact* function. Letting R be a price return associated with a trade of size V, the market impact a time l after the trade occurred is

$$\mathcal{I}(V,l) = E[R|V,l]$$

For many purposes it is useful to separate the dependence on volume from the dependence on time. One can make the hypothesis that the impact function can be written as a product of two functions, that is,

$$\mathcal{I}(V,t) = \mathcal{S}(V)\mathcal{R}(l)$$

In this section we primarily discuss the dependence on volume, saving the discussion of time dependence for Section 2.6.

At this stage we are intentionally being vague about the definition of the return R and the volume V; defining these more precisely is one of the main points of this chapter. The way in which the market impact behaves depends on the market structure as well as on what one means by "return" and "volume." Many studies have empirically investigated market impact with a range of different results; we argue that in many cases these differences stem from differences in what is being studied. The important distinctions that should be made are:

- First, one can consider market impact of an individual transaction vs. an aggregate of many transactions. Aggregation here means that the market impact is conditioned on a given number of trades or to a given interval of time. We discuss the volume dependence of individual impact in Section 2.5.1 and the time dependence in Sections 2.6.2–2.6.5, and we study the properties of aggregate impact in Sections 2.5.2 and 2.6.8.
- A second important aspect is the type of market exchange in which the transactions take place. As we said, most financial markets have upstairs or block markets as well as downstairs or order book markets. In the downstairs market, trades are made by placing orders in a limit order book, and it is quite common to aggressively split large trading orders into many small pieces. The upstairs market trades are arranged bilaterally between individuals. As a result of the varying market structures, the impacts can be quite different.
- A third factor that must be kept in mind is that large trading orders, which we will call *hidden orders*, are typically split into small pieces and executed incrementally. This is in contrast to *realized orders*, which are the actual orders that are traded—for example, the pieces into which hidden orders are split. For realized orders the impacts may be part of a larger process of order splitting that is invisible with the data that we have here. The impacts of hidden orders may be quite different than those of realized orders. The impacts of individual orders behave much like those of individual transactions, as described in the next bullet. We will discuss the impact of hidden orders in Section 2.6.7.
- Finally, even if we have discussed market impact in terms of transacted volume, other events in the market have an impact on price. Specifically, in double auction

market limit, orders and cancellations can have a market impact that is different from the impact of a market order.

In the following sections we discuss the empirical regularities in these different types of market impact.

2.5.1. Impact of Individual Transactions

We now discuss the impact of individual transactions in limit order book markets, whose volume we will denote by v. Many studies have examined the market impact for a single transaction, and all have observed a concave function of the transaction volume v, that is, one that increases rapidly for small v and more slowly for larger v. The detailed functional form, however, varies from market to market and even period to period. Early studies by Hasbrouck (1991) and Hausman, Lo, and MacKinlay (1992) found strongly concave functions but did not attempt to fit functional forms. Keim and Madhavan (1996) also observed a concave impact function for block trades.

Based on Trades and Quotes (TAQ) data for a set of 1000 NYSE stocks, the concavity of the market impact was interpreted by Lillo et al. (2003b) using the functional form

$$E[r|v] = \frac{\epsilon v^{\psi}}{\lambda} \tag{2.8}$$

The exponent $\psi(v)$ is approximately 0.5 for small volumes and 0.2 for large volumes. Even normalizing the volume v by daily volume, the liquidity parameter λ varies for different stocks; there is a clear dependence on market capitalization M that is well approximated by the functional form $\lambda \sim M^{\delta}$, with $\delta \approx 0.4$.

Potters and Bouchaud (2003) analyzed stocks traded at the Paris Bourse and NASDAQ and found that a logarithmic form gave the best fit to the data. For the London Stock Exchange, Farmer and Lillo (2004) and Farmer et al. (2005) found that for most stocks Eq. 2.8 was a good approximation with $\psi = 0.3$, independent of v. Hopman (2007) studied market impact on a 30-minute time scale in the Paris Bourse for individual orders and found $\psi \approx 0.4$, depending on the urgency of the order. Thus all the studies find strongly concave functions but report variations in functional form that depend on the market and possibly other factors as well. Figure 2.4 shows the price impact of buy market orders for five highly capitalized LSE stocks, that is, AZN, DGE, LLOY, SHEL, and VOD. The price impact is well fit by the relation $E[r|v] \propto v^{0.3}$.

2.5.2. Impact of Aggregate Transactions

Studies of aggregated market impact have produced variable results, reaching different conclusions that we will argue depend substantially on the time scale for aggregation. The BARRA market impact model, an industry standard, uses the TAQ data aggregated on a half-hour time scale (Torre, 1997). They compare fits using Eq. 2.8 and find $\psi \approx 0.5$; they obtain similar results using individual block data. Kempf and Korn (1999) studied data for futures on the DAX (the German stock index) on a five-minute time

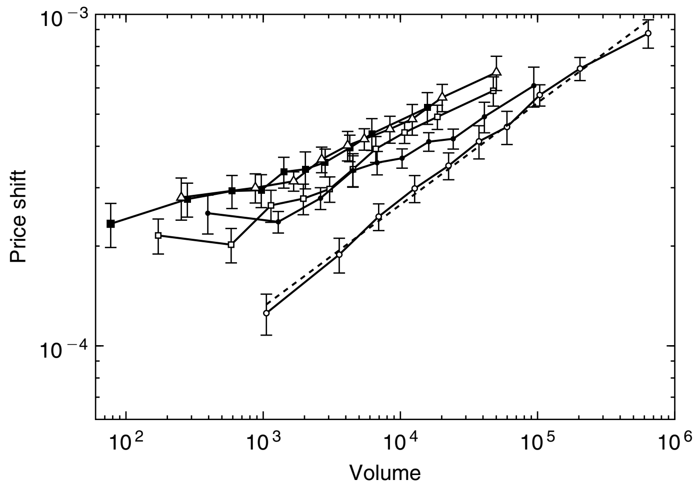

FIGURE 2.4 Market impact function of buy market orders for a set of five highly capitalized stocks traded in the LSE, specifically AZN (*filled squares*), DGE (*empty squares*), LLOY (*triangles*), SHEL (*filled circles*), and VOD (*empty circles*). Trades of different sizes are binned together, and the logarithmic price change's average size for each bin is shown on the vertical axis. The *dashed line* is the best fit of the market impact of VOD with a functional form as described in Eq. 2.8. The value of the fitted exponent for VOD is $\psi = 0.3$.

scale and found a very concave functional form. Plerou et al. (2002) studied data from the NYSE during 1994 and 1995 ranging from 5- to 195-minute time scales and fit the market impact function with a hyperbolic tangent. They noted that at shorter time scales this functional form did not work well for small v; $\tanh(v)$ is linear for small v, but at short time scales (e.g., 5 or 15 minutes) they observed a nonlinear impact function becoming more linear as they went toward longer time scales.

Evans and Lyons (2002) studied foreign exchange rate transactions data for DM and Yen against the dollar at the daily scale over a four-month period. They used the number of buyer-initiated transactions minus the number of seller-initiated transactions as a proxy for the signed order flow volume v and found a strong positive relationship to concurrent returns. Chordia and Subrahmanyam (2004) study impacts of stocks in the S&P 500 at a daily time scale and perform linear regressions but do not compare to other functional forms. For the Paris Bourse Hopman (2007) measures aggregate order flow as $\sum_i \epsilon_i v_i^{\psi}$, where the sum is taken over fixed time intervals. At a daily scale he finds that he gets the best linear regression against contemporary daily returns with $\psi \approx 0.5$. He also documents that the slope of the regression decreases with increasing time scale. Finally, as discussed in more detail later, Gabaix et al. (2003, 2006) have made extensive studies of data from the New York, London, and Paris stock markets on a 15-minute time scale and find exponents $\psi \approx 0.5$.

What is the origin of these differences in the observed functional form of the aggregate market impact? Part of the difference comes certainly from the fact that these studies consider different markets, different assets, and different time periods. However, another important difference across studies is the time scale of aggregation. There is no reason that the aggregate market impact over a ten-minute time interval should have the same functional form of that over a one-hour time interval or over an interval that is defined by 30 trades.

To get an idea of how the market impact changes its shape with aggregation scale, consider a specific example. Let v_t be the volume of transaction happening at time t (in event time). Let $r_t = \log(p_{t+1}/p_t)$ be the corresponding log-return, where p_t is the price of transaction t. For a sequence of N successive transactions beginning at time t, let $Q_N = \sum_{i=1}^{N} \epsilon_{t+i} v_{t+i}$ be the aggregate volume and $R_N = \sum_{i=1}^{N} r_{t+i}$ be the aggregate return. The average market impact conditioned on volume is

$$R(Q, N) = E[R_N | Q_N = Q] \tag{2.9}$$

that is, it is the expected return associated with a signed volume fluctuation Q. We write $R(Q, N)$ to emphasize that this can depend both on the signed trading volume imbalance Q and the number of transactions N. In Figure 2.5 we show empirical estimates for the market impact of the stock AZN, which is traded on the London Stock Exchange, from Lillo et al. (2008a). Figure 2.5 shows the market impact for different values of N with offsets added to the vertical axis to aid visualization. As one would expect, the scale increases with N. The shape of $R(Q, N)$ also changes, becoming more linear with increasing N. This is illustrated more clearly in Figure 2.5(b), where we rescale the horizontal and vertical axes using a rescaling factor based only on Q_N. The renormalization makes the increasing linearity clearer. As N increases, the market impact near $Q = 0$ becomes linear, and the size of the region that can be approximated as linear grows with increasing N. It also illustrates a surprising feature: The slope of the linear region decreases with N. These same basic features (increasing linearity and decreasing slopes) hold for all the stocks in our sample, in both the New York and London Stock Exchanges. This result shows that the shape and the scale of the aggregate market impact change with the aggregation scale. At short time scales the function is significantly nonlinear, but at large aggregation scales the market impact becomes close to linear, and the slope of the impact decays with the aggregation scale. For this reason it is in general misleading to compare aggregate impact curves with different scales unless one has a theory for how the market impact depends on aggregation scale. This also shows why the studies mentioned previously found different forms of the market impact. In Section 2.6.8 we present some models that help explain the behavior of aggregate impact observed in real data.

2.5.3. Hidden Order Impact

Because data for hidden orders, which are sometimes also called *trading packages*, are difficult to obtain, there are only a few studies (Chan and Lakonishok, 1993, 1995; Almgren et al., 2005; Vaglica et al., 2008). These studies show that hidden orders can be extremely long, involving thousands of realized trades spread over periods of many weeks or even months. As reviewed in Section 2.11.1, the most recent

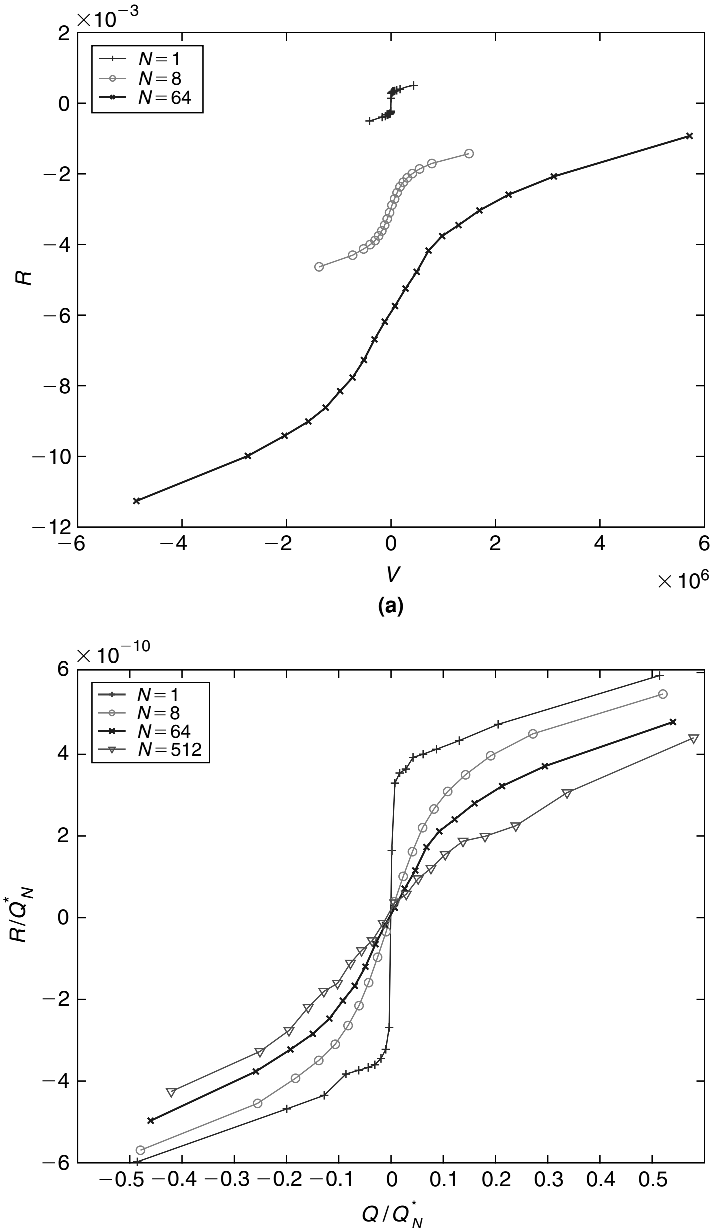

FIGURE 2.5 Aggregate market impact $R(Q, N)$ for the LSE stock Astrazeneca for 2000–2002. **(a)** Plot of the shifted aggregate return $R(Q, N) + R_0$ vs. the aggregate signed volume Q for three values of N. The arbitrary constant R_0 is added to aid visualization; its values are $R_0 = \{0, -3 \times 10^{-3}, -6 \times 10^{-3}\}$ for $N = 1, 8$ and 64, respectively. **(b)** A rescaling for each N of both the horizontal and vertical axes by $Q^*_N = Q^{(95)}_N - Q^{(5)}_N$, where $Q^{(5)}_N$ is the 5% quantile and $Q^{(95)}_N$ is the 95% quantile of Q.

study by Vaglica et al. confirms that hidden orders obey a power-law distribution of size, which, as we argue in Section 2.6, plays an important role in determining their impact.

The theoretical considerations for treating hidden orders are quite different from those for individual orders, and they also very different from those of aggregated anonymous orders. The reason is that such orders come from the same agent, creating bursts of orders in the order flow which are all of the same sign. As we argued in Section 2.4.3, this generates strong correlations in order flow that have to be compensated for, as discussed in Section 2.6. The volume dependence of hidden order impact is intimately connected to the temporal aspects, and so we save the development of the theory for hidden order impact for the next section.

2.5.4. Upstairs Market Impact

Market impact in the upstairs market has been studied by Keim and Madhavan (1996). As in other cases, they find empirically that market impact is concave. They explain this based on a model for the difficulty of finding counterparties for trading. Ultimately, as pointed out by Gabaix et al. (2006), upstairs market impact should match hidden order impact, for the simple reason that the upstairs market is competing with the downstairs market, and if costs in the upstairs market are too high, they have the option of splitting up their trades in the downstairs market. This is convenient because it implies that a theory for either market automatically gives a theory for the other.

2.6. THEORY OF MARKET IMPACT

In this section we develop theoretical explanations for both the volume dependence and the temporal dependence of market impact. As stressed in the previous section, there are several distinct types of impact that require a different approach to their analysis. We begin in Section 2.6.1 by explaining why the impacts associated with individual trades are so concave, arguing that the dominant cause is selective liquidity taking.

Then, in Sections 2.6.2 through 2.6.5, we develop a theoretical approach to understanding the temporal behavior of impacts associated with individual trades. We show that the long memory of order flow and market efficiency play a crucial role, which one can take into account one of two ways. One can either assume a fixed impact, in which case the future contribution to the impact of each trade must decay to zero with time, or one can assume a varying but permanent impact, which implies asymmetry liquidity. We show that these two approaches are equivalent.

In Section 2.6.6 we present empirical results supporting these ideas. In Section 2.6.7 we develop a theory for the impact of hidden orders—that is, linked sets of trades made by large investors. Finally, in Section 2.6.8, we develop a theory for the aggregate impact of successive trades and show that it does a good job of explaining the empirical results of Section 2.5.2.

2.6.1. Why Is Individual Transaction Impact Concave?

Let us first consider the impact of individual transactions. Several different theories have been put forth to explain why market impact for single transactions is concave. These can be grouped into three classes: (1) size-dependent informativeness of trades (e.g., due to stealth trading, as postulated by Barclay and Warner, 1993), (2) average depth vs. price in the limit order book (Daniels et al., 2003), and (3) selective liquidity taking (Farmer et al., 2004).

The standard reason given for the concavity of market impact is that it reflects the informativeness of trades. If small trades carry almost as much information as large trades, the price changes caused by small trades should be nearly as big as those for large trades. For example, this could be due to "stealth trading,"; that is, because informed traders keep their orders small to avoid revealing their superior knowledge (Barclay and Warner, 1993). Hypothesis 2, due to Daniels et al. (2003), is that it reflects the accumulation of liquidity in the limit order book—that is, the depth in the order book as a function of the price will determine the market impact for a market order as a function of its size. Hypothesis 3 is that this is due to selective liquidity taking; that is, that liquidity takers submit large orders when liquidity is high and small orders when it is low (see Farmer et al., 2004; Weber and Rosenow, 2006; and Hopman, 2007).

Theory 2 is easily ruled out by computing the average virtual market impact as a function of volume. This is defined as the average price change that would instantaneously occur for an effective market order of size v (Weber and Rosenow, 2006; Farmer and Zamani, 2007). In Figure 2.6 we show the virtual impact for AZN, computed by

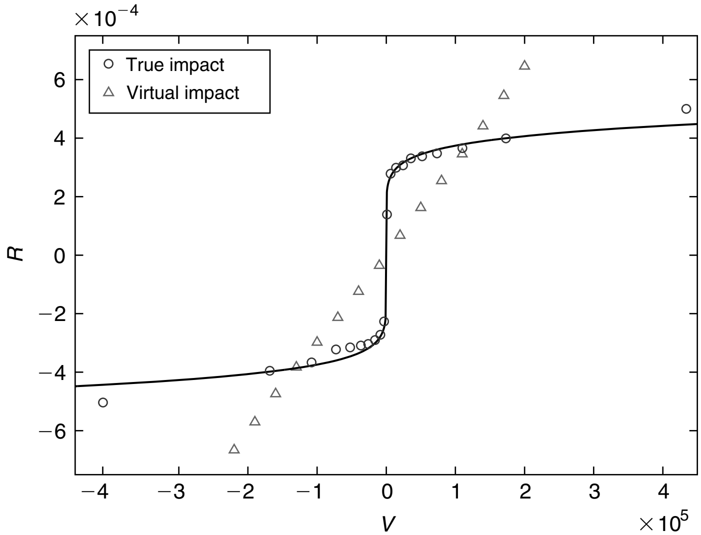

FIGURE 2.6 Comparison of virtual to true market impact: true impact (*circles*), virtual impact (triangles). The fitted curve for true impact (*black line*) is of the form $f(v) = Av^{\psi}$, with $\psi = 0.3$.

hypothetically submitting orders for a range of different values of v and measuring the immediate price response. This is done for each time when real effective market orders were submitted. The resulting price response is a direct probe of the depth of the limit order book. The fact that the mechanical impact is linear to very good degree of approximation makes it clear that this is not the cause of the concavity of the real market impact function.

The selective liquidity taking (Hypothesis 3) means that agents condition the size of their transactions on liquidity, making large transactions when liquidity is high and small transactions when it is low. As shown by Farmer et al. (2004), for LSE stocks it is rare that a trade penetrates more than one price level.[14] For example, for Astrazeneca, approximately 87% of the market orders creating an immediate price change have a volume equal to the volume at the opposite best. Moreover, approximately 97% of the market orders creating an immediate price change have a volume that is either equal to the opposite best or larger than this value but smaller than the sum of volume at the second best opposite price. This means that to a good approximation the market impact can be written in the very simple form

$$E[r|v] = P(+|v)E[r] \tag{2.10}$$

where $P(+|v)$ is the probability that a trade of size v generates a nonzero return—that is, the probability that $v \geq \Phi_b$, where Φ_b is the volume offered or bid at the opposite best price. $E[r]$ is the expected return given that there is a nonzero return, which is of the order of the bid–ask spread (see Section 2.7 for more precise statements). This demonstrates that trading orders that penetrate the opposite best are rare. This is because agents do not like to suffer price degradation more than the opposite best and so condition the size of their orders on what is being offered there.

We have now to explain why $P(+|v)$ is a concave function. An explanation in terms of selective liquidity taking is the following. Suppose that the volume at the best is drawn from a distribution $P_b(\Phi_b)$ and suppose that the liquidity taker draws the volume v she would like to trade from another distribution and independently from Φ_b. If $v < \Phi_b$ she places a market order of size v, whereas if $v > \Phi_b$ she places a market order of size Φ_b. What is the probability $P(+|v)$ under this simple model? A straightforward calculation shows that $P(+|v) = \int_0^v P_b(\Phi_b)d\Phi_b$; that is, it is equal to the cumulative distribution of the volume at the best. This is an increasing and concave function of v that could be used to fit the empirical $P(+|v)$. Under this model the shape of the market impact is explained by $P(+|v)$, that is, by the conditioning of trading orders on the liquidity that is offered. In other words, Theory 3 does a good job of explaining the data, at least qualitatively.

It is a matter of interpretation, however, whether this is also consistent with Theory 1; that is, that smaller trades are proportionately more informative than larger trades. From one point of view, one can simply say that the market impact *defines* the informativeness of trades. If so, then it is obviously consistent. However, if it means that price changes are a response to the new information contained in trades, the evidence presented here

[14] See Table 2 of Farmer et al.

is inconsistent with Theory 1. In the LSE the quoted volume is visible to all, and so, except for occasional latency problems, in which the quote changes just before a trade is placed, the trader is aware of the quote when she places the trade. The fact that the size of the trade is strongly correlated with the size of the best quote implies that the size of the trade carries little new information. This does not mean that the trade is based on inferior information; it merely means that other market participants do not learn much from its size when it occurs. It is the conditioning of trade size on best quotes that drives concavity and not because the smaller trades are nearly as "informed" as the larger trades.

2.6.2. A Fixed Permanent Impact Model

In the previous section we described how midquote prices react on average to market orders of a given volume v. The preceding discussion was restricted to the immediate impact, that is, the impact that is felt immediately after a trade is completed. In general this can have both temporary and permanent components. In this section we discuss the impact of individual transactions—that is, the average midquote price change between just before the n^{th} trade and just before the $n+1^{\text{th}}$ trade. It is an empirical fact that this immediate impact, defined as $E[r_n|\epsilon_n v_n]$, is nonzero and can be written as $E[r|\epsilon v] = \epsilon f(v)$, where f is a function that grows with v. Clearly, it is important to understand if and how this immediate impact evolves with time (which we will measure in terms of the sequence number of the trades). Is the impact of a trade permanent or transient? Is it fixed or is it variable? How does it depend on the past order flow history?

The simplest situation is that of a usual random walker, where position at any time is the sum over all past steps—however far in the past they might be. In financial language, this corresponds to the case where the impact of a transaction is permanent, which translates into the following equation for the midquote price m_n at time n:

$$r_n = m_{n+1} - m_n = \epsilon_n f(v_n; \Omega_n) + \eta_n, \tag{2.11}$$

where η_n is an additional random term describing price changes not directly attributed to trading itself—for example, the impact of news where quotes could instantaneously jump without any trade. We will assume here that η_n is independent of the order flow and we set $E[\eta] = 0$ and $E[\eta^2] = \Sigma^2$. We have included a possible dependence of the impact on the instantaneous state Ω_n of the order book. We expect such a dependence on general grounds: A market order of volume v_n, hitting a large queue of limit orders, will in general impact the price very little. On the other hand, one expects a very strong correlation between the state of the book Ω_n and size of the incoming market order: Large limit order volumes attract larger market orders.

The preceding equation can be written as:

$$m_n = \sum_{k<n} \epsilon_k f(v_k; \Omega_k) + \sum_{k<n} \eta_k \tag{2.12}$$

which makes explicit the nondecaying nature of the impact in this model: $\epsilon_k \partial m_n / \partial v_k$ (for $k < n$) does not decay as $n - k$ grows. This simple model makes the following predictions for the lagged impact function $\mathcal{R}_\ell$ and the lagged return variance $\mathcal{V}_\ell$:

$$\mathcal{R}_\ell \equiv E[\epsilon_n \cdot (m_{n+\ell} - m_n)] = E[f];$$
$$\mathcal{V}_\ell \equiv E[(m_{n+\ell} - m_n)^2] = \left(E[f^2] + \Sigma^2\right) \ell \quad \textbf{(2.13)}$$

that is, constant price impact and pure price diffusion, close to what is indeed observed empirically on small-tick, liquid contracts. However, if we consider the autocovariance of price returns within this model, we find that

$$E[r_n r_{n+\tau}] \propto E[\epsilon_n \epsilon_{n+\tau}] \sim \tau^{-\gamma} \quad \textbf{(2.14)}$$

which means that price returns are strongly autocorrelated in time. This fact would violate market efficiency because price returns would be easily predictable even with linear methods. We therefore come to the conclusion that the empirically observed long memory of order flow is incompatible with the previous random walk model if prices are efficient (Bouchaud et al., 2004; Lillo and Farmer, 2004; Challet, 2007). In other words one of the assumptions of the random walk model must be relaxed. Among the various possibilities we will relax either the assumption that price impact is permanent or the assumption that price impact is independent of the order flow. As we will see, these two possibilities are related one to each other, but for the sake of clarity we present them in two different subsections.

2.6.3. The MRR Model

To illustrate the preceding concepts, let us discuss a slight variant of a model due to Madhavan, Richardson, and Roomans (Madhavan et al., 1997) that helps define various quantities and hone in on relevant questions. The assumptions of the model are (1) that all trades have the same volume $v_n = v$ and (2) the ϵ_n's are generated by a Markov process with correlation ρ, which means that the expected value of ϵ_n conditioned on the past only depends on ϵ_{n-1} and is given by:

$$E\left[\epsilon_n | \epsilon_{n-1}\right] = \rho \epsilon_{n-1} \quad \textbf{(2.15)}$$

The case $\rho = 0$ corresponds to independent trade signs, whereas $\rho > 0$ describes positive autocorrelations of trade signs. Note that in this model, correlations decay exponentially fast, that is,

$$C_\ell = E[\epsilon_i \epsilon_{i+\ell}] = \rho^\ell \quad \textbf{(2.16)}$$

which, as we discussed in Section 2.4, does not conform to reality.

The MRR model postulates that the midpoint m_n evolves only because of unpredictable external shocks (or news) and because of the surprise component in the order flow. This postulate, of course, automatically removes any predictability in the price returns and ensures efficiency. Under the assumption that the surprise component of the order flow at the n^{th} trade is given by $\epsilon_n - \rho\epsilon_{n-1}$, one writes the following evolution equation for the price[15]

$$m_{n+1} - m_n = \eta_n + \theta[\epsilon_n - \rho\epsilon_{n-1}] \quad \textbf{(2.17)}$$

where η is the shock component and the constant θ measures the size of trade impact.

These equations make it possible to compute several important quantities such as the lagged impact function defined earlier (Eq. 2.13). One may write:

$$m_{n+\ell} - m_n = \sum_{j=n}^{n+\ell-1} \eta_j + \theta \sum_{j=n}^{n+\ell-1} [\epsilon_j - \rho\epsilon_{j-1}] \quad \textbf{(2.18)}$$

The full impact function is found to be constant, equal to:

$$\mathcal{R}_\ell = \theta(1-\rho^2), \quad \forall \ell \quad \textbf{(2.19)}$$

We can also define the "bare" impact of a single trade $G_0(\ell)$, which measures the influence of a single trade at time $n - \ell$ on the the midpoint at time n. In terms of $G_0(\ell)$, the midpoint is therefore written as:

$$m_n = \sum_{j=-\infty}^{n-1} \eta_j + \sum_{j=-\infty}^{n-1} G_0(n-j-1)\,\epsilon_j \quad \textbf{(2.20)}$$

here found to be given by $G_0(\ell = 0) = \theta$ and $G_0(\ell \geq 1) = \theta(1-\rho)$; a part $\theta\rho$ of the impact instantaneously decays to zero after the first trade, whereas the rest of the impact is permanent. The instantaneous drop of part of the impact compensates the sign correlation of the trades. Finally, the volatility, within this simplified version of the MRR model, reads:

$$\left[\theta^2(1-\rho^2) + \Sigma^2\right]\ell \quad \textbf{(2.21)}$$

2.6.4. A Transient Impact Framework

Compared to the simplifying assumptions of the MRR model, the data shows that (1) the volumes v of the incoming market orders are very broadly distributed, with a power-law tail (see Section 2.4.5); (2) the sign time series ϵ_n has long-range correlations C_ℓ that decays again as a power-law $\sim c_0\ell^{-\gamma}$ with $\gamma < 1$, defining a long-memory process. The smallness of γ makes the correlation function C_ℓ nonsummable: The average relaxation time is infinite, whereas the correlation time of the Markovian sign process in the preceding MRR model is finite, equal to $(1-\rho)^{-1}$.

In this section we relax the assumption that impact of a single trade is permanent in time. Rather, we find that long-range correlations in trades imply that the impact

[15] The assumption that prices respond linearly to the order flow is a very strong assumption.

itself has to decay slowly with time. In the next section, we discuss an alternative but equivalent model, where the impact is permanent but asymmetric and history dependent.

Transient Impact and Mean Reversion

What would happen if the impact of each trade was purely transient—for example, an exponential decay in time? Eq. 2.11 would now read:

$$m_n = \sum_{k<n} \alpha^{n-k-1} \epsilon_k f(v_k; \Omega_k) + \sum_{k<n} \eta_k, \quad (0 \leq \alpha < 1) \tag{2.22}$$

The lagged impact and the return variance would then be given by:

$$\mathcal{R}_\ell = \alpha^{\ell-1} E[f] \quad \mathcal{V}_\ell = 2E[f^2]\frac{1-\alpha^\ell}{1-\alpha^2} + \Sigma^2 \ell \tag{2.23}$$

That is, a short-time volatility $\approx E[f^2] + \Sigma^2$ larger than its long-time value Σ^2, in which only the "news" component survives. The price would exhibit significant high-frequency mean reversion: Impact kicks it temporarily up and down, but the long-term wandering of the price is unrelated to trading. Of course, one could be in a mixed situation where the impact decays exponentially but toward a positive value, in which case the long-term volatility still involves an impact component. This conforms with conventional wisdom about efficient markets: an increased value of high-frequency volatility driven by the *tâtonnement* process and a long-term volatility made up both of unexpected news and long-term impact of market orders, which translates private information into prices. However, recall that this does not conform to observations, which show volatility very nearly constant across all time scales (see Section 2.3.9).

What is the relation between the average $\mathcal{R}_\ell$ and the impact of a single trade that we call $G_0(\ell)$ henceforth? If trades were uncorrelated, the two quantities would be identical, but trade correlations, as we shall see, change the picture in a rather interesting way.

Mathematical Theory of Long-Term Resilience

The long-term memory of trades is *a priori* paradoxical and hints of a nontrivial property of financial markets, which can be called *long-term resilience*. Take again Eq. 2.20 with the assumption that single trade impact is lag independent, $G_0(\ell) = G_0$, and that volume fluctuations can still be neglected. The midprice variance is easily computed to be:

$$\mathcal{V}_\ell \equiv \langle (m_{n+\ell} - m_n)^2 \rangle = [\Sigma^2 + G_0^2]\ell + 2G_0 \sum_{j=1}^{\ell} (\ell - j) C_j \tag{2.24}$$

When $\gamma < 1$, the second term of the *rhs* can be approximated, when $\ell \gg 1$, by $2c_0 G_0 \ell^{2-\gamma}/(1-\gamma)(2-\gamma)$, which grows faster than the first term. In other words, the price would *superdiffuse*, or trend, at long times, with a volatility diverging with the lag ℓ. This, of course, does not occur: The market reacts to trade correlations so as to prevent the occurence of such trends. In fact, within the present linear model, the impact to single trades must be transient rather than permanent. Before explaining why and how this occurs in practice, let us first express mathematically how the efficiency of prices imposes strong constraints on the shape of the single trade impact function. For an arbitrary function $G_0(\ell)$, the lagged price variance can be computed explicitly and reads:

$$\mathcal{V}_\ell = \sum_{0\le j<\ell} G_0^2(\ell - j) + \sum_{j>0} [G_0(\ell + j) - G_0(j)]^2 + 2\Delta(\ell) + \Sigma^2\ell \qquad \textbf{(2.25)}$$

where $\Delta(\ell)$ is the correlation-induced contribution:

$$\begin{aligned}\Delta(\ell) = &\sum_{0\le j<k<\ell} G_0(\ell - j)G_0(\ell - k)C_{k-j} \\ &+ \sum_{0<j<k} [G_0(\ell + j) - G_0(j)]\,[G_0(\ell + k) - G_0(k)]\,C_{k-j} \\ &+ \sum_{0\le j<\ell}\sum_{k>0} G_0(\ell - j)\,[G_0(\ell + k) - G_0(k)]\,C_{k+j}\end{aligned} \qquad \textbf{(2.26)}$$

Assume that $G_0(\ell)$ itself decays at large ℓ as a power law, $\Gamma_0\ell^{-\beta}$. When $\beta, \gamma < 1$, the asymptotic analysis of $\Delta(\ell)$ yields:

$$\Delta(\ell) \approx \Gamma_0^2 c_0 I(\gamma, \beta)\ell^{2-2\beta-\gamma} \qquad \textbf{(2.27)}$$

where $I > 0$ is a certain numerical integral. If the single trade impact does not decay ($\beta = 0$), we recover the above superdiffusive result. But as the impact decays faster, superdiffusion is reduced, until $\beta = \beta_c = (1 - \gamma)/2$, for which $\Delta(\ell)$ grows exactly linearly with ℓ and contributes to the long-term value of the volatility. However, as soon as β exceeds β_c, $\Delta(\ell)$ grows sublinearly with ℓ, and impact only enhances the high-frequency value of the volatility compared to its long-term value Σ^2, dominated by "news." We therefore reach the conclusion that the long-range correlation in order flow does not induce long-term correlations nor anticorrelations in the price returns if and only if the impact of single trades is transient ($\beta > 0$) but itself nonsummable ($\beta < 1$). This is a rather odd situation in which the impact is not permanent (since the long-time limit of G_0 is zero) but is not transient either because the decay is extremely slow. The convolution of this semipermanent impact with the slow decay of trade correlations gives only a finite contribution to the long-term volatility. The mathematical constraint $\beta = \beta_c$ will be given more financial flesh later.

Within this framework, one can also compute the average impact function $\mathcal{R}_\ell$. From Eq. 2.13 one readily obtains:

$$\mathcal{R}_\ell = G_0(\ell) + \sum_{0<j<\ell} G_0(\ell - j)C_j + \sum_{j>0} [G_0(\ell + j) - G_0(j)]\,C_j \qquad \textbf{(2.28)}$$

This equation can be understood as a way to extract the impact of single trades G_0 from directly measurable quantities, such as $\mathcal{R}_\ell$ and C_n; see Section 2.6.6 and Appendix 2.2. From a mathematical point of view, the asymptotic analysis can again be done when $G_0(\ell)$ decays as $\Gamma_0 \ell^{-\beta}$. When $\beta + \gamma < 1$, one finds:

$$\mathcal{R}_\ell \approx_{\ell \gg 1} \Gamma_0 c_0 \frac{\Gamma(1-\gamma)}{\Gamma(\beta)\Gamma(2-\beta-\gamma)} \left[\frac{\pi}{\sin \pi\beta} - \frac{\pi}{\sin \pi(1-\beta-\gamma)} \right] \ell^{1-\beta-\gamma} \qquad \textbf{(2.29)}$$

where we have explicitly given the numerical prefactor to show that it exactly vanishes when $\beta = \beta_c$, which means that in this particular case one cannot satisfy oneself with the leading term. When $\beta < \beta_c$, one finds that $\mathcal{R}_\ell$ diverges to $+\infty$ for large ℓ, whereas for $\beta > \beta_c$, $\mathcal{R}_\ell$ diverges to $-\infty$, which is perhaps counterintuitive but means that when the decay of single trade impact is too fast, the accumulation of mean reverting effects leads to a negative long-term average impact—see Figure 2.7. When β is precisely equal to β_c, $\mathcal{R}_\ell$ tends to a finite positive value $\mathcal{R}_\infty$: The decay of single trade impact precisely offsets the positive correlation of the trades.

In this framework, volume fluctuations have been neglected. An extended version of the model, which is directly related to the discussion in the next section, is presented in Appendix 2.2 (see also Bouchaud et al., 2004).

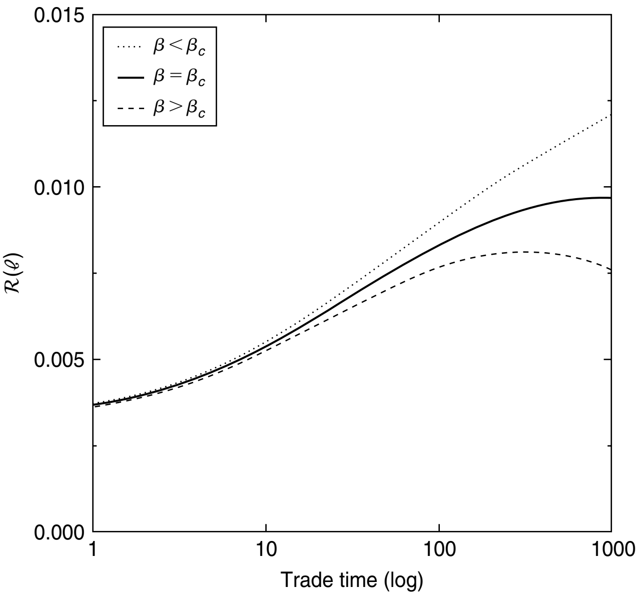

FIGURE 2.7 Theoretical impact function $\mathcal{R}_\ell$, from Eq. 2.28, and for values of β close to β_c. When $\beta = \beta_c$, $\mathcal{R}_\ell$ tends to a constant value as ℓ becomes large. When $\beta < \beta_c$ (slow decay of G_0), $\mathcal{R}_{\ell\to\infty}$ diverges to $+\infty$, whereas for $\beta > \beta_c$, $\mathcal{R}_{\ell\to\infty}$ diverges to $-\infty$.

2.6.5. History Dependent, Permanent Impact

An alternative interpretation of the preceding formalism is to assume that price impact is permanent, but history dependent as to ensure statistical efficiency of prices (Lillo and Farmer, 2004; Farmer et al., 2006; Gerig, 2007).

Predictable Order Flow and Statistical Efficiency

Let us consider a generalized MRR model:

$$r_n = m_{n+1} - m_n = \eta_n + \theta(\epsilon_n - \hat{\epsilon}_n), \quad \hat{\epsilon}_n = E_n[\epsilon_{n+1}|I] \tag{2.30}$$

where I is the information set available at time n. In line with our discussion in Section 2.3.8, we assume that in the market there are three types of traders. First, there are directional traders (liquidity takers) that have large hidden orders to unload and, by placing many consecutive orders with the same sign, create a correlated order flow. The second group of agents consists of the liquidity providers, who post bids and offers and attempt to earn the bid–ask spread. The third group is made by noise traders—that is, traders placing uncorrelated order flow. Anticipating the discussion in Section 2.7.3, it is indeed reasonable to assume that the strategies of the first two types of agents will adjust in such a way as to remove any predictability of the midpoint change; in other words, that $E_{n-1}[r_n|I] = 0$ as implied by Eq. 2.30. This is a plausible first approximation, although one can expect (and indeed observe) deviations from strict unpredictability at high frequencies.

Within the preceding simplified model, in which we have neglected volume fluctuations (see Appendix 2.2 for an attempt to include them), there are only two possible outcomes. Either the sign of the n^+ transaction matches the sign of the predictor $E_n[\epsilon_{n+1}|I]$ or they are opposite. Let us call r_n^+ and r_n^- the expected ex-post absolute value of the return of the n^{th} transaction, given that ϵ_n either matches or does not match the predictor. If we indicate with φ_n^+ and (φ_n^-) the *ex ante* probability that the sign of the n^{th} transaction matches (or disagrees) with the predictor ϵ_n, we can rewrite $E_{n-1}[r_n|I] = 0$ as:

$$\varphi_n^+ r_n^+ - \varphi_n^- r_n^- = 0 \tag{2.31}$$

Within the MRR model as described previously, this means

$$r_n^+ = \theta(1 - \hat{\epsilon}_n) \tag{2.32}$$

$$r_n^- = \theta(1 + \hat{\epsilon}_n) \tag{2.33}$$

This result shows that the most likely outcome has the smallest impact. We call this mechanism *asymmetric liquidity*: Each transaction has a permanent impact, but the impact depends on the past order flow and its predictability. The price dynamics and the impact of orders therefore depend on (1) the order flow process, (2) the information set I available to the liquidity provider, and (3) the predictor used by the liquidity provider to forecast the order flow.

Equivalence with the Transient Impact Model

In the following we consider the case where the information set available to liquidity providers is restricted to the past order flow. We call this information set *anonymous* because liquidity providers do not know the identity of the liquidity takers and are unable to establish whether or not two different orders come from the same trader. We assume also that the predictor used by liquidity takers to forecast future order flow comes from a linear model. In some cases, such as for an order flow generated according to the model presented in Section 2.4.3, this may not be an optimal predictor. However, linear time series models are probably the most widely used forecasting tools. Here we analyze a linear time series model based on the signs of executed transactions, and we assume a K^{th} order autoregressive AR model of the form

$$\hat{\epsilon}_n = \sum_{i=1}^{K} a_i \epsilon_{n-i} \tag{2.34}$$

where a_i are real numbers that can be estimated on historical data using standard methods (see Lillo and Farmer, 2004; Bouchaud et al., 2004; and Appendix 2.2). The MRR model corresponds to an AR(1) order flow, with $a_1 = \rho$ and $a_k = 0$ for $k > 1$, with an exponential decay of the correlation.

The resulting impact model, Eq. 2.30 with a general linear forecast of the order flow, is in fact *equivalent*, when $K \to \infty$, to the temporary impact model of the previous section (see the appendix of Bouchaud et al., 2004). It is easy to show that one can rewrite the generalized MRR model in terms of a propagator as

$$m_n = m_{n-1} + \theta\epsilon_n + \sum_{i=1}^{\infty}[G(i+1) - G(i)]\epsilon_{n-i} + \eta_n, \quad \theta = G(1) \tag{2.35}$$

The equivalence is obtained with the relation:

$$\theta a_i = G(i+1) - G(i) \quad \text{or} \quad G(i) = \theta[1 - \sum_{j=1}^{i-1} a_j] \tag{2.36}$$

More General Information Models

In the previous section we saw that the fixed/temporary impact model is equivalent to the variable/permanent impact model under the additional assumptions that (1) the information set available to the liquidity provider is the set of the past order flow and (2) that liquidity providers use a linear forecast model to predict the future order flow from the past and to adjust price response. These two assumptions of the variable/permanent impact model are far from general. In the following we discuss the more general situations in which a different information set and forecast model can arise.

In most financial markets order flow is available in real time to all market participants; thus it is clear that any liquidity provider could use the past order flow time series to trade efficiently. However, in some cases participants can use information other than the time series of order flow signs. There are often indirect clues about the identity of orders such as the consistent use of particular round lots for orders that arrive at regular intervals. Activity in block markets can also provide clues about the activity of large orders. Another case is when a trader is trying to execute his large order by a so-called "slicing and dicing" algorithm. The liquidity provider could be able to detect the presence of this trader, and therefore the liquidity provider has additional information to add to his information set.

The algorithm used by the liquidity provider to forecast the future order flow depends on the information set and on the degree of sophistication of the liquidity provider. Even if linear forecasting methods are widespread, they can lead to suboptimal predictions if the time series one is trying to forecast is strongly nonlinear. For example, in Section 2.4.3 we discussed a microscopically based order flow model that reproduces the correlation properties observed in the real order flow. This model (Lillo, Mike, and Farmer, 2005) is clearly nonlinear. Despite the fact that an optimal forecast method for this order flow model is not easily available, one can find suboptimal nonlinear forecast models that outperform the linear forecast method. When one incorporates nonlinear forecast models in the variable/permanent impact model, the price dynamics will not be equivalent to the fixed/temporary model.

In conclusion, the variable/permanent model sets a general framework for describing the interaction between order flow and price dynamics. In a paper in progress, Gerig et al. (2008) show how different assumptions on the information set and on the forecast method lead to different functional forms of the impact of hidden orders and on the dynamical properties of prices.

Mechanisms for Asymmetric Liquidity

Let us rephrase in more intuitive terms the results established earlier. Due to the small outstanding liquidity, order flow must develop temporal correlations. This is such an obvious empirical fact that high-frequency traders/market makers quickly come to learn about it and adapt to it. In the simple MRR model where signs are exponentially correlated, the probability that a buy follows a buy is $p_+ = (1+\rho)/2$. The unconditional impact of a buy is θ (see Eq. 2.60); however, a second buy immediately following the first has a reduced impact equal to $\mathcal{R}_1^+ = \theta(1-\rho)$. The second buy is not as surprising as the first and therefore should impact the price less. A sell immediately following a buy, on the other hand, has an *enhanced* impact equal to $\mathcal{R}_1^- = \theta(1+\rho)$, in such a way that the conditional average impact of the next trade is zero: $p_+\mathcal{R}_1^+ + (1-p_+)\mathcal{R}_1^- \equiv 0$ Gerig (2007). This is the "asymmetric liquidity" effect explained previously (Lillo and Farmer, 2004; Farmer et al., 2006; and Gerig, 2007; see also Bouchaud et al., 2006, where it is called "liquidity molasses"). This mechanism is expected to be present in general; because of the positive correlation in order flow, the impact of a buy following a buy should be less than the impact of a sell following a buy—otherwise, trends would appear.

But what are the mechanisms responsible for asymmetric liquidity, and how can they fail (in which case markets cease to be efficient)? This is still an open empirical question that started to be investigated only recently. For example, Lillo and Farmer (2004) showed that when the order flow becomes more predictable, the probability that a market order triggers a price change is larger for market orders with the unexpected sign than for those with the expected one. Moreover, the same authors showed that the ratio between the volume of the market order and the volume at the opposite best is lower (higher) for market orders with an expected (unexpected) sign.

Another related basic mechanism is "stimulated refill": Buy market orders trigger an opposing flow of sell limit orders, and vice versa (Bouchaud et al., 2006). This rising wall of limit orders decreases the probability of further upward moves of the price, which is equivalent to saying that $\mathcal{R}_1^+ < \mathcal{R}_1^-$, or else that the initial impact of the first trade reverts at the second trade. This dynamical feedback between market orders and limit orders is therefore fundamental for the stability of markets and for enforcing efficiency. It can be directly tested on empirical data. For example, Weber and Rosenow (2005) have found strong evidence for an increased limit order flow compensating market orders.

Since such a dynamical feedback is so important to reconcile correlation in order flow with the diffusive nature of price changes, it is worth detailing its intimate mechanism a little further and insisting on cases where this feedback may break down. Recall our discussion of the market ecology in Section 2.3.8: Market participants can be, in a first approximation, classified as a function of their trading frequencies. Large latent demand arises from low-frequency participants; the decision to buy or sell can be considered as fixed over a time scale of a few hours or a few days, much longer than the average time between trades. These participants create long-term correlations in the sign of the trades. Higher-frequency traders try to profit from microstructural effects and short time predictability. Even if institutionally designated market makers are no longer present in most electronic markets, these high-frequency strategies are in fact akin to market making—they make money from providing liquidity to lower-frequency traders. This is why we often (incorrectly) call this category of participants *market makers*.[16] So, one should think of two rather large latent supply and offer quantities that await favorable conditions, in terms of both price and quantity, to be executed on the market. Then begins a kind of hide-and-seek game, where each side attempts to guess the available liquidity on the other side. A "tit-for-tat" process then starts, whereby market orders trigger limit orders and limit orders attract market orders. A buy trade at the ask (say) is a signal that an investor is indeed willing to trade at that particular price. But the seller who placed a limit order at the ask is also, by definition, willing to trade at that price. The natural consequence is that a flow of refill orders is expected to occur at the ask immediately after a buy trade (and at the bid after a sell).

[16]Of course, the preceding distinction between participants must be taken with a grain of salt: Low-frequency decisions may be executed using smart high-frequency algorithmic trading. In this case, the same participant is at the same time a low-frequency trader and a market maker.

In other words, optimized execution strategies that look for micro-opportunities impose strong correlations between market order flow of one sign and limit order flow of the opposite sign. Imagine a case where buy market orders eat up sell limit orders at the ask, with no refill. The ask then moves up one tick. By making the price more expensive, the flow of buy market orders slows and the probability that a sell limit order reappears at the previous ask increases. Imagine now that the refill process is too intense; sell limit orders at the ask now pile up. This has two effects: (1) the probability of a large market order that executes a large volume in one shot increases; (2) the large volume at the ask decreases the probability of further sell limit orders joining the queue because the priority of these new orders is low. Both cases (no refill or intense refill) therefore induce a clear feedback mechanism ensuring local stability of the order book.

The previous mechanism can be thought of as a dynamical version of the supply-demand equilibrium, in the following sense: Incipient up trends quickly dwindle because as the ask moves up, the buy pressure goes down, while the sell pressure increases. Conversely, liquidity induced mean reversion—keeping the price low—attracts more buyers and soon gives way. Such a balance between liquidity taking and liquidity providing is at the origin of the subtle compensation between correlation and impact explained previously. It is interesting to notice that several other dynamical systems operate similarly, with a competition between two antagonist systems; heartbeats is an interesting example: The sympathetic and parasympathetic system act in opposition to speed/slow the cardiac rhythm.

One easily envisions that such a subtle dynamical equilibrium can quickly break down; for example, an upward fluctuation in buy order flow might trigger a momentary panic, with the opposing side failing to respond immediately. These liquidity micro-crises are probably responsible for the large number of price jumps; if the feedback mechanism changes sign, this can even lead to crashes. The tug of war is a vivid illustration of this phenomenon. A major challenge of microstructure theory is to turn the previous qualitative story into a quantitative model for heavy-tailed return distributions and volatility clustering, with interesting potential ideas on how to limit the occurence of these liquidity micro-crises. We are convinced that a consistent theory of hidden liquidity and stimulated refill is well within reach at this stage.

2.6.6. Empirical Results

The section reviews how the preceding ideas can be directly tested and measured on high-frequency data.

We start with the full impact function, defined by Eq. 2.13, which is easily measured, at least when the lag ℓ is not too large. When ℓ becomes of the order of the number of daily trades or more, the error bar on $\mathcal{R}_\ell$ quickly becomes large. The main features of $\mathcal{R}_\ell$ are, however, quite robust from stock to stock and also across different markets. For example, $\mathcal{R}_\ell$ for France Telecom in 2002 is shown in Figure 2.8. One sees a mild increase by a factor $\lambda \sim 2$ between $\ell = 1$ and $\ell = 1000$ before a saturation or maybe a decline for larger lags. This behavior is quite typical, in particular the

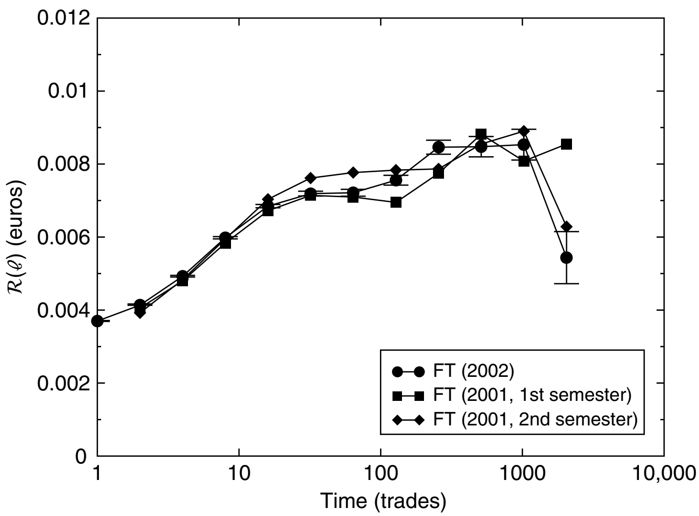

FIGURE 2.8 Average empirical response function $\mathcal{R}_\ell$ for FT during three different periods (1st and 2nd semester of 2001 and 2002); error bars are shown for the 2002 data. For the 2001 data, the y axis has been rescaled such that $\mathcal{R}_1$ coincides with the 2002 result. $\mathcal{R}_\ell$ is seen to increase by a factor $\sim$2 between $\ell = 1$ and $\ell = 100$.

roughly twofold increase between small lags and large lags. So $\mathcal{R}_\ell$ reveals some non-trivial temporal structure; recall that $\mathcal{R}_\ell$ is constant within models where the midpoint reacts to surprise in order flow. In an MRR setting, the amplification factor λ should be $1/(1 - C_1)$, which is found to be in the range of 1.2 to 1.4, still too small to explain $\lambda \sim 2$.

As noted, one can in fact extract the theoretical impact of single trades $G_0(\ell)$ from the empirically measured impact $\mathcal{R}_\ell$ and the correlation between the sign of the trades C_ℓ, using Eq. 2.28. This was done in Bouchaud et al. (2006) and indeed produces nice, power-law decaying $G_0(\ell)$'s; see Figure 2.9 for a few examples. Within the previous restrictive theoretical framework, this provides a direct proof of the transient nature of the impact of single market orders and the long-term resilience of markets. This is quite important as far as execution strategies are concerned; see Section 2.10.

We should, however, list a number of caveats. One is the assumption that the impact is time translation invariant—that is, only the lag ℓ is relevant. This is clearly questionable, since strong intraday seasonality effects are expected. For example, there are indications that the trade sign correlation function C_ℓ for a given lag ℓ is quite different intraday and from one day to the next (Eisler et al., 2008). Similarly, we expect that the single trade impact should decay differently intraday and overnight. Second, we have to a large extent discarded the interesting correlations between the state of the order book Ω_n, the incoming volume v_n and the resulting impact (see Eq. 2.11). All this complexity was replaced by an average description: $\epsilon_n f(v_n; \Omega_n) \longrightarrow \epsilon_n \ln v_n$. Certainly, a refined version is needed, in particular because the fluctuations of $f(v_n; \Omega_n)$ will contribute to

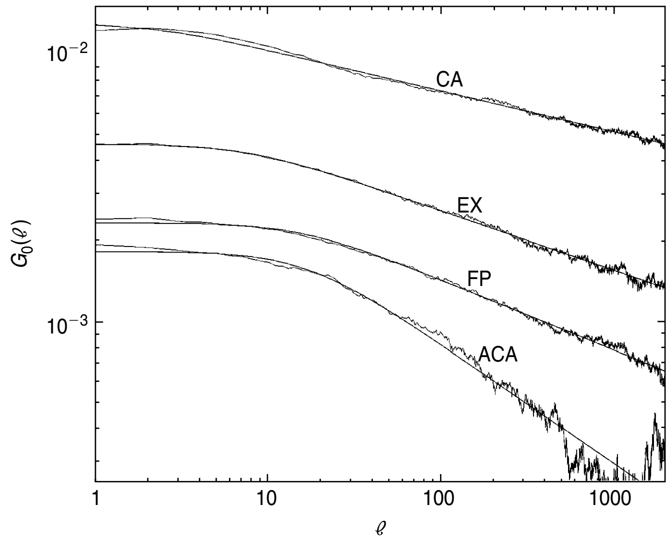

FIGURE 2.9 Comparison between the empirically determined $G_0(\ell)$, extracted from $\mathcal{R}$ and $\mathcal{C}$ using Eq. 2.28, and the power-law fit $G_0^f(\ell) = \Gamma_0/(\ell_0^2 + \ell^2)^{\beta/2}$ for a selection of four stocks: ACA, CA, EX, and FP.

the diffusion properties (see Eq. 2.25). Finally, we have chosen from the start to give a special role to market orders, as if only those impact the price. But this is not true: Obviously, limit orders also impact the price. In fact, it is precisely the impact of limit orders that offsets that of market orders and leads to a decay of the single trade impact $G_0(\ell)$. In other words, we have studied an effective model in terms of market orders only, dumping into $G_0(\ell)$ the counteracting effect of limit orders. A more symmetric version of the model, which treats market and limit orders on an equal footing, would be quite enticing (Eisler et al., 2008).

We now consider some empirical evidence for asymmetric liquidity. Figure 2.10 shows the behavior of the conditional returns r^+ and r^- defined in Eq. 2.33 as a function of the sign predictor $\hat{\epsilon}$. The data we show in Figure 2.10 is for Astrazeneca, a stock traded at the LSE. The sign predictor is the linear predictor defined in Eq. 2.33. The larger the absolute value of $\hat{\epsilon}$, the stronger the predictability of the next market order sign. We have plotted the average value of the return conditioned to be in the direction of the predictor, r^+, and the average return when the sign of the predictor is wrong, r^-. We see that r^- is indeed larger than r^+ and this difference increases with the predictability of the order flow. This is a clear evidence for asymmetric liquidity. Note also that both r^+ and r^- are approximately described by a linear function of the predictor $\hat{\epsilon}$. This is expected under the model described in the "Predictable Order Flow and Statistical Efficiency" section (see Eq. 2.33). However, the slopes of r^+ and r^- vs. $\hat{\epsilon}$ are different, challenging the implicit symmetric assumption (Eq. 2.33) in the MRR model. Other evidence for the buildup of the "liquidity molasses" accompanying the flow of market order can be found in Bouchaud et al. (2006) and Weber and Rosenow (2005).

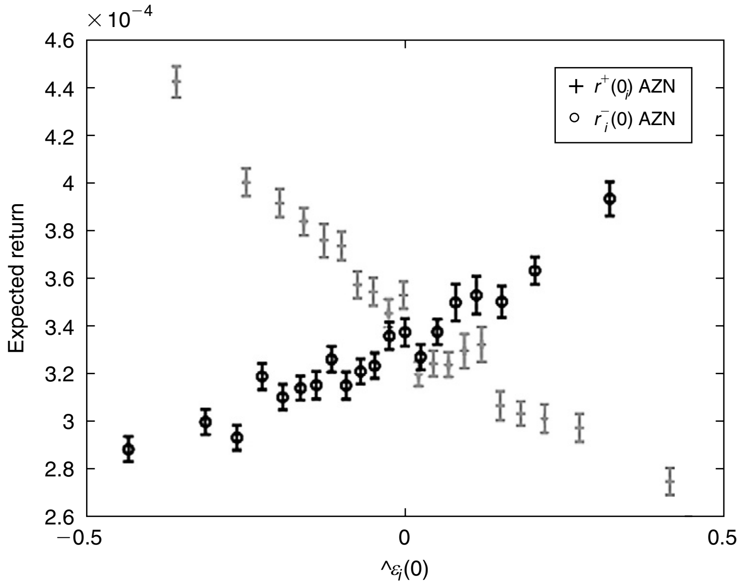

FIGURE 2.10 Expected return as a function of the sign predictor $\hat{\epsilon}$. The quantity r^+ (r^-) refers to trades with a sign that is equal (opposite) to the one of the predictor. Data are binned in such a way that each point contains an equal number of observations. Error bars are standard errors. (*Source:* Adapted from Gerig, 2007.)

2.6.7. Impact of a Large Hidden Order

We now want to calculate, within the stated theoretical framework, the impact of a hidden order of size N. For simplicity, let us first assume that the hidden order is made of N consecutive trades made by the same institution, though this remains "hidden" if trades are anonymous. Let us call m_0 the price at the beginning of the hidden order and compute the average price m_{N+t} observed t transactions after the completion of the hidden order. Within the generalized MRR model with a linear predictor of the order flow, a straightforward calculation shows that

$$E[m_{N+t}] - m_0 = \epsilon\theta \sum_{i=t+1}^{t+N} [1 - \sum_{j=1}^{i-1} a_j] \tag{2.37}$$

For $t = 0$ this expression gives the (temporary) total impact of the hidden order, whereas for $t > 0$ we can calculate the price reversion after the completion of the hidden order, and the permanent impact (if any) for $t \to \infty$.

This result can be generalized to the case where there is only one hidden order active at a given time, which mixes with a flow of uncorrelated orders with a constant participation rate π. The total time needed to execute the hidden order is then $T = N/\pi$. It is possible to show in this case that (Farmer, Gerig, Lillo, and Waelbroeck, 2008):

$$E[m_N] - m_0 = \epsilon\theta \sum_{i=1}^{N} \left(1 - \sum_{k=1}^{i/\pi} a_k\right) \tag{2.38}$$

Let us estimate this formula in the case where the autocorrelation C_τ of order flow asymptotically decays as a power law $C_\tau \sim \tau^{-\gamma}$ for large τ. There are several different ways of generating and forecasting long-memory processes. Here we assume that the participants observing public information model the time series with a FARIMA process. It is known (Beran, 1994) that for large k the best linear predictor coefficients of a FARIMA process satisfy $a_k \approx k^{-\beta-1}$, where $\beta = (1-\gamma)/2$. For large k we can go into the continuum limit, and from Eq. 2.38 the impact is:

$$E[m_N] - m_0 = \epsilon\theta \left[1 + \sum_{i=1}^{N-1} \left(1 - \left(1 - (n/\pi)^{-\beta}\right)\right)\right] \tag{2.39}$$

Converting the sum to an integral gives:

$$E[m_N] - m_0 \approx \epsilon\theta \left(1 + \frac{2^{\beta-1}\pi^{\beta}}{1-\beta}[(2N-1)^{1-\beta} - 1]\right) \sim \pi^{\beta} N^{1-\beta} \tag{2.40}$$

Thus, for a fixed participation rate, the market impact asymptotically increases with the length of the hidden order as $N^{1-\beta}$. A typical decay exponent for the autocorrelation of order signs is $\gamma \approx 0.5$ (Lillo and Farmer, 2004; Bouchaud et al., 2004), which means that $\beta \approx 0.25$. This means that according to the linear time series model, the impact should increase as roughly the $\frac{3}{4}$ power of the order size. An interesting property of this solution is that it depends on the speed of execution. The size of the impact varies as π^{β}. This means that the more slowly an order is executed, the less impact it has, and in the limit as the order is executed infinitely slowly, the impact goes to zero. Note, however, that if the execution time $T = N/\pi$ is *fixed*, the impact becomes linear with N but decays as $T^{-\beta}$.

To investigate the reversion dynamics, we again make use of the Eq. 2.37. We assume that the liquidity provider uses a FARIMA model to forecast order signs, and for the sake of simplicity in the following, we assume that $\pi = 1$—that is, that there are no noise traders. Realistically, the regression made by the liquidity provider on past signs will use a finite lag K, leading to:

$$\hat{\epsilon}_n = \sum_{i=1}^{K} a_i^{(K)} \epsilon_{n-i} \tag{2.41}$$

where (Beran, 1994):

$$a_i^{(K)} = -\binom{K}{i} \frac{\Gamma(i-H+1/2)\Gamma(K-H-i+3/2)}{\Gamma(1/2-H)\Gamma(K-H+3/2)} \tag{2.42}$$

and $H = 1/2 - \beta$ is the Hurst exponent of the FARIMA process. It is possible to derive an analytical exact result for the permanent impact. In fact, from Eq. 2.37, one can obtain:

$$E[m_\infty] - m_0 = \epsilon\theta N(1 - \sum_{j=1}^{K} a_j^{(K)}) = \epsilon\theta N \frac{4^{H-1}\sqrt{\pi}\,\Gamma[H]\sec[(K-H)\pi]}{\Gamma(3/2+K-H]\Gamma[2H-1-K]} \quad \textbf{(2.43)}$$

Using the Stirling formula and the reflection formula for the Gamma function, one can show that for large K, the permanent impact scales as:

$$E[m_\infty] - m_0 \sim \epsilon\theta \frac{N}{K^\beta} \quad \textbf{(2.44)}$$

If K is infinite, $E[m_\infty] - m_0 = 0$; that is, the impact is completely temporary. This can be shown in the mathematically equivalent propagator model Bouchaud et al. (2004, 2006). For a FARIMA forecast model with finite K (or equivalently, if the sign autocorrelation function decays fast beyond time scale K), the permanent impact is nonzero and is linear in N. Even if for large K the permanent impact is small, the convergence to zero with the memory K is very slow.

Another interesting issue that can be discussed within the model is the decay of the impact immediately after the end of the hidden order (defined by Eq. 2.37). One finds that the initial drop for $t \ll N$ is in fact very sharp for $\beta < 1$: $m_{N+t} - m_N \propto -t^{1-\beta}$, such that the slope of the decay is infinite when $t \to 0$ (in the continuous limit).

2.6.8. Aggregated Impact

Impact is often measured not on a trade-by-trade level but rather on a coarse-grained time scale, say five minutes or a day. One then speaks of positive correlations between signed order flow and price returns. At the level of single trades, impact is strongly concave in volume and decays in time. How does this translate at a coarse-grained level? In Section 2.5.2 we have discussed this from an empirical point of view. Here we show how the impact theories we have developed so far make predictions about the impact function, following the approach of Lillo et al. (2008a).

Suppose one aggregates the returns and volumes of N consecutive trades (not necessarily from the same hidden order). Using the same notation as in Section 2.5.2, the total volume imbalance is $Q_N = \sum_{n=0}^{N-1} \epsilon_n v_n$. Conditioned to a particular value $Q_N = Q$, what is the average price return $R(Q)$? The answer to this question depends on the order flow and the properties of the impact function. In the following discussion, we consider two extreme cases. In the first case we consider an unrealistic model where the order flow is described by an independent identically distributed random process, and the impact is fixed and permanent. In the second case we consider a correlated order flow and a fixed/temporary impact model.

Independent Identically Distributed Order Flow

If the unconditional distribution of market order volume and the functional form of the impact function are known, it is possible to find a closed expression for the impact $R(Q)$. Consider a series of N transactions with signed[17] volumes v_i corresponding to total return $R = \sum_{i=1}^{N} r_i$ and total signed volume $Q = \sum_{i=1}^{N} v_i$. The expected return given Q can be written:

$$R(Q, N) \equiv E[R|Q] = \int RP(R|Q, N)\, dR$$
$$= \frac{1}{P_N(Q)} \int RP(R, Q, N)\, dR \tag{2.45}$$

where $P_N(Q)$ is the probability density for Q. We assume that the N individual price impacts r_i due to the IID signed volumes v_i are given by a deterministic function[18] $r_i = f(v_i)$. Let the distribution of individual v_i be $p(v_i)$. Then the joint distribution of v_i is $P(v_1, \ldots, v_N) = p(v_1) \ldots p(v_N)$. The integral above becomes:

$$\int RP(R, Q, N)\, dR = \int dv_1 \ldots dv_N p(v_1) \ldots p(v_N) \sum_{i=1}^{N} f(v_i)\delta(Q - \sum_{i=1}^{N} v_i) \tag{2.46}$$

where we introduced the Dirac delta function.

By making use of the integral representation of the Dirac delta function, after some manipulations it is possible to rewrite $R(Q, N)$ as:

$$R(Q, N) = \frac{N}{2\pi} \frac{1}{P_N(Q)} \int d\lambda e^{(N-1)h(\lambda)} g(\lambda) e^{-i\lambda Q} \tag{2.47}$$

where $h(\lambda)$ is the logarithm of the Fourier transform of the volume distribution and $g(\lambda)$ is the Fourier transform of the product of the volume distribution and the impact function. Moreover, $P_N(Q)$ is the probability density that the total signed volume in the N trades is Q.

The functional form of the aggregate impact $R(Q, N)$ can be calculated by integrating this expression. It is possible to show that many of the properties of the solution are robust, independent of the details of the model. For small values of Q the aggregate impact $R(Q)$ is always linear with a slope that depends on N and on the details of the volume distribution and the impact function. For example, if the impact function is a power-law function $\epsilon|v|^{\psi}$ and the volume distribution decays asymptotically as $P(V) \sim V^{-\alpha-1}$, for large N the aggregate impact behaves for small Q as:

$$R(Q, N) \sim \frac{Q}{N^{\kappa}} \tag{2.48}$$

where κ depends in a nontrivial way on α and ψ (see Lillo et al., 2008a). For example, if volumes have a finite second moment and the impact function is concave, then

[17] Only in this subsection we indicate with v_i the signed and not the absolute value of volume.
[18] The results remain the same if a noise term is added to the impact function.

$\kappa = 0$; in contrast, if the second moment of the volume does not exist and the impact function is sufficiently concave, then $\kappa > 0$. The latter case agrees with what is seen in Figure 2.5, where the slope of the aggregate impact decreases with N. Thus theories for the aggregate impact make falsifiable predictions connecting volumes, order flow, and impact.

Transient Impact Model

Within the model of Section 2.6.4, the aggregate impact reads:

$$R(Q,N) = \sum_{n=0}^{N-1} G_0(N-n)E[q_n|Q] + \sum_{m<0}[G_0(N-m) - G_0(-m)]E[q_m|Q] \tag{2.49}$$

where $q_n = \epsilon_n \ln v_n$, and we assume that volumes are lognormally distributed (see Appendix 2.2). Because trades are long ranged correlated, the second term is nonzero. But one can show it is subdominant when $N \gg 1$, so we discard it in a first approximation. In the first term, one can compute $E[q_n|Q] = x$ using the fact that the q_ns are, within the model, Gaussian with $rms = s$. Noting also that typical values of Q are of order $N^{1-\gamma/2} \ll N$, one finds:

$$x \approx \frac{sQ}{\mathcal{I}N}, \quad \mathcal{I} = 2\int_0^\infty \mathrm{d}u\, u\, e^{us-u^2/2} \tag{2.50}$$

With $R(Q,N) \approx \Gamma_0 N^{1-\beta} x/(1-\beta)$ and the previous relation between $\beta = (1-\gamma)/2$, we finally find the following result, written in a suggestive scaling form:

$$R(Q,N) = \sqrt{N}\frac{s\Gamma_0}{\mathcal{I}(1-\beta)}\left(\frac{Q}{N^{1-\gamma/2}}\right) \tag{2.51}$$

This means that by rescaling the return and the signed volume by their respective root mean square value, one obtains at large N a limiting curve that is a straight line. Whereas for small N impact is strongly concave, impact becomes linear when $N \gg 1$. One can go one step further and compute the leading nonlinear correction in Q when N is large. One finds that it is negative, as a remnant of the small N concavity, and becomes noticeable at increasingly larger values of $Q \sim N$, as seen in empirical data; see Figure 2.5.

The important conclusion of this model is that although the impact of individual trades is concave and decays in time, the compensating effect of correlated trades leads to a well-defined *linear relation* between order imbalance and returns at an aggregated level. This is important because such a relation is often interpreted as a manifestation of the permanent component of the impact.

Is this linear relation telling us that part of the trades have indeed correctly *predicted* the aggregated return (in Hasbrouck's words); see Hasbrouck (2007)? In light of all the previous results, it looks much more plausible to us that anonymous trades in fact statistically *induce* price changes, although in a quite nontrivial and perhaps unexpected fashion.

2.7. THE DETERMINANTS OF THE BID–ASK SPREAD

In modern electronic markets, liquidity is *self-organized* in the sense that any agent can choose, at any instant of time, to either provide liquidity or consume liquidity. The liquidity of the market is partially characterized by the bid–ask spread S, which sets the cost of an instantaneous round trip of one share (a buy instantaneously followed by a sell, or vice versa).[19] A liquid market is such that this cost is small. A question of both theoretical and practical crucial importance is to know what fixes the magnitude of the spread in the self-organized setup of electronic markets and the relative merit of limit vs. market orders. In the economics literature (O'Hara, 1995; Biais et al., 1997; Madhavan, 2000; Glosten and Milgrom, 1985), the existence of the bid–ask spread is often attributed to three types of liquidity providing costs (Stoll, 1978):

- Order processing costs (this includes the profit of the market maker)
- Adverse selection costs—liquidity takers may have superior information on the future price of the stock, in which case the market maker loses money
- Inventory risk—market makers may temporarily accumulate large long or short positions that are risky; if agents are risk sensitive and have to limit their exposure, this may add extra costs

A somewhat surprising conclusion of early econometric studies is that order processing costs account for a large fraction of the spread. This may make sense in illiquid markets where market makers exploit a monopolistic situation to open large spreads but cannot be the correct picture in highly liquid, electronic markets in which market making is highly competitive. What we argue is that the main determinant of the spread is in fact impact.

2.7.1. The Basic Economics of Spread and Impact

What are the basic economics behind a trade—that is, the encounter between a liquidity taker and one (or several) liquidity provider(s)?

The Average Gain of Market Makers

Consider the sequence of all trades (not necessarily coming from the same hidden order). Let the n^{th} trade have volume v_n and sign ϵ_n. The profit collectively made by

[19] Other determinants of liquidity discussed in the literature are the depth of the order book and market resiliency, see Black (1971), Kyle (1985).

liquidity providers on that given trade, marked to market at time $n + \ell$, is given by:

$$\mathcal{G}_L(n, n+\ell) = v_n \epsilon_n \left[(m_n + \epsilon_n \frac{S_n}{2}) - m_{n+\ell} \right] \quad \textbf{(2.52)}$$

where S_n is the value of the spread at that moment in time. Think of a buy trade $\epsilon_n = +1$. This equation compares the money received by the liquidity provider when the trade occurs ($v_n(m_n + \frac{S_n}{2})$) to its mark-to-market (midpoint) price at time $n + \ell$. Symmetrically, the profit made by the liquidity taker using market orders is $\mathcal{G}_L(n, n+\ell) = -\mathcal{G}_M(n, n+\ell)$. This equation clearly shows that the profitability of market making comes from the spread ($+S_n/2$), whereas the losses are induced by market impact ($-\epsilon_n(m_{n+\ell} - m_n)$), which may or may not come from more informed traders (see the following discussion).

Neglecting for simplicity volume fluctuations at this stage ($v_n \equiv v$) and using Eq. 2.13, we see that the average gain of the market maker in the absence of extra costs is given by:

$$E[\mathcal{G}_L](\ell) = v \left(E[\frac{S}{2}] - \mathcal{R}_\ell \right) \quad \textbf{(2.53)}$$

which shows explicitly that for a given total market impact $\mathcal{R}_\ell$, the spread S should be larger than a minimum value for market-making strategies to be at all profitable on a time scale ℓ—or else, for a given value of S, the impact function $\mathcal{R}_\ell$ should be as small as possible. We recover here the idea that it is in the interest of liquidity providers to control the growth of $\mathcal{R}_\ell$ by tuning the liquidity asymmetry.

In fact, this reasoning neglects the cost of unwinding the market-maker position, and a better estimate will be provided later in this article. But the main message of the preceding simple computation is that the spread compensates for the impact of market orders. In the microstructure literature, this is refered to as *adverse selection*; as alluded to previously, this implies that market orders originate from better informed traders, with an information on the future price on average worth $\mathcal{R}_\ell$. But the same result would hold if impact was purely statistical, with no information content whatsoever. In fact, one could even revert the logic and claim that it is the spread that determines the impact: If *some* traders accepted to pay $m_n + S_n/2$ for the stock, it is natural that the market as a whole revises its fair-price estimate from m_n to $m_n + \alpha S_n/2$, where $\alpha \geq 0$ is a number measuring how trades influence the participants beliefs, leading to $\mathcal{R}_\infty = \alpha S/2$. The MRR model with spread (see The MRR Model with a Bid–Ask Spread section), in this context, assumes that market participants believe that the last traded price is indeed the correct price ($\alpha = 1$). Clearly, in that model, the cost of a market order or the gain of a limit order are exactly zero. This leaves us, by the way, in the familiar but uncomfortable situation of the "no trade theorem": If the spread is such that the information content of a market order is compensated, why would the informed trader trade at all?

How Informed Are the Trades?

So, are some market orders informed? Can one find convincing ex-post signatures of informed trades? A minimal definition of an informed trade is a trade that earns a profit significantly larger than the transaction costs (including both brokerage fees and market slippage). Introducing the signed return $r(n, n+\ell) \equiv \epsilon_n(m_{n+\ell} - m_n)$, the profit of the n^{th} market order on time scale ℓ is:

$$\mathcal{G}_M(n, n+\ell) = v_n \left[r(n, n+\ell) - \frac{S_n}{2} \right] \qquad \textbf{(2.54)}$$

Note that by definition the average of $r(n, n+\ell)$ is equal to the total impact $\mathcal{R}_\ell$, which is positive. If one averages this equation over *all* trades, one in fact finds that $E[\mathcal{G}_M]$ is close to zero, which means that the spread compensates for the average impact, at least measured on short time scales ℓ (between a few seconds to a few days). More precisely, on liquid NYSE stocks in 2005 (when market makers were still present), one finds that $E[\mathcal{G}_M]$ is zero within error bars, which means that, after transaction costs, market orders lose money, on average. The situation is slightly better for liquid PSE stocks in 2002, where one finds $E[\mathcal{G}_M] = gE(S)/2$ with $g \approx 0.3$ (see Figure 2.14 later). This amounts to 3 to 5 bp per trade, close to the transaction costs. So, on average, and although market orders do impact prices, there does not seem to be much *short-term* information in these orders, at least judging from their ex-post profitability. The question of longer-term information is, of course, left open here, simply because the statistics are not sufficient to judge the average profitability of trades on long time scales and because long-term drift effects cannot be neglected. (On average, buy trades are profitable in the long run!)

We can look in more detail at the full distribution of $r_\ell \equiv r(n, n+\ell)$, $P(r_\ell)$, which contains much more information. Note that its second moment $E[r^2|\ell]$ is very close to the volatility on scale ℓ, which soon becomes much larger than $\mathcal{R}^2$ when ℓ increases. Concerning the shape of $P(r_\ell)$, two extreme scenarios could occur (see Figure 2.11 for an illustration):

- A small proportion of well-informed trades *predict* the future price while a majority of trades are uninformed and do not impact the price at all. The distribution of r_ℓ should then be composed of a broad blob, symmetric around $r_\ell = 0$, corresponding to uninformed trades, plus a hump (or more plausibly, a broad shoulder) on the positive side, corresponding to well-informed trades. The nonzero value of $E[r_\ell]$ comes from these informed trades. This is the scenario behind, for example, the Kyle model or the Glosten-Milgrom model.
- All trades are equally weakly informed or even not informed, but *all* statistically impact prices. In this case one expects a symmetric broad blob but around the average impact $E[r_\ell]$.

Empirically, the distribution of r_ℓ is found to be very close to the second picture for ℓ corresponding to intraday time scales. In particular, no noticeable asymmetry

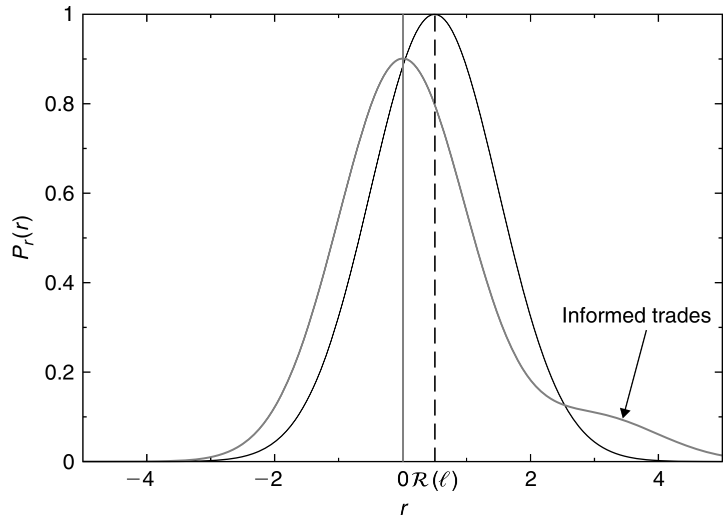

FIGURE 2.11 Two extreme cases for the distribution $P(r_\ell)$ of signed returns r_ℓ. Thick curve means nearly all trades are uninformed but impact prices, leading to a symmetric $P(r_\ell)$ around a nonzero average impact. Narrow curve means most trades are uninformed and do not impact prices, while some trades are informed and predict correctly the future return, leading to a thick tail in the $r_\ell > 0$ region.

(beyond the existence of a nonzero value of $E[r_\ell]$) is observed on liquid stocks; see Figure 2.12 for an example. This suggests that trades, on average, impact prices but do not seem to "predict" future prices—at least not on short time scales. The strong relation between order imbalance and price returns would then be a tautological consequence of this impact (see Section 2.6) and not a signature of "true" information revelation.

2.7.2. Models for the Bid–Ask Spread

The Glosten-Milgrom Model

One of the earliest theories of the spread that makes the preceding discussion is the sequential trade model of Glosten and Milgrom (Glosten and Milgrom, 1985). One assumes that market orders are either due (with some probability q) to informed traders, who know the end-of-day price p_f, or (with probability $1-q$) to noise traders. The value of q is assumed to be known by the market maker, which is not necessarily very realistic (a similar assumption is made within the Kyle model). The end-of-day price p_f can either be above ($p_>$) or below ($p_<$) the open price. The probabilities for either outcome at the start of the day are $\delta_+ = \delta_- = 1/2$ for simplicity. But as trading occurs, either at the bid or at the ask, the market maker updates in a Bayesian way the value of $\delta_+ = 1 - \delta_-$: trades at the ask increase the value of δ_+, whereas trades at the bid increase δ_-.

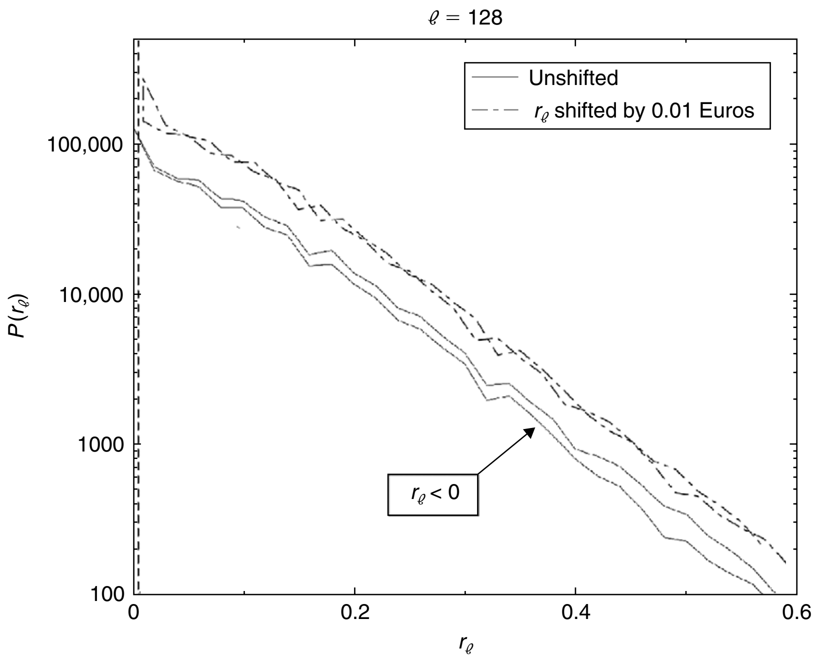

FIGURE 2.12 Probability distribution $P(r_\ell)$ of the quantity $r = (m_{n+\ell} - m_n).\varepsilon_n$ (in Euros) for $\ell = 128$; data are again for France Telecom during 2002. The negative part of the distribution has been folded back to positive r to highlight the small positive assymetry of the distribution. The average value $\mathcal{R}_\ell = E[r] \approx 0.01$ is shown by the vertical dashed line. The dashed-dotted line corresponds to the distribution of $r - 0.01$, for which no asymmetry of the type shown in Figure 2.11 can be detected. This curve has been shifted upwards for clarity.

This leads to a certain update rule for δ_+ as a function of the sign of the next trade, which we do not write here explicitly. Anticipating the value of $\delta_\pm$ after the next trade allows the market maker to position his quotes in such a way as not to have *ex-post* regrets. More precisely:

$$a = \delta_+(+)p_> + \delta_-(+)p_<, \quad b = \delta_+(-)p_> + \delta_-(-)p_< \tag{2.55}$$

where $(\pm)$ refers to the sign of the next trade. This leads to the following prediction for the bid–ask S_n after the n^{th} trade:

$$S_n = 4q\delta_+^{(n)}\delta_-^{(n)}(p_> - p_<) \tag{2.56}$$

where $\delta_\pm^{(n)}$ is the updated value of $\delta_\pm$ after n trades (with $\delta_\pm^{(0)} = 1/2$), and we have neglected terms of order q^2, which must be small if this model is to be realistic. This model is by construction compatible with a random walk for the midpoint, with a volatility per trade σ_1 proportional to the bid–ask spread, as reported later in this chapter. It also predicts that the bid–ask spread declines on average throughout the day, since the update rule drives δ_+ either to zero or to one: As trading occurs, the market maker discovers more accurately which outcome is more likely.

A detailed comparison of this model with empirical data is given in Wiesinger et al. (2008). Here we simply note that as far as order of magnitude goes, the spread at the beginning of the day (when $\delta_\pm = 1/2$) is typically 0.1%, whereas the daily volatility fixes the order of magnitude of $p_> - p_<$ to typically 2%, leading to $q \sim 0.05$. Within this framework, one finds again that the fraction of short-time "informed trades" must be small. One also finds that in this model the spread decays exponentially fast with time, at variance with the slow, power-law relaxation that has been observed (see Section 2.7.4).

The MRR Model with a Bid–Ask Spread

The original MRR model is in fact slightly different from the model described in Section 2.6.3. MRR model rather assumes that it is the "true" fundamental price p_n rather than the midpoint m_n, which is impacted by the surprise in order flow, and hence:

$$p_{n+1} - p_n = \eta_n + \theta[\epsilon_n - \rho\epsilon_{n-1}] \tag{2.57}$$

MRR then specifies a rule for the bid and ask price, which in turn allows one to *compute* the midpoint m_n. Since market makers cannot guess the surprise of the next trade, they post a bid price b_n and an ask price a_n given by:

$$a_n = p_n + \theta[1 - \rho\epsilon_{n-1}] + \phi, \quad b_n = p_n + \theta[-1 - \rho\epsilon_{n-1}] - \phi \tag{2.58}$$

where ϕ is the extra compensation claimed by the market maker, covering processing costs and the shock component risk. This rule ensures no *ex-post* regrets for the market maker: Whatever the sign of the trade, the traded price is always the "right" one. The midpoint $m \equiv (a + b)/2$ immediately before the n^{th} trade is now given by:

$$m_n = p_n - \theta\rho\epsilon_{n-1} \tag{2.59}$$

whereas the spread is given by $S = a - b = 2(\theta + \phi)$.

More generally, assuming that only the sign surprise matters, one can write, for arbitrary correlations between signs:

$$m_{n+\ell} - m_n = \sum_{j=n}^{n+\ell-1} \eta_n + \theta \sum_{j=n}^{n+\ell-1} \left\{\epsilon_j - E_j\left[\epsilon_{j+1}\right]\right\} \tag{2.60}$$

where the last term is the conditional expectation of the next sign. In the Markovian case, $E_j[\epsilon_{j+1}] = \rho\epsilon_j$, and we recover the previous result. The impact function, in the general case, reads:

$$\mathcal{R}_\ell = \theta\,[1 - C_\ell] \tag{2.61}$$

Using Eq. 2.53, one sees that the long-term profit of market makers is zero. However, due to correlations between trades, the longtime impact is enhanced compared to the short-term impact by a factor:

$$\lambda = \frac{1}{1 - C_1} > 1 \tag{2.62}$$

As we've discussed very generally, spread and impact are two sides of the same coin. This is particularly clear within the MRR model, where the half-spread $S/2$ is set to be equal to the long-term impact $\mathcal{R}_\infty = \theta$. This means that the profit of market makers is exactly zero (provided $\phi = 0$), but also, as noted previously, that the profit of putatively informed market orders is zero. The spread in the MRR model is

$$S = 2(\theta + \phi) = 2(\mathcal{R}_\infty + \phi) = 2\lambda\mathcal{R}_1 + 2\phi \tag{2.63}$$

where $\lambda = (1 - \rho)^{-1}$. Appendix 2.3 provides an alternative, enlightening derivation.

One computes the midpoint volatility on scale ℓ, defined as

$$\sigma_\ell^2 = \frac{1}{\ell}\langle(m_{\ell+i} - m_i)^2\rangle \tag{2.64}$$

One finds a sum of a trade-induced volatility $\theta^2(1-\rho)^2$ and a "news"-induced volatility Σ^2:

$$\sigma_1^2 = \langle(m_{n+1} - m_n)^2\rangle = \Sigma^2 + \theta^2(1-\rho)^2 \tag{2.65}$$

and

$$\sigma_\infty^2 = \Sigma^2 + \theta^2(1-\rho)^2(1 + 2\frac{\rho}{1-\rho}) = \Sigma^2 + \theta^2(1-\rho^2) \geq \sigma_1^2 \tag{2.66}$$

The MRR model therefore leads to two simple relations among spread, impact, and volatility per trade

$$S = 2\lambda\mathcal{R}_1 + 2\phi \quad \sigma_1^2 = \mathcal{R}_1^2 + \Sigma^2 \tag{2.67}$$

where $\lambda = (1 - \rho)^{-1}$ and ϕ is any extra compensation claimed by market makers. These relations are generalized to more realistic assumptions and tested empirically in the next two sections.

2.7.3. Limit vs. Market Orders: The Microstructure Phase Diagram

Market Order Strategies

As we have mentioned, the gain (or cost) of a given market order can be defined as $v_n[r(n, n+\ell) - \frac{S_n}{2}]$. This definition in fact marks the trade to market after ℓ trades and is often referred to as the *realized spread* (Bessembinder, 2003; Stoll, 2000). The volume-weighted averaged gain (over a large number of trades) of market orders over a long horizon $\ell \gg 1$ is therefore:[20]

$$E[\mathcal{G}_M] \approx \lambda \frac{E[v\mathcal{R}_1(v)]}{E[v]} - \frac{E[vS]}{2E[v]} \tag{2.68}$$

In this expression we have introduced the volume-dependent lagged impact function

$$\mathcal{R}_\ell(v) = E[\epsilon_n(m_{n+\ell} - m_n)|v_n = v] \tag{2.69}$$

and we have used the previous definition of the amplification factor λ: $\mathcal{R}_{\ell \gg 1} = \lambda \mathcal{R}_1$. In the plane $x = E[v\mathcal{R}_1(v)]/E[v]$, $y = E[vS]/E[v]$ (which will repeatedly be used later), the condition $E[\mathcal{G}_M] = 0$ defines a straight line of slope 2λ separating an upper region where market orders are on average costly from a region where single market orders are favored: see Line a in Figure 2.13. For large spreads, the positive average cost of the market orders would deter their use; limit orders would then pile up and reduce the spread.

Below Line a of slope 2λ, market orders have a negative cost, and one might be able to devise profitable strategies based solely on market orders. The idea would be to try to benefit from the impact term $\mathcal{R}_\infty$ in the previous balance equation. The growth of $\mathcal{R}_\ell$ ultimately comes from the correlation between trades—that is, the succession of buy (sell) trades that typically follow a given buy (sell) market order. The simplest "copycat" strategy one can rigorously test on empirical data is to imagine placing a market order with vanishing volume fraction (so as not to affect the subsequent history of quotes and trades), immediately following another market order. This strategy suffers on average from the impact of the initial trade, used as a guide to guess the direction of the market. Therefore, the profit $\mathcal{G}_{CC}$ of such a copycat strategy, marked to market after a long time and neglecting further unwinding costs, is reduced to:

$$\mathcal{G}_{CC} = [\lambda - 1]\frac{E[v\mathcal{R}_1(v)]}{E[v]} - \frac{E[vS]}{2E[v]} \tag{2.70}$$

By requiring that this gain is nonpositive, one obtains a lower line in the plane x, y, of slope $2(\lambda - 1)$. Only below Line d can the preceding infinitesimal copycat strategy be profitable. We therefore expect markets to operate above this line and below Line a of slope 2λ.

Note also that the longtime impact of an isolated market order, uncorrelated with the order flow, is given by $G_0(\ell \gg 1)$, which is small (see Section 2.6.2). These isolated market orders thus also have a positive cost equal to half the spread. The only way to benefit from the average impact $\mathcal{R}_\ell$ is to free-ride on a wave of orders launched by others, as in the copycat strategy. Let us now take the complementary point of view of limit orders and determine the region of profitable market-making strategies.

[20]Note that this definition neglects the fact that one single large market order may trigger transactions at several different prices, up the order book ladder, and pay more than the nominal spread. Nevertheless, this situation is empirically quite rare in the markets we are concerned with and corresponds to only a few percent of all cases; see Farmer et al. (2004).

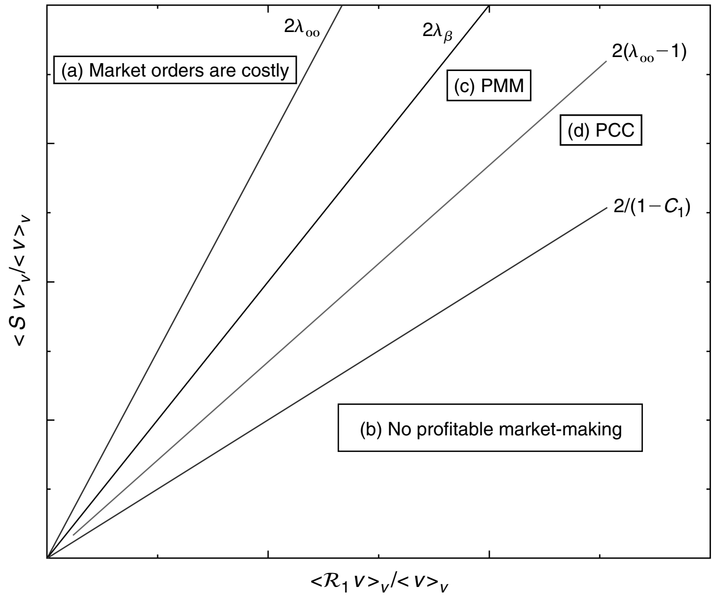

FIGURE 2.13 General "phase diagram" in the plane $x = E[v\mathcal{R}_1(v)]/E[v]$, $y = E[vS]/E[v]$, showing several regions: **(a)** above the line of slope 2λ, market orders are costly (on average) and market-making is profitable; **(b)** below the line of slope $\approx 2/(1 - C_1)$, limit orders are costly and no market-making strategy is profitable; **(c)** above the thick line of slope $2\overline{\lambda}_\beta$, market-making on time scale β^{-1} (or faster) is profitable (PMM); **(d)** below the fine line of slope $2(\lambda - 1)$, copycat strategies can be profitable (PCC). Since neither market orders nor liquidity providing should be systematically penalized to ensure steady trading, we expect that markets should operate in the "neutral wedge" in between the (b) and (a) lines. Competition between liquidity providers should push the market toward the (b). Since copycat strategies should not be profitable either, the PCC (d) line cannot lie above (b). Note that the (b), (a), and (c) all coincide within the MRR model.

An Infinitesimal Market-Making Strategy

We now compute the gain of a simple market-making strategy, which amounts to participating in a vanishing fraction of all trades through limit orders. The simplest strategy is to consider a market maker with a certain time horizon who provides an infinitesimal fraction φ of the total available liquidity. As illustrated by Eq. 2.68, the cost incurred by the market maker comes from market impact: The price move is anticorrelated with the accumulated position. When the crowd buys, the price goes up while the market-making strategy accumulates a short position, which would be costly to buy back later, and vice versa.

We consider a *steady-state* market-making strategy that avoids explicit unwinding costs. The strategy is such that tendered volume dynamically depends on the accumulated position, which ensures that the inventory is always bounded. We choose the tendered fraction φ to be given by $\varphi_i = \varphi_0(1 + \alpha V_i \epsilon)$, where V_i is the (signed) position accumulated up to time i^-, and $\epsilon = +1$ for orders placed at the ask and $\epsilon = -1$ for orders placed at the bid. This mean-reverting strategy ensures that the typical position is always bounded. One can now use this strategy for an arbitrary long time T; its profit and loss is simply given by

$$\mathcal{G}_L = \sum_{i=0}^{T-1} \varphi_i \epsilon_i v_i \left(m_i + \epsilon_i \frac{S_i}{2} \right) \tag{2.71}$$

For large T one can replace this expression by its average:

$$\mathcal{G}_L = TE\left[\varphi_i \epsilon_i v_i \left(m_i + \epsilon_i \frac{S_i}{2} \right) \right] \tag{2.72}$$

with $O(T^0)$ corrections due to the residual position at T. This quantity has been computed in Wyart et al. (2008) and depends on the value of $\beta = 1 - \alpha\varphi_0 E[v]$ that fixes the typical time scale $\ell^* = (1-\beta)^{-1}$ of the market-making strategy. When $\beta \to 0$ (fast market-making), the gain per unit time and unit volume reduces to

$$\frac{\mathcal{G}_L(\beta \to 0)}{T\varphi_0 E[v]} \approx \frac{E[vS]}{2E[v]}[1 - C_1] - \frac{E[v\mathcal{R}_1(v)]}{E[v]} \tag{2.73}$$

whereas $\beta \to 1$, corresponding to slow market making, yields:

$$\frac{\mathcal{G}_L(\beta \to 1)}{T\varphi_0 E[v]} = \frac{E[vS]}{2E[v]} - \frac{E[v\mathcal{R}_1(v)]}{E[v]} \tag{2.74}$$

The competition between impact and spread is more favorable to limit orders when the strategy is fast ($\beta = 0$) than when it is slow ($\beta = 1$). Imposing that there is a certain frequency β such that the gain of market-making strategies is zero leads to a linear relation between spread and impact, generalizing the prerious MRR relation Eq. 2.67:

$$\frac{E[vS]}{E[v]} = 2\lambda_\beta \frac{E[v\mathcal{R}_1(v)]}{E[v]} \tag{2.75}$$

Using the empirical shape of $\mathcal{R}_\ell$ and C_ℓ, the slope $2\lambda_\beta$ is found to increase between $\approx 2/(1-C_1)$ and 2λ when β increases from zero to one. When $\beta \to 1$, $\lambda_\beta \to \lambda$ and the lower limit of profitability of very slow market making is precisely Line a of Figure 2.13, where market orders become profitable. Faster strategies correspond to smaller values of λ_β, closer to $1/(1-C_1)$, leading to an extended region of profitability for market making.

From the assumption that the preceding market-making strategy for any value of β should be at best marginally profitable (since one might find more sophisticated strategies that take full advantage of the correlations between signs and volumes), we finally obtain the following bound between spread and impact:

$$\frac{E[vS]}{E[v]} \leq \frac{2}{1-C_1} \frac{E[v\mathcal{R}_1(v)]}{E[v]} \tag{2.76}$$

defining Line b of slope $2/(1-C_1)$ in the x, y plane of Figure 2.13. Consistent with the MRR model, when $\lambda = 1/(1-C_1)$, the Lines a and b of Figure 2.13 exactly coincide. Using that fact that $\mathcal{R}_1^{n+} \leq \mathcal{R}_1^{(n-1)+}$, a simple generalization of the argument presented in Appendix 2.3 allows one to show directly that the cost of limit orders is indeed negative above Line b.

Eqs. 2.68 and 2.76 and the resulting microstructural “phase diagram” of Figure 2.13 are the central results of this section. The preceding analysis delineates, in the impact-spread plane, a central wedge bounded from above by a slope 2λ and from below by a slope $\approx 2/(1-C_1)$, within which both market orders and limit orders are viable. In the upper wedge, market orders would always be costly and would be substituted by limit orders. In the lower wedge, market-making strategies, even at high frequencies, would never eke out any profit. Such a market would not be sustainable in the absence of any incentive to provide liquidity. But if the spread happened to fall in this region, the enhanced flow of market orders would soon reopen the gap between bid and ask. In the MRR model, this wedge reduces to a single line.

Comparison with Empirical Data

In conclusion of the preceding theoretical section, one expects electronic markets to operate in the vicinity of Line b of Figure 2.13, that is, there should be a linear relation between spread and market impact with a slope close to $2/(1-C_1)$. This prediction has been tested on empirical data in Wyart et al. (2008), where different markets were considered. The prediction can be tested in two different ways: for a given stock across time and across all different stocks. In both cases, a rather convincing agreement with the theory is obtained. We show, for example, in Figure 2.14 the cross-sectional test of Eq. 2.75 over 68 different stocks of the PSE in 2002. The relative values of the spread and the average impact vary by a factor of five between the various stocks, which makes it possible to test the linear relations Eqs. 2.70 and 2.76. A linear fit with zero intercept gives a slope of 2.86,[21] while the average of $2/(1-C_1)$ over all stocks is found to be ≈ 2.64.

However, the situation appears to be different on the NYSE, where specialists are present. Plotting the data corresponding to the 155 most actively traded stocks on the NYSE in 2005 in the spread-impact plane, one now finds that the empirical results cluster around the upper Line a limit where market orders become costly; see Figure 2.15. The regression has a significantly larger slope of 3.3, larger than $2/(1-C_1) \approx 2.78$, and a positive intercept $2\phi \approx 1.3$ basis points.[22] This suggests the existence of monopoly rents on NYSE; even if there is some competition to provide liquidity with other

[21]The intercept of a two-parameter regression is in fact found to be slightly negative.

[22]This is five times smaller than the average spread, leading to $\phi/\theta \sim 0.25$, much smaller than the result $\phi/\theta \sim 1-2$ found within the MRR model in 1990 or a similar value reported in Stoll (2000).

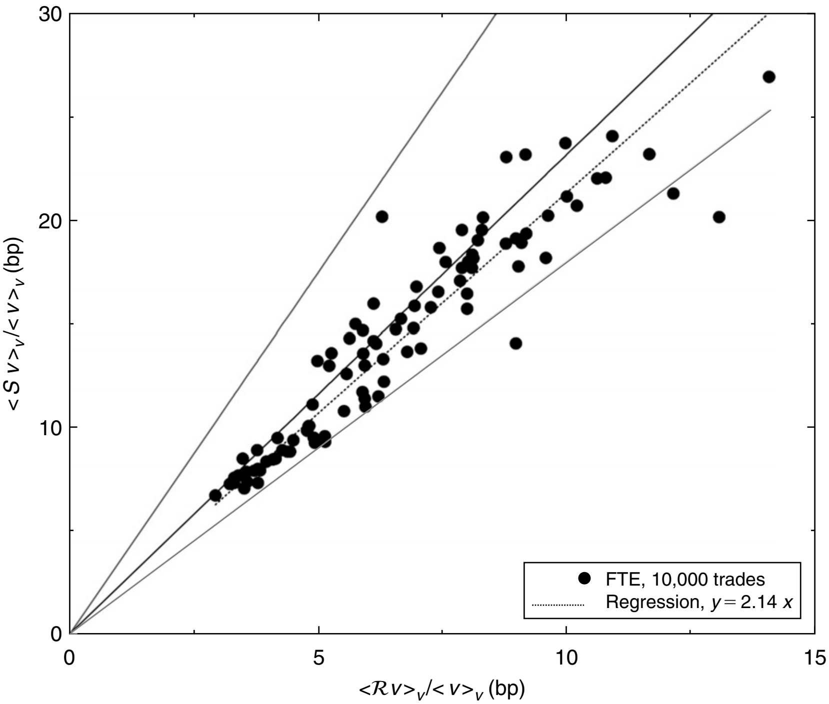

FIGURE 2.14 Plot for 68 stocks on the Paris Stock Exchange in 2002. Each point corresponds to a pair $(y = \langle vS \rangle / \langle v \rangle, x = \langle v\mathcal{R}_1 \rangle / \langle v \rangle)$, computed by averaging over the year. Both quantities are expressed in basis points. We also show the different bounds, from largest to smallest slope: Eqs. 2.68, 2.76, and 2.70, and a linear fit that gives a slope of 2.86, while $\langle 2/(1 - C_1) \rangle \approx 2.64$. The correlation is $R^2 = 0.90$.

market participants. Market makers post spreads that are systematically overestimated compared to the situation in electronic markets, with a nonzero extrapolated spread 2ϕ for zero market impact. This result is in agreement with older studies on the NYSE: Harris and Hasbrouck (1996) used data from the early 1990s to show that limit orders were more favorable than market orders; and Handa and Schwartz (1996) showed that pure limit order strategies were indeed profitable. We refer to Wyart et al. (2008) for more discussion.

The empirical analysis therefore shows that for liquid markets, an approximate symmetry between limit and market orders holds, in the sense that neither market orders nor limit orders are systematically unfavorable. Markets operate in the “neutral wedge” of Figure 2.13. In fully electronic markets, competition for providing liquidity is efficient in keeping the spread close to its lowest value. For markets with specialists, such as the NYSE, spreads appear to be significantly larger and market orders are now marginally costly on average.

Note that the preceding analysis does not require any model-specific assumptions such as the nature of order flow correlations or the fraction of informed trades. In fact,

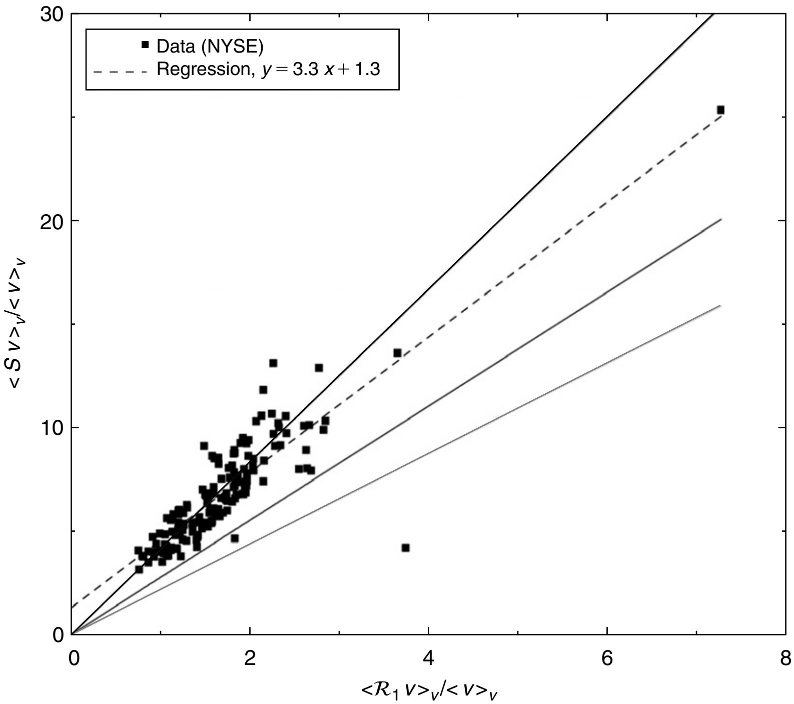

FIGURE 2.15 Plot of 155 stocks on the NYSE in 2005. Each point corresponds to a pair ($y = \langle vS\rangle/\langle v\rangle$, $x = \langle v\mathcal{R}_1\rangle/\langle v\rangle$), computed by averaging over the year. Both quantities are expressed in basis points; also shown are bounds, from largest to smallest slope: Eqs. 2.68, 2.76, and 2.70. Data clearly show that market orders are less favorable than in the electronic Paris Bourse. The regression now has a positive intercept of 1.3 bp with an $R^2 = 0.87$.

the preceding results hold even if trades are all uninformed but still mechanically impact the price.

2.7.4. Spread Dynamics After a Temporary Liquidity Crisis

The preceding analysis has shown the existence of relations between market impact and the unconditional value of the spread. The spread, however, is a variable with interesting temporal dynamics. Several studies have characterized the statistical properties of spread. Generally these studies have found that the spread distribution is fat tailed and the time correlation properties are consistent with a long-memory process (Plerou et al., 2005; Mike and Farmer, 2008; Gu, Chen, and Zhou, 2007).

It is also interesting to ask how the spread responds after a temporary liquidity crisis. As we describe in more detail in Section 2.8.1, even at the scale of individual transactions, price returns are heavy tailed; that is, it is not infrequent to observe individual transactions triggering large price changes. This often happens because a market order removes all the volume at the best, and the next-to-best occupied price level has a price very different from the price at the best (Farmer et al., 2004). As a consequence, even a small order can create a large price change, creating a very large spread. A large spread is what we mean here by a "temporary liquidity crisis."

We now describe the average dynamics followed by the spread as it converges to its "typical" value. First, a large spread is a strong incentive for limit orders inside the spread and a strong disincentive for market orders. Direct measurements of the order flow conditional on the spread value confirm this intuition (Mike and Farmer, 2008; Ponzi et al., 2008). The limit order flow inside the spread has a limit price distribution that is roughly independent of spread size and monotonically decreasing when one moves from the same best toward the opposite best. This suggests that the typical spread dynamics is not a fast reversion to its typical value, but rather it is a slow process where each liquidity provider competes with the others to close the spread. Each player tries to do this as slowly as possible to get a more favorable price from the incoming market orders, but at the same time competition prevents this process from being too slow. Empirically this slow decay has been measured in Zawadowski et al. (2006) and Ponzi et al. (2008). One way of quantifying the average dynamics is by computing the quantity, Ponzi et al. (2008),

$$G(\tau|\Delta) = E(S_{t+\tau}|S_t - S_{t-1} = \Delta) - E(S_t) \tag{2.77}$$

where S_t is the spread at time t (in seconds). This quantity is the expected value of the spread at time $t + \tau$ conditional to the fact that at time zero there is a spread change of size Δ. Figure 2.16 shows this quantity for the stock AZN traded at the LSE as a

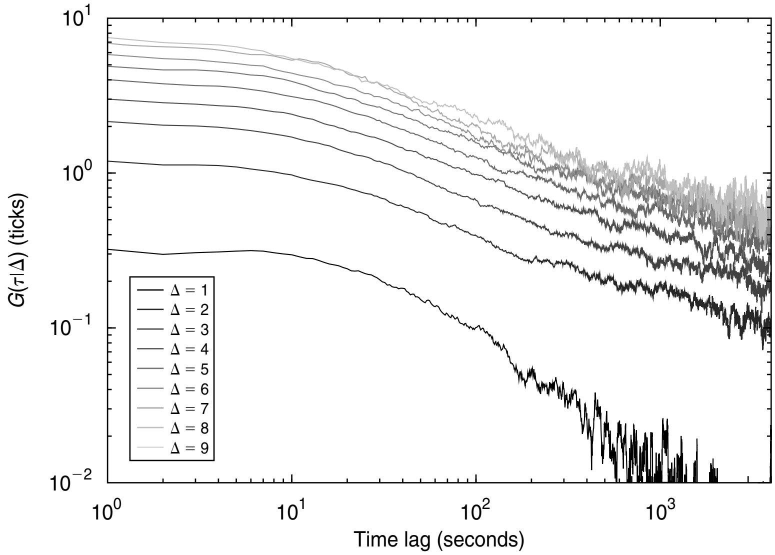

FIGURE 2.16 Conditional spread decay $G(\tau|\Delta)$ defined in Eq. 2.77 for the stock AZN, showing $G(\tau|\Delta)$ for different positive values of Δ (in ticks) corresponding to an opening of the spread at time lag $\tau = 0$. (*Source*: Adapted from Ponzi et al., 2008.)

function of τ for different positive and negative values of Δ. The decay of $G(\tau|\Delta)$ as a function of τ is very slow and for large values of τ is compatible with a power-law decay with a fitted exponent in the range 0.4–0.5. A similar slow decay of the volatility after a shock has been reported in Lillo and Mantegna (2003), Zawadowski et al. (2006), and Joulin et al. (2008).

2.8. LIQUIDITY AND VOLATILITY

One of the best-known statistical regularities of financial time series is the fact that the empirical distribution of asset price changes is heavy tailed; that is, there is a higher probability of extreme events than in a Gaussian distribution. This property has been verified by many authors on many different financial time series (e.g., Mandelbrot, 1963; Lux, 1996; Gopikrishnan et al., 1998). Extensive empirical analyses have shown that the distribution of price change over time intervals ranging from a few minutes to one or a few trading days is asymptotically distributed in a way that is approximately independent of the time interval size.

2.8.1. Liquidity and Large Price Changes

Many estimates indicate that the part of the distribution describing large price changes is a power law. For larger time intervals, the tail behavior of the return distribution becomes slowly consistent with a Gaussian tail in accordance with the central limit theorem. The heavy-tailed property of large price change is important for financial risk, since it means that large price fluctuations are much more common than one might expect under a Gaussian hypothesis.

There have been several conjectures about the origin of heavy tails in prices. Two theories that make testable hypotheses about the detailed underlying mechanism are the subordinated random process theory of Clark (1973) and the recent theory of Gabaix et al. (2003). The first model has its origins in a proposal of Mandelbrot and Taylor (1967) that was developed by Clark. Mandelbrot and Taylor proposed that prices could be modeled as a subordinated random process $Y(t) = X(\tau(t))$, where Y is the random process generating returns, X is Brownian motion, and $\tau(t)$ is a stochastic time clock whose increments are independent and identically distributed and uncorrelated with X. Clark hypothesized that the time clock $\tau(t)$ is the cumulative trading volume in time t. In simple terms, the subordination hypothesis states that price changes would be Gaussian if one measured them in equal intervals of volume (or number of trades) rather than in real time intervals.

Gabaix et al.'s proposal, in contrast, is that high-volume orders cause large price movements. They argue that the distribution of large trade size scales as $P(V > x) \sim x^{-\alpha}$, where v is the volume of the trade and $\alpha \approx 1.5$. Based on the assumption that agents maximize a first-order utility function, with a risk penalty term that is proportional to standard deviation rather than variance, they claim that the average market impact function has the form $\Delta p \propto V^{\psi}$, where $\psi \approx 0.5$. From this follows that large price changes have a power-law distribution with exponent $\alpha/\psi \approx 3$. For a critique of the empirical results and a rebuttal, see Farmer and Lillo (2004) and Plerou et al. (2004).

Both the Clark and Gabaix theories emphasize the role of trading volume as the determinant of large price changes. Even if it is clear that volume has some role in determining price changes, recent studies show that trading volume could not be the key factor. In a recent paper, Farmer et al. (2004) considered the distribution of returns generated by individual market orders. They showed that even at this microscopic time scale, price returns are heavily tailed, and more important, the size of price moves is essentially independent of the volume of the orders; see also Joulin et al. (2008). Both these facts seriously challenge the explanation of fat tails based on volume fluctuations. In that paper Farmer et al. showed that price returns associated with individual transactions are driven by liquidity fluctuations. The authors proposed and tested a mechanism for explaining how liquidity fluctuations determine large price changes. Even for the most liquid stocks in the London Stock Exchange, the limit order book often contains large gaps, corresponding to a block of adjacent price levels containing no quotes. When such a gap exists next to the best price, a new market order can remove the best quote and generate a large price change. At this time scale the distribution of large price changes merely reflects the distribution of gap sizes in the order book. The LSE data indicate that approximately 85% of the trades having a nonzero price impact have a volume equal to the volume at the best. Moreover, 97% of the trades having a nonzero price impact generate a price change equal to the first gap. In summary, the fluctuations of the gap sizes in the book are a key determinant of large price changes. The gap size is a measure of the liquidity available in the market as limit orders. Thus fluctuations of liquidity—that is, in the market's ability to absorb new market orders—are the origin of large price changes, whereas the trading volume plays a minor role.

The previously proposed mechanism raises the question of the importance of temporary liquidity crises, evidenced by large gaps in the book, for price changes over long time intervals. Although a definite answer is not available, there are three indications that short time scale and long time scale price fluctuations may be related. First, the gap size displays long-memory properties in time; see Lillo and Farmer (2005). This means that the gap size—that is, the liquidity availability—is strongly correlated in time. Periods when the typical gap size is large are likely to be followed by periods of large gaps; that is, liquidity availability is a persistent quantity. Second, it has been shown that the permanent component of the price impact is roughly proportional to the immediate impact caused by the trade (Ponzi et al., 2008). Thus the distribution of permanent price impacts, which is closely related to the distribution of price changes over relatively long time intervals, is approximately the same as the distribution of temporary price impacts, that is, of gaps in the order book. The third indication concerns the relative importance of volume and liquidity in explaining aggregate price changes, as discussed in more detail in the next section.

2.8.2. Volume vs. Liquidity Fluctuations as Proximate Causes of Volatility

The existence of a relation between volume and volatility has been known for a long time. This relation has been often interpreted as a causal relation, suggesting that volume (or number of transactions) is the driving factor determining volatility (Ane and Geman, 2000). In the previous section we discussed the subordination hypothesis, which states that returns would be Gaussian if measured in equal intervals of volume rather than in equal intervals of real time. The theory by Gabaix et al. (2003, 2006) reaches the same conclusion. Here we present some evidence challenging this view and indicating that liquidity fluctuations may be more important than volume in explaining volatility fluctuations. The question can be posed in terms of Eq. 2.1—that is, $\Delta p = \mathcal{T}(I)/\lambda$: Which is more important in determining the size of price movements, $\mathcal{T}(I)$ or λ?

In a recent paper, Gillemot et al. (2006) have presented evidence based on several different tests involving comparisons of long memory and regressions of the volatility in specific time intervals, showing that liquidity is a more important determinant of volume. Even when one aggregates returns over a fixed number of transactions (or volume), the return probability density function remains heavy tailed with properties very similar to those in fixed intervals of time. A simple way to see this effect is given in Figure 2.17, which shows the empirical probability $P(|r| > x)$ as a function of x for

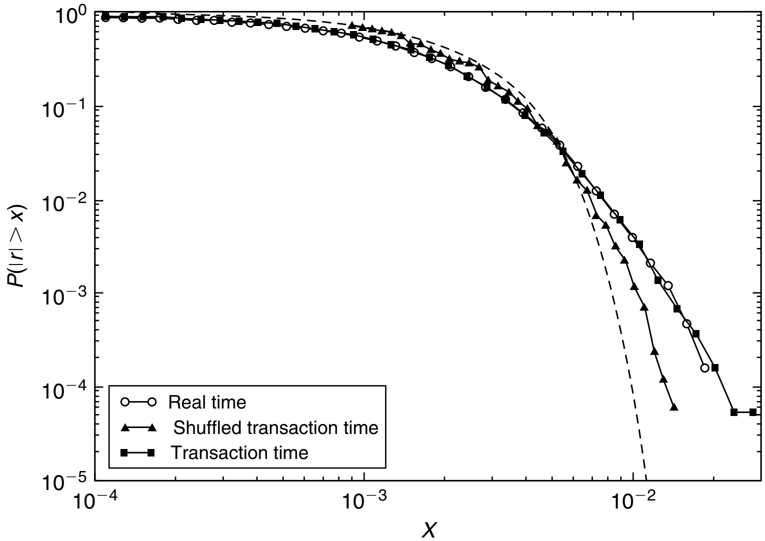

FIGURE 2.17 Cumulative distribution of absolute (log) returns $P(|r| > x)$ for the NYSE Procter & Gamble stock under different time clocks, plotted on double logarithmic scale. The circles refer to 15-minute returns, the squares refer to returns aggregated with a fixed number of transactions, and the triangles show the cumulative distribution obtained by randomly shuffling individual transaction returns and then aggregating them in a way that matches the number of transactions in each real-time interval. The dashed line corresponds to a normal distribution.

the NYSE stock Procter & Gamble. Here r is the price return over a 15-minute time interval. Suppose returns are measured in transaction time; that is, every 87 transactions rather than every 15 minutes, where 87 is chosen because it is the average number of transactions in 15 minutes (during the period from January 29, 2001, to December 31, 2003). The empirical distribution of transaction time returns matches that of real-time returns very well. Since in this case the number of transaction is held constant, this shows that the heavy tail of the return distribution is not due to variations in the number of transactions. The same effect is seen by aggregating transactions with volume rather than the number of transactions fixed (see Gillemot et al., 2006, for details).

This result shows that the fluctuation in number of trades or volume associated with a fluctuating trading activity is not the main determinant of the heavy tails of the return distribution. To highlight this effect, Figure 2.17 also shows the distribution of returns obtained from a surrogate distribution, constructed by randomly shuffling the returns of individual transactions and by aggregating them in a way that matches the number of transactions in each real-time interval. In doing so the unconditional distribution of returns of individual transactions is preserved, as are the fluctuation properties of trading frequency, but any temporal correlations of individual trade returns are destroyed. The figure shows that the tail of the surrogate distribution is less heavy than the real one, indicating that fluctuations and the time correlation properties of the reaction of prices to trades—that is, liquidity—are more important than fluctuations in trading frequency.

More supporting evidence for the importance of liquidity in determining volatility comes from a recent paper testing the microscopic random walk hypothesis against real data (Laspada et al., 2008). The price dynamics can be described as a random walk in which the increments are due to individual transactions. Under the assumption that the sign and the size of the price increments are mutually independent stochastic processes, it is possible to derive an exact expression for the volatility expected in a time interval with a given number of transactions. When one tests this expression on real data, it is found that for one-hour intervals the model consistently overpredicts the volatility of real price by about 70% and that this effect becomes stronger as the length of the time interval increases. This fact suggests that the assumption of independence of size and sign of price changes is wrong. However, data show that the contemporaneous correlation between size and sign of returns is nonstatistically significant. By performing a series of shuffling experiments, Laspada et al. (2008) show that the discrepancy between the volatility of the model and of the data is caused by a subtle but long-memory noncontemporaneous correlation between the signs and sizes of individual returns. Therefore, even after controlling for the number of transactions and the order imbalance in a given time interval, the random walk model has a strong bias in predicting the volatility, which is caused by the long memory of liquidity. This once again indicates that volume is not the key factor in explaining volatility. The neglected subtle relation between return signs and sizes shows that fluctuating liquidity is an important factor in explaining volatility.[23]

[23] In Section 2.6 we discuss how such a correlation is a consequence of the long memory of order flow and of market efficiency. The asymmetric liquidity models described in Section 2.6 predict a reduction of volatility relative to what one would expect under an unconditional permanent impact model.

Finally, the correlation between large volumes and large returns was directly studied in Joulin et al. (2008), both for trade-by-trade data and for one-minute bins, with the conclusion that such a correlation is totally absent from the data.

2.8.3. Spread vs. Volatility

It is worth investigating the relation between spread and volatility in the framework of the MRR model discussed previously. In fact, this model predicts a simple relation between volatility and impact, as shown in Eq. 2.67. Together with the relation between spread and impact we've discussed at length, this suggests a direct link between volatility per trade and spread, which we motivate and test in this section.

By definition of the volatility per trade $\sigma_1^2 = E[(m_{\varrho+1} - m_{\varrho})^2]$ and of the instantaneous impact $r_{i,i+1} \equiv (m_{i+1} - m_i) \cdot \epsilon_i$, one has as an identity:[24]

$$\sigma_1^2 \equiv E\left[r_{i,i+1}^2\right] \tag{2.78}$$

The instantaneous impact $r_{i,i+1}$ is expected to fluctuate over time for several reasons. First, the volume of the trade, the volume in the book, and the spread strongly fluctuate with time (Mike and Farmer, 2008; Wyart et al., 2008). Large impact fluctuations may also arise from quote revisions due to addition or cancellation of limit orders. Second, there might also be important news affecting the "fundamental price" of the stock. These may result in large, instantaneous jumps of the midpoint with virtually no trade at all. In order to account for both effects, one may generalize the previous MRR relation (Eq. 2.67), as in Bouchaud et al. (2004), Rosenow (2002), and Wyart et al. (2008):

$$\sigma_1^2 = A\mathcal{R}_1^2 + \Sigma^2 \tag{2.79}$$

where $\mathcal{R}_1 \equiv E[\mathcal{R}_1(v)]$ is the average impact after one trade, A is a coefficient accounting for the variance of impact fluctuations, and Σ^2 is the news component of the volatility (see Section 2.6.2). This relation holds quite precisely across different stocks of the PSE, with a correlation of $R^2 = 0.96$ (see Figure 2.18). Perhaps surprisingly, the exogenous "news volatility" contribution Σ^2 is found to be small. (The intercept of the best affine regression is even found to be slightly negative.) This could be related to the observation made in Farmer et al. (2004)—and discussed earlier—that for most price jumps, some limit orders are cancelled too slowly and get "grabbed" by fast market orders. This means that most of these events also contribute to the impact component $\mathcal{R}_1$.[25] We can neglect Σ^2 in the preceding equation; in this sense the volatility of the stocks can be mostly attributed to market activity and trade impact. This is in agreement with the conclusions of Evans and Lyons on currency markets (Evans and Lyons, 2002); see also the discussion in Bouchaud et al. (2004) and Hopman (2007).

[24]Neglecting the extremely small drift contribution.

[25]One could argue that our results simply show that the news volatility Σ itself is proportional to $\overline{\mathcal{R}}_1$ and thus to the spread S. However, there is no reason that this should *a priori* be the case. For example, a model where rare jumps of typical amplitude J and probability per trade $p \ll 1$ lead to $\Sigma = \sqrt{p}J$, whereas the cost of such jumps, contributing to S, is $pJ \ll \Sigma$.

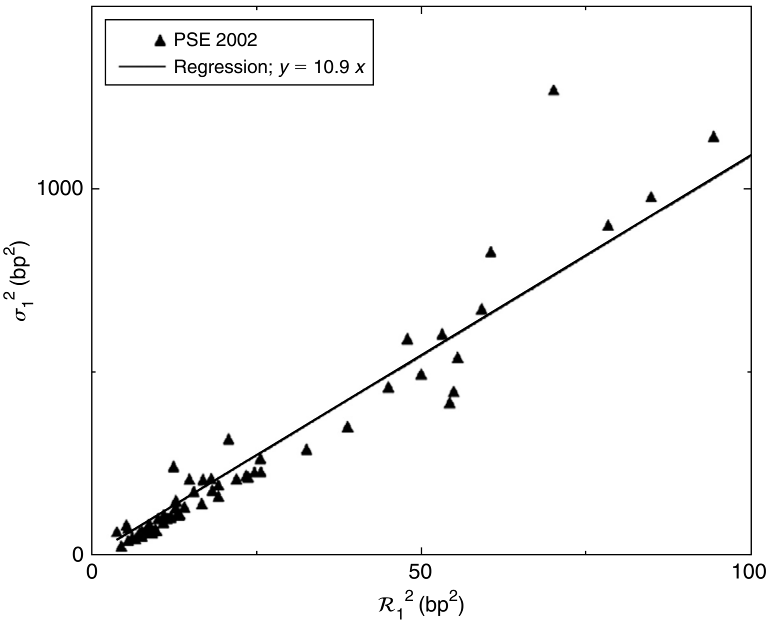

FIGURE 2.18 Plot of σ_1^2 vs. $\overline{\mathcal{R}}_1^2$, showing that the linear relation Eq. 2.79 holds quite precisely with $\Sigma^2 = 0$ and $a \approx 10.9$. (The intercept of the best affine regression is even found to be slightly negative.) Data here correspond to the 68 stocks of the PSE in 2002. The correlation is very high: $R^2 = 0.96$.

A final important assumption is that of *universality*. When the tick size is small enough and the typical number of shares traded is large enough, all stocks within the same market should behave identically up to a rescaling of the average spread and the average volume. In particular we assume that the statistics of (1) the volume of market orders (2) the spread S, and (3) the impact $\mathcal{R}_1$ and the various correlations between these quantities are independent of the stock when these quantities are normalized by their average value. Empirical evidence for (at least approximate) universality can be found in Lillo et al. (2003b) and Bouchaud et al. (2002). However, one expects that universality holds only for large-cap, small-tick stocks; large-tick stocks are not covered by the following analysis.

Universality then implies that:

$$E[vS] = BE[v]E[S], \quad E[v\mathcal{R}_1(v)] = B'E[v]\mathcal{R}_1 \tag{2.80}$$

where B, B' are stock independent numbers. Equation 2.80 accounts well for the Paris Stock Exchange data studied in Wyart et al. (2008), where it was found that $B \approx 1.02$ and $B' \approx 1.80$: The incoming volume and the spread are nearly uncorrelated, whereas the volume traded and the impact are correlated ($B' > 1$), as expected.

Therefore, using Eq. 2.76 as an equality and Eqs. 2.79 and 2.80 with $\Sigma^2 = 0$, we obtain the main result of this section:

$$E[S] = C\,\sigma_1 \tag{2.81}$$

where C is a stock-independent numerical constant, which can be expressed using the constants introduced as $C = 2\lambda B'/\sqrt{AB}$. This very simple relation between volatility *per trade* and average spread was noted in Bouchaud et al. (2004), Zumbach (2004), and Wyart et al. (2008), and we present further data to support this conjecture. Therefore, the fact that the cost of limit and market orders should be nearly equal on average Eqs. 2.68 and 2.76 and the absence of a specific contribution of news to the volatility lead to a particularly simple relation between liquidity and volatility. As an important remark, we note that the preceding relation is not expected to hold for the volatility *per unit time* σ, since it involves an extra stock-dependent and time-dependent quantity, namely the trading frequency f, through:

$$\sigma = \sigma_1\sqrt{f} \tag{2.82}$$

The predicted linear relation between spread and volatility per trade was tested empirically in Wyart et al. (2008) on small-tick stocks. For example, the results for the Paris Stock Exchange are shown in Figure 2.19. One finds that Eq. 2.81 describes the data very well, with R^2 values over 0.9. One can also check that there is an average intraday pattern that is followed in close correspondence both by $E[S]$ and σ_1: Spreads

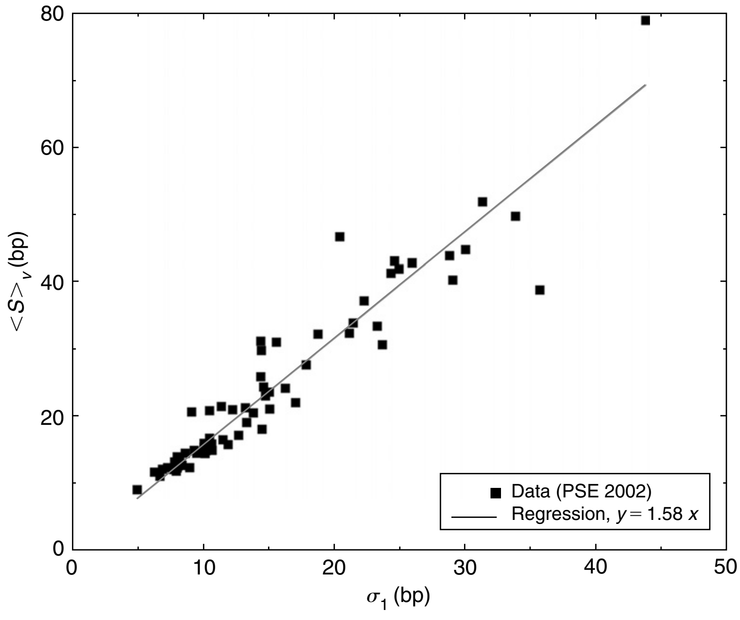

FIGURE 2.19 Test of Eq. 2.81 for 68 stocks from the Paris Stock Exchange in 2002, averaged over the entire year. The value of the linear regression slope is $c \approx 1.58$, with $R^2 = 0.96$.

are larger at the opening of the market and decline throughout the day. Note that the trading frequency f increases as time elapses, which, using Eq. 2.82, explains the familiar U-shaped pattern of the volatility per unit time.

Note that there are two complementary economic interpretations of the relation $\sigma_1 \sim S$ in small-tick markets:

- Since the typical available liquidity in the order book is quite small, market orders tend to grab a significant fraction of the volume at the best price; furthermore, the size of the "gap" above the ask or below the bid is observed to be on the same order of magnitude as the bid–ask spread itself, which therefore sets a natural scale for price variations. Hence both the impact and the volatility per trade are expected to be on the order of S, as observed.
- The relation can also be read backward as $S \sim \sigma_1$: When the volatility per trade is large, the risk of placing limit orders is large and therefore the spread widens until limit orders become favorable.

Therefore, there is a clear two-way feedback that imposes the relation $\sigma_1 \sim S$, and that can in fact lead to liquidity instabilities: Large spreads create large volatilities, which in turn may open the spread more. A detailed study of such effects would be highly valuable. On average, however, any deviation from the balance between spread and volatility tends to be corrected by the resulting relative flow of limit and market orders.

The result $\sigma_1 \sim S$ therefore appears as a fundamental property of the market organization, which should be satisfied within any theoretical description of the microstructure. This is an important constraint on models of order flow; however, none of the simple models studied in the past (zero intelligence models, Daniels et al., 2003; bounded-range models, Foucault et al., 2005, Luckock, 2003, and Rosu, 2008; or diffusion-reaction models, Slanina, 2001) is able to predict the preceding structural relation between S and σ_1 (see, however, Mike and Farmer, 2008, for recent developments using a "low intelligence" model, as discussed in the "A Simple Empirical Agent-Based Model for Liquidity Fluctuations" section).

2.8.4. Market Cap Effects

It is interesting to study the systematic dependence of the volatility and spread as a function of market capitalization M. Across stocks, the volatility per unit time shows a systematic slow decrease with M, $\sigma \propto M^{-\varphi}$, where φ is small. The trading frequency f, on the other hand, increases with M as $f \propto M^{\zeta}$. For stocks belonging to the FTSE-100, Zumbach finds $\zeta \approx 0.44$ (Zumbach, 2004), whereas for U.S. stocks the scaling for f is less clear (Eisler and Kertecz, 2006), with apparently two regimes, one for $M > 10$ B\$, where $\zeta \approx 0.44$, and the other for $M < 10$ B\$, for which $\zeta \approx 0.86$. The average amount per trade v_m, on the other hand, also increases with M in such a way that $f \times v_m$ is directly proportional to M. This last scaling holds with rather good accuracy and merely states that the total volume of transactions is proportional to market capitalization, which is somewhat expected a priori. What is interesting is that this is insured by having both the frequency of trades *and* the volume per trade increase with M, and not, for example, the transaction frequency at fixed amount per trade. The constant of proportionality is such that $\sim 10^{-3}$ of the total market cap is exchanged per day, on average, both in London and in New York (Zumbach, 2004; Eisler and Kertecz, 2006).

Combining the above two relations for the volatility per trade $\sigma_1 = \sigma/\sqrt{f}$ results in the following scaling law for the spread S,

$$S \sim \sigma_1 \propto M^{-\omega} \quad \omega = \varphi - \frac{\zeta}{2} \approx 0.22 \tag{2.83}$$

The average spread therefore decreases with market capitalization. This result is in good agreement with data from the LSE, Zumbach (2004), and from the PSE, Wyart et al. (2008). It can also be directly compared with the impact data of Lillo et al. (2003b) in the NYSE, where it was established that

$$\mathcal{R}_1(v) \approx M^{-0.3} F\left(M^{0.3}\frac{v}{\overline{v}}\right) \tag{2.84}$$

where $\overline{v}$ is the average volume per trade for a given stock and F a master curve that behaves approximately as a power law with exponent ψ. Since spread and impact are proportional, this last result is directly comparable to Eq. 2.83. The average over v of the preceding result then leads to $E[\mathcal{R}_1] \sim M^{-\omega}$ with $\omega \approx 0.3(1-\psi)$, which is in the range 0.15–0.25 (see Section 2.5.1 for a discussion of the value of ψ).

2.9. ORDER BOOK DYNAMICS

The previous section stresses the key role that liquidity plays in price formation. In double auction markets, prices are formed in the limit order book. Thus one obvious approach to understanding liquidity is to investigate the causes of liquidity fluctuations in the limit order book. Although the dynamics of liquidity are still very much an open question, several studies have identified statistical regularities in the behavior of limit order books and give some insight into the relationship between order flow and liquidity.

2.9.1. Heavy Tails in Order Placement and the Shape of the Order Book

There are several statistical regularities of limit orders placement. First, as mentioned, limit order signs are also well described by a long-memory process with a Hurst exponent very close to the one for market order signs. Lillo and Farmer (2004) reported a value of $H = 0.69$ for market orders and of $H = 0.71$ for limit orders.

Limit orders are characterized also by the limit price. The absolute value of the difference between the limit price and the best available price is a measure of the patience of the trader. Patient (impatient) traders submit limit orders very far from (close to) the spread. One of the statistical regularities recently observed in the microstructure of financial markets is the power-law distribution of limit order price in continuous double auction financial markets (Bouchaud et al., 2002; Zovko and Farmer, 2002). Let

$b(t) - \Delta$ denote the price of a new buy limit order and $a(t) + \Delta$ the price of a new sell limit order. Here $a(t)$ is the best ask price and $b(t)$ is the best sell price. The Δ is measured at the time when the limit order is placed. It is found that $\rho(\Delta)$ is very similar for buy and sell orders. Moreover, for large values of Δ the probability density function is well fitted by a single power law:

$$\rho(\Delta) \sim \frac{1}{\Delta^{1+\mu}} \tag{2.85}$$

There is no consensus on the value of the exponent μ. Zovko and Farmer (2002) estimated the value $\mu = 1.5$ for stocks traded at the London Stock Exchange, whereas Bouchaud et al. (2002) estimated the value $\mu = 0.6$ for stocks traded at the Paris Stock Exchange. More recently, Mike and Farmer (2008) fitted the limit order distribution for LSE stocks with a Student distribution, with 1.3 degrees corresponding to a value $\mu = 1.3$. This power law extends from 1 tick to over 100 ticks (sometimes even 1000 ticks), corresponding to a relative change of price of 5% to 50%. Such a broad distribution of limit order prices tells us that the opinion of market participants about the price of the stock in the near future could be anything from its present value to 50% above or below this value, with all intermediate possibilities. This means that market participants, quite oddly, anticipate the existence of large price jumps that would lead to trading opportunities.

A heavy tail in the distribution of relative limit price Δ indicates that there is a large heterogeneity in the limit price—that is, in the patience associated with each limit order. Patience is in turn related to the time scale the investor is willing to wait before her order is filled. The typical time to fill[26] of a limit order grows with Δ. In a recent study Lillo (2007) suggested that the origin of the heavy tails in the distribution of the relative limit price Δ can be attributed to a heterogeneity of time scales characterizing the trading behavior of individual utility maximizer investors and tested this theory using brokerage data from the LSE.

The order flow and the interaction of orders determine the instantaneous state of the book Ω_t. By averaging over time, empirical studies consistently show that the average shape of the order book is roughly symmetric between the bid and offer side of the book and is consistent across various stocks (Bouchaud et al., 2002; Zovko and Farmer, 2002; Mike and Farmer, 2008). They show that the maximum of the averaged book is not the best price, as shown in the top panel of Figure 2.20, even though this is the most likely place for an order to be placed. In Section 2.9.3 we present statistical models explaining this fact.

It is important to stress that the average shape of the book is very different from the "typical" shape of the book. As Farmer et al. (2004) showed, for most LSE stocks the typical shape of the book is extremely sparse (see the bottom panel of Figure 2.20). This occurs when the ratio between tick size and price is small so that there are often

[26]The mean time to fill of a limit order is infinite if the price process can be approximated by a random walk. "Typical" means some other measure such as the median time to fill.

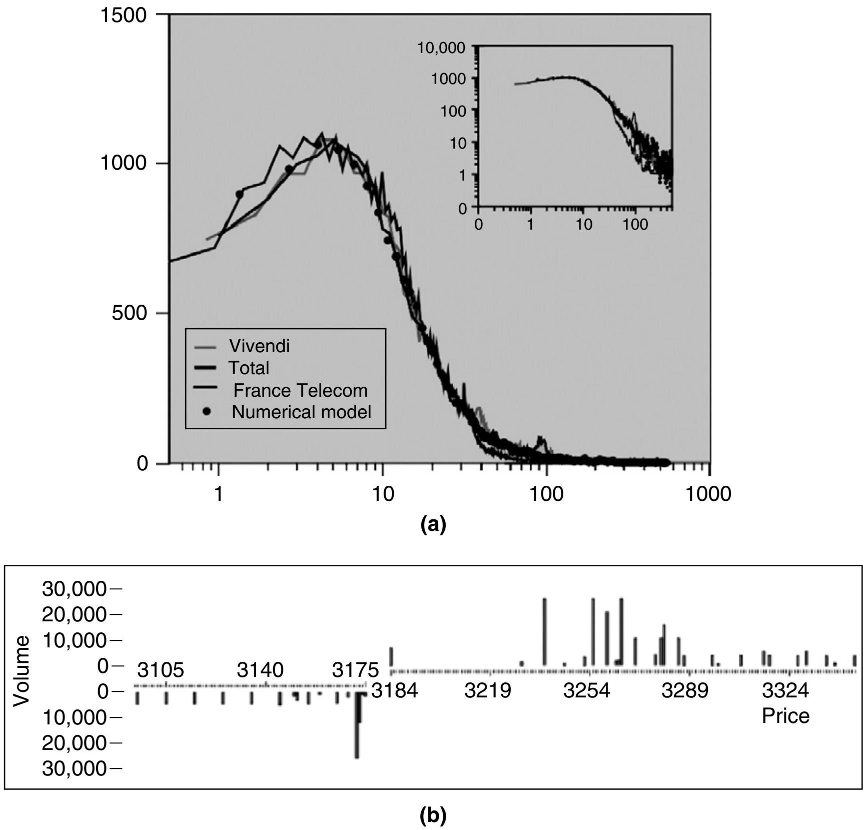

FIGURE 2.20 **(a)** Average shape of the order book. **(b)** Instantaneous shape of the order book.

many unoccupied price levels. As we discussed in Section 2.8.1, this fact has important consequences for the price impact of individual transactions and on the origin of large price fluctuations.

2.9.2. Volume at Best Prices: The Glosten-Sandas Model

The distribution of available volumes at best can be fitted by a gamma distribution with an exponent less than unity, meaning that the most probable value of the volume is much smaller than the average value. Both the value of the spread S and the quantity available at the bid and the ask, Φ_b, Φ_a, give information on the willingness of liquidity providers to enter a trade. One would like to understand the relation between these quantities—intuitively, large spreads are more favorable to liquidity providers and should attract larger volumes. More generally, it would be extremely interesting to have a theory for the shape of the whole order book, that is, the relation between the available volume and the distance from the best price.

The approach of Glosten and Sandas attempts to answer these questions, within a framework where market orders are informed trades (Glosten, 1994; Sandas, 2001). The idea is now that information is time dependent and modeled as a random variable that gives the predicted future variation of the midpoint, which we call (in conformity with the previous notation) $\epsilon_n r(n, n+\ell)$. Just before the n^{th} trade, a liquidity provider considers the volume of the queue at the ask, $\Phi_{a,n}$ and decides to add an extra (infinitesimal) limit order if its expected gain, conditional on execution, is greater than some minimum value $\mathcal{G}_{\min} \geq 0$. This reads:

$$E[m_{n+\ell} - m_n | \epsilon_n = 1, v_n \geq \Phi_{a,n}] \leq \frac{S_n}{2} - \mathcal{G}_{\min} \qquad \textbf{(2.86)}$$

At this stage, Glosten and Sandas add several questionable assumptions. A crucial one is that the volume that the informed trader chooses to trade is proportional to the information he has: $v_n = \alpha r(n, n+\ell)$, *independently* of the shape of the book at that moment, and in particular of the available volume at the ask. He is prepared to walk up the book if necessary, which occurs with only a very small probability in practice: As discussed in Section 2.6.1, trading is, in fact, discretionary. Introducing the probability of information content $P_\ell(r)$ and dropping the index n for convenience, the previous conditional expectation inequality reads:

$$\int_{\Phi_a/\alpha}^{+\infty} r P_\ell(r)\mathrm{d}r \leq \left[\frac{S}{2} - \mathcal{G}_{\min}\right] \int_{\Phi_a/\alpha}^{+\infty} P_\ell(r)\mathrm{d}r \qquad \textbf{(2.87)}$$

where we have used the fact that information is assumed to be reliable, that is, the expected midpoint change is indeed given by the informed trader prediction. To achieve a quantitative prediction, Sandas further assumes that $P_\ell(r)$ has an exponential shape:[27]

$$P_\ell(r) = \beta e^{-\beta r} \longrightarrow \frac{\Phi_a}{\alpha} + \frac{1}{\beta} \leq \frac{S}{2} + \mathcal{G}_{\min} \qquad \textbf{(2.88)}$$

In fact, this calculation can be reinterpreted to give the total volume of orders available $\Phi_<$ at a price less or equal to $p = m + S/2$ and therefore makes a prediction for the shape of the order book:

$$\Phi_<(p) = \alpha(p - m) - \alpha\mathcal{G}_{\min} - \frac{\alpha}{\beta} \qquad \textbf{(2.89)}$$

that is, a linear order book with slope α and, in principle, a *negative* intercept. (The prediction for the buy side of the book is obvious by symmetry.) Note that within this framework, the volume-dependent impact of market orders is by assumption linear: $\mathcal{R}_\ell(v) = v/\alpha$, which we already know is quite a bad representation of real data, where impact is always strongly sublinear (see Section 2.5.1). Altogether, this model fares quite badly compared with empirical data:

[27] This exponential assumption is in fact not so important. For example, a pure power-law distribution $P_\ell(r) \propto r^{-1-\mu}$ when $r > r_0$ would lead to the following result instead: $\Phi_a/\alpha \leq (1-\mu^{-1})[S/2 + \mathcal{G}_{\min}]$ $(\mu > 1)$.

- The order book intercept, which should be negative according to the model, is found to be positive when the model is fitted to empirical data, suggesting negative costs for placing limit orders.
- The slope α, when obtained from the slope of the order book, is found to be ten times larger than when obtained from direct impact estimates.
- As mentioned, the empirical shape of order books is nonmonotonic, exhibiting a maximum away from the best price. This is not accounted for by the model.

The reason for such a failure is essentially that, as discussed in Section 2.6.1, as shown by Farmer et al. (2004), the volume of the incoming market order is in fact strongly correlated with the available volume at the best price. This is in fact why impact is sublinear in volume and is at the heart of the liquidity game we have been detailing in the previous pages. One cannot consider that the market order flow is an exogenous process to which the limit order flow must adapt; rather, the two coevolve in a strongly intertwined manner.

One can, however, directly test Eq. 2.86 on empirical data, without any further theoretical assumptions, much as we did in the previous section. We choose $\ell = 1, 10, 100$ and identify a "neutral line" in the S, Φ plane separating the region (above that line) where executed limit orders are profitable from a region where they are costly (see Figure 2.21 and Eisler et al., 2008). One sees that after the $\ell = 1$ trade the separation line is flat and is located around the value of the average spread. This means that the value of the spread is such that limit orders and market orders break even on average at high frequencies, as discussed in Section 2.7.3. However, judged on longer time scales, the profitability of a limit order behind a large preexisting order only becomes positive for spreads significantly larger than the average. In other words, correlations between spread and volume, of the type predicted by the Glosten-Sandas model (Eq. 2.86), indeed appear on longer time scales.

2.9.3. Statistical Models of Order Flow and Order Books

An alternative point of view is to model the order flow directly as a stochastic process, decomposed into three components: market orders, addition of limit orders, and cancellation of limit orders.

Zero-Intelligence Models

There is a long list of literature that develops models of this type.[28] We will describe the approach of Daniels et al. (2003) (see also Smith et al., 2003), which has the advantage that it makes predictions that can be tested against real data. They assume that each elementary event is independent and concerns a fixed "quantum" of volume v. Buy

[28]Examples of stochastic process models of limit order books include Mendelson (1982), Cohen et al. (1985), Domowitz and Wang (1994), Bak et al. (1997), Bollerslev et al. (1997), Eliezer and Kogan (1998), Maslov (2000), Slanina (2001), Challet and Stinchcombe (2001), Daniels et al. (2003), Chiarella and Iori (2002), Bouchaud et al. (2002), Smith et al. (2003), Farmer et al. (2005), and Mike and Farmer (2008). See Smith et al. (2003) for a more detailed survey.

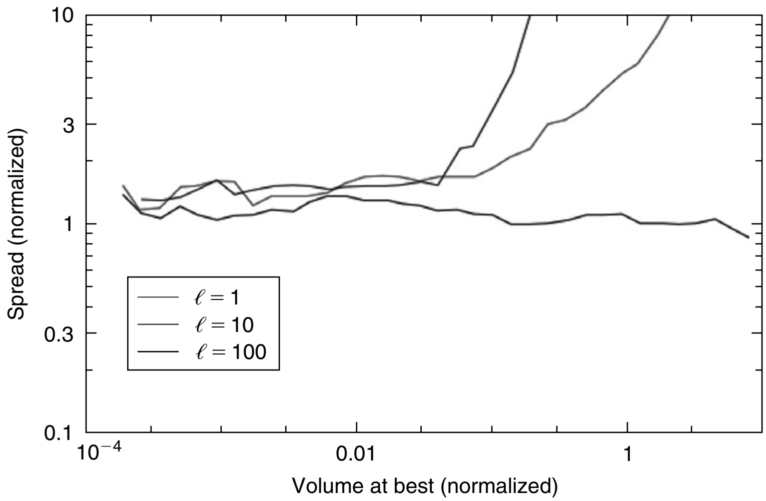

FIGURE 2.21 Neutral line for the profitability of adding a new limit order at the best price, for three different values of the horizon ℓ. The profit is positive above and negative below the curves. The indicated time horizons are given in number of transactions. The curves were gained by averaging for the ten most liquid stocks traded in the Paris Stock Exchange during 2002. Both volume and spread were normalized by their means for each stock before averaging. (*Source*: From Eisler et al., 2008).

and sell market orders are described by two Poisson processes of rate μ. Limit orders have a constant probability per unit time ρ to land anywhere they will not generate an immediate transaction, and existing limit orders have a probability ν to be canceled. This model is of course highly schematic, since it neglects all correlations between market and limit orders, in particular, the "stimulated refill" effect that we argued to be so important. Another important effect neglected is the dependence of the canceling rate on the size of the queue: One can actually observe that the probability of cancellation decreases as the number of orders at that price increases.

A simple self-consistent argument makes it possible to estimate the size of the spread S in this model. The total flux of limit orders between the midpoint and $S/2$ is by definition $\int_0^{S/2} d\Delta\rho(\Delta)$, where Δ is the distance from the midpoint and we are allowing here for the possibility that ρ might depend on Δ. If we assume that S is sufficiently small so that ρ is approximately constant, one finds that this incoming flux is $\approx \rho(0)S/2$. Whenever $\mu > \rho(0)S/2$, the rate of market order eats up the limit orders that appear within the spread completely, and the average volume present is close to zero. The cancellation term can be safely neglected if removal by market orders is more efficient—that is, when $\mu \gg \nu(0)$. But the argument breaks down when $S \sim 2\mu/\rho(0)$, which sets the typical position of the best price, provided the tick size is small compared to S. The spread is therefore larger for larger market order rates and smaller when the flow of limit orders is larger, as expected intuitively. The above "scaling" result for the spread has been derived more quantitatively when ρ and ν are independent of Δ. One finds for the average spread:

$$E[S] = \frac{\mu}{\rho} F\left(\frac{\nu}{\mu}\right) \tag{2.90}$$

where $F(u)$ is a monotonically increasing function that can be approximated as $F(u) \approx 0.28 + 1.86u^{3/4}$. Therefore, in the limit where cancellation can be neglected, one recovers the previous result $S \approx 0.28\mu/\rho(0)$. This prediction can be compared with empirical data by independently measuring the spread and the rates of the various processes (Farmer et al., 2005). In view of the simplicity of the model, the agreement with data is quite good, but systematic deviations remain. In view of the importance of feedback mechanisms that are neglected, this is hardly surprising.

These results are interesting because they demonstrate that some properties of the limit order book are dictated more or less automatically by the structure of the continuous double auction itself. In particular, Eq. 2.90 is an "equation of state" relating statistical properties of price formation to those of order flow. This equation of state is clearly inaccurate due to the extreme assumptions that must be made to derive it. However, it has some reasonable level of empirical validity, suggesting that such a relationship indeed exists for real markets. See the discussion concerning attempts to find a more realistic equation of state in the section "A Simple Empirical Agent-Based Model for Liquidity Fluctuations."

Statistical Model of Order Book

The preceding model can also explain the hump shape of the average order book. From a theoretical point of view, however, the problem is difficult to handle: If one chooses a fixed reference frame, the rates of incoming orders and cancellations change with the midpoint, whereas if one chooses the reference frame where the midpoint is fixed, limit orders that are already present get shifted around. The main difficulty comes from the fact that the motion of the midpoint is dictated by the order flow. To make progress, one can artificially decouple the motion of the midpoint and impose that it follows a random walk. An approximate quantitative theory of the volume in the book $\Phi(\Delta)$ can then be written as follows: Sell orders at distance Δ from the current midpoint at time t are those that were placed there at a time $t' < t$ and have survived until time t. These orders (1) have not been cancelled, and (2) have not been crossed by the ask at any intermediate time t'' between t' and t.

An order at distance Δ at time t in the reference frame of the midpoint $m(t)$ appeared in the order book at time t' at a distance $\Delta + m(t) - m(t')$. The average order book can thus be written, in the long time limit $t \to \infty$, as

$$\Phi_{\rm st}(\Delta) = \lim_{t\to\infty} \Phi(\Delta, t) = \int_{-\infty}^{t} {\rm d}t' \int du \rho\,(\Delta + u)\, \mathcal{P}\left(u|C(t,t')\right) {\rm e}^{-\nu(t-t')} \tag{2.91}$$

where $\mathcal{P}\,(u|C(t,t'))$ is the conditional probability that the time evolution of the price produces a given value of the midpoint difference $u = m(t) - m(t')$, given the condition that the path always satisfies $\Delta + m(t) - m(t'') \geq 0$ at all intermediate times $t'' \in [t', t]$.[29] The evaluation of $\mathcal{P}$ requires the knowledge of the statistics of the price process, which we assume to be purely diffusive. In this case, $\mathcal{P}$ can be calculated using the method of images. One finds:

$$\mathcal{P}\left(u|C(t,t')\right) = \frac{1}{\sqrt{2\pi D\tau}}\left[\exp\left(-\frac{u^2}{2D\tau}\right) - \exp\left(-\frac{(2\Delta+u)^2}{2D\tau}\right)\right] \tag{2.92}$$

where $\tau = t - t'$ and D is the diffusion constant of the price process.

After a simple computation, one finally finds, up to a multiplicative constant that only affects the overall normalization,

$$\Phi_{\text{st}}(\Delta) = \Phi(\Delta, t \to \infty) = \mathrm{e}^{-\alpha\Delta}\int_0^{\Delta} \mathrm{d}u\rho(u)\sinh(\alpha u) + \sinh(\alpha\Delta)\int_{\Delta}^{\infty} \mathrm{d}u\rho(u)\mathrm{e}^{-\alpha u} \tag{2.93}$$

where $\alpha^{-1} = \sqrt{D/2\nu}$ measures the typical variation of price during the lifetime of an order ν^{-1}.

The preceding formula depends on the statistics of the incoming limit order flow, modeled by $\rho(u)$. When $\rho(u) = e^{-\beta u}$, all integrals can be perfomed explicitly and one finds:

$$\Phi_{\text{st}}(\Delta) = \Phi_0\frac{\alpha\beta}{\alpha-\beta}\left[e^{-\beta\Delta} - e^{-\alpha\Delta}\right] \tag{2.94}$$

which can easily be seen to be zero for $\Delta = 0$, reach a maximum and decay back to zero exponentially at large Δ. Here Φ_0 is the total volume in the sell side of the book.

We have seen that the limit order price distribution is characterized by a power law with exponent μ (see Eq. 2.85). When $\mu < 1$, the parameter α in the preceding formula can be rescaled away in the limit where α^{-1} is much larger than the tick size (this is relevant for small-tick stocks, where $\alpha^{-1} \sim 10$ ticks). In this case, the shape of the average order book only depends on μ. In rescaled units $\delta = \alpha\Delta$, it is given by the following convergent integral:

$$\Phi_{\text{st}}(\Delta) = \mathrm{e}^{-\delta}\int_0^{\delta} \mathrm{d}u\, u^{-1-\mu}\sinh(u) + \sinh(\delta)\int_{\delta}^{\infty} \mathrm{d}u\, u^{-1-\mu}\mathrm{e}^{-u} \tag{2.95}$$

For $\Delta \to 0$, the average available volume vanishes in a singular way, as $\Phi_{\text{st}}(\Delta) \propto \Delta^{1-\mu}$, whereas for $\Delta \to \infty$, the average volume simply reflects the incoming flow of orders: $\Phi_{\text{st}}(\Delta) \propto \Delta^{-1-\mu}$. We have shown in Figure 2.22 the average order book obtained numerically from the previous Poisson model with a power-law order flow and compared it with Eq. 2.95, for various choices of parameters and $\mu = 0.6$, as found for various stocks of the Paris Stock Exchange. After rescaling the two axes, the numerical models lead to very similar average order books, and the analytic approximation, although crude, appears rather effective. The average shape of the order book therefore

[29] We neglect here the fluctuations of the spread. The condition should in fact read $\Delta + a(t) - a(t'') = \Delta + m(t) - m(t'') + (S(t) - S(t''))/2 \geq 0$.

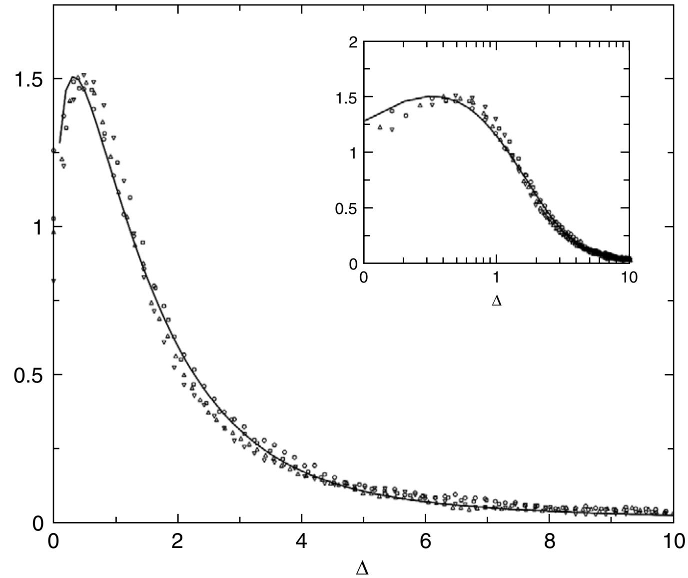

FIGURE 2.22 Average order book for a Poisson rate model with various choices of parameters (see Bouchaud et al., 2002) and $\mu = .6$. After rescaling the axes, the various results roughly fall on the same curve, which is well reproduced by the simple analytic approximation leading to Eq. 2.95, shown as the *solid line*.

reflects the competition between a power-law flow of limit orders with a finite lifetime and the price dynamics that remove the orders close to the current price.

A Simple Empirical Agent-Based Model for Liquidity Fluctuations

We now return to discuss the problem of the relationship between order flow and liquidity. The pure zero intelligence model of Daniels et al. (2003) was limited by its extreme assumptions of Poisson processes and the use of a highly stylized simplified model for order placement. A model based on more realistic assumptions was made by Mike and Farmer (2008). They made simple econometric models for order placement and cancellation and showed that by simulating this model it was possible to reproduce many of the empirical features of prices, including a quantitative match for the distribution of returns and the distribution of the spread.

In Section 2.9.1 we discussed the remarkable heavy tails in order placement. This result applied only to orders placed inside the same best.[30] Mike and Farmer also studied the distribution of order prices for orders placed inside the spread or crossing the opposite best (i.e., those generating immediate transactions). Remarkably they found, in a certain sense, that the same power-law behavior applied there as well. The frequency

[30]For example, for buy orders the same best is the best bid; the power law applies to orders placed at prices less than the best bid. The "opposite best" for buy orders is the best ask.

of order placement peaks at the same best and dies out on either side and can be reasonably well fit by a Student distribution (which has a power-law tail). Under the rule that orders that cross the opposite best price are executed, this simple rule does a reasonably good job of explaining execution frequency. One of the predictions that emerges automatically is that when the spread is small it is more likely for an order to cross the opposite best; that is, market orders become more likely. This at least partially explains the "stimulated refill" process mentioned earlier, since when the spread is large, orders chosen at random are more likely to fall inside the spread (and therefore accumulate in the limit order book), whereas when the spread is small executions are more likely. In fact, the model based on this process relied on this effect to preserve stability in the number of orders accumulating in the order book.

In this model the rate of cancellation was empirically found to depend on factors such as the number of orders in the order book, the imbalance in the order book, and the position of a given order relative to the opposite best price. Finally, it takes as an exogenous input the long memory of order signs discussed in Section 2.4. When these three elements (order placement, order sign, and cancellation) are put together, it is possible to simulate this model, generating a time series of order books with the corresponding prices. Note that it is critical that there is feedback between price formation and the order placement process. The resulting series of prices are not efficient, which is not surprising given that no effort was made to make them so and there are no agents who can take advantage of inefficiencies.

Nonetheless, for a subset of stocks with properties similar to those that were used to build the model, which were called "Type I" stocks, it does a good job of reproducing many of the properties of real prices.[31] In particular, it provides a good quantitative match with the distribution of returns and the distribution of the spread. This match includes not just the shape of the distributions but their scale, including the absolute level of volatility. That is, for Type I stocks a simulation of prices based on the measured parameters of the order flow produces forecasts of volatility that make a good match in absolute terms; in other words the predictions and measured values lie along the identity line. This provides further evidence for the existence of an equation of state relating order flow and prices. (It remains to extend this model so that it also works well for Types II and III.)

To summarize, the interesting point about this model is that it suggests that volatility is directly related to fairly simple properties of the order flow.

2.10. IMPACT AND OPTIMIZED EXECUTION STRATEGIES

The fact that trades impact prices is obviously detrimental to trading strategies: Since, again, liquidity is so small, trades must typically be divided into small chunks and spread

[31] Type I stocks are those with reasonably low volatility and small tick size. Type II stocks are those with high volatility, and Type III are stocks with large tick sizes. At this stage the model only performs well for Type I stocks.

throughout the day. But because of impact, the price paid for the last lot is on average higher than the price for the first lot. This poses a well-defined problem: What is the optimal trading profile as a function of time of day, such that the average execution price is as low as possible compared to the decision price (a quantity often called "implementation shortfall")?

Assume that a trader has a total volume V to execute; he decides to cut his order into N trades, each of size v, with $nv = V$. His trading profile $\phi(t)$ is such that the number of trades between t and $t + dt$ is $\phi(t)dt$. His own impact on the price of the stock at time $t' \geq t$ is modeled as

$$p(t') - p_0(t') = P(0) \int_0^{t'} dt \phi(t) G_0(t' - t) \ln v \tag{2.96}$$

where G_0 is the continuous time version of the single trade impact discussed in Section 2.6.4. Using all the results obtained previously, one has

$$G_0(t - t') = \frac{g_0 S}{f^\beta |t' - t|^\beta} \tag{2.97}$$

where g_0 is a number of order unity (since impact and spread are proportional) and f the number of trades per unit of time. We neglect here the possible dependence of the spread S and of f on time of day.

The total extra cost due to impact for a given profile $\phi(t)$ is therefore given by:

$$\int_0^T dt \int_0^t dt' \phi(t) G_0(t - t') \phi(t') \equiv \frac{1}{2} \int_0^T dt \int_0^T dt' \phi(t) G_0(|t - t'|) \phi(t') \tag{2.98}$$

where T is the trading period (say, one day). The previous quantity should be minimized with the constraint that the total trading volume is fixed, that is:

$$\int_0^T dt \phi(t) v = V \tag{2.99}$$

This problem can be easily solved using the method of functional derivatives with a Lagrange multiplier z to enforce the constraint. This leads to the following linear equation for the profile:

$$\int_0^T dt' G_0(|t - t'|) \phi(t') = z \tag{2.100}$$

where z is such that Eq. 2.99 is satisfied.

As a pedagogical example, let us assume that the impact decays exponentially as:

$$G_0(\tau) = G_0 \exp(-\alpha\tau) + G_\infty \tag{2.101}$$

Thanks to the constraint Eq. 2.99, the value of G_∞ can be reabsorbed in z and drops out of the computation; the permanent part of the impact is irrelevant to the optimization of execution costs (although the resulting implementation shortfall, of course, depends on G_∞). The solution of the optimization problem then reads:

$$G_0\phi^*(t) = z\delta(t) + z\delta(T - t) + \frac{z\alpha}{2} \quad \textbf{(2.102)}$$

and the constraint is:[32]

$$\frac{1}{G_0}\left[2\frac{z}{2} + \frac{z\alpha T}{2}\right] = V \longrightarrow z = \frac{G_0 V}{1 + \alpha T/2} \quad \textbf{(2.103)}$$

so finally:

$$\phi^*(t) = \frac{V}{1 + \alpha T/2}\left[\delta(t) + \delta(T - t) + \frac{\alpha}{2}\right] \quad \textbf{(2.104)}$$

the optimal profile is composed of two peaks at the open and at the close of the day and a flat profile in between. The ratio of the volume traded within the day to the volume traded at the open and at the close is $\alpha T/2$; for a fast-decaying impact (αT large), most of the volume should be spread out evenly intraday, whereas for a slowly decaying impact, trading should mostly concentrate at the open and at the close.

More generally, it can be shown that the solution to Eq. 2.100 is symmetric around $T/2$ and U-shaped (this is also mentioned in Hasbrouck, 2007, Chapter 15). In particular, when $G_0(\tau)$ is given by Eq. 2.97, one finds that the optimal profile diverges at both $t = 0$ and $t = T$, respectively, as $t^{\beta-1}$ and $(T - t)^{\beta-1}$. An approximate solution to Eq. 2.100 in that case reads:

$$\phi^*(t) \approx V\frac{\Gamma[2\beta]}{T^{2\beta-1}\Gamma^2[\beta]}t^{\beta-1}(T - t)^{\beta-1} \quad \textbf{(2.105)}$$

It is interesting to note that none of the parameters g_0, S, f entering in the numerical evaluation of G_0 appear in the shape of the profile, since again these can be reabsorbed in the definition of z at an early stage of the computation.

A generic U-shape solution for the optimized execution profile suggests an interesting interpretation of the observed U-shaped total traded volume as a function of the time of day.

2.11. TOWARD AN EMPIRICAL CHARACTERIZATION OF A MARKET ECOLOGY

The description of financial markets we have depicted is based on the assumption of the existence of different degrees of heterogeneity among market participants. The first level

[32]Note that the two delta functions only contribute half of their "area" to the total volume, since they are at the edge of the integration range.

of heterogeneity is due to the existence of a broad distribution of scales among market participants. Here scale refers to any quantity that measures the typical size of the trades of an investor. Moreover, the size of the hidden order determines the time horizon over which the order is worked and the number of transactions needed to complete the order.

As described in Section 2.3.8, the second degree of heterogeneity is due to the existence of (at least) two classes of agents acting systematically on opposite sides of the market. One group corresponds to liquidity providers and the other to liquidity takers. It would be extremely valuable to have a comprehensive empirical study that connects the heterogeneity of market participants with their strategy and with the properties of price dynamics. Unfortunately, it is not easy to obtain databases containing this level of information. Some data providers are starting to release datasets containing information about the financial institutions involved in the transaction and/or submission or cancellation of orders from the book.

It is important to stress that such financial institutions are not individual traders or agents but rather are usually credit entities and investment firms that are members of the stock exchange and are entitled to trade at the exchange. Very often these institutions are acting both as brokers for other clients and trading for their own account. Although an institution may act on behalf of many individuals and institutions with different strategies, recent findings show that in most cases it is possible to characterize an institution with an overall strategy, corresponding to that of the bulk of their trades. In the following two sections we present the results of two recent papers investigating the behavior of institutions in the Spanish Stock Exchange.

2.11.1. Identifying Hidden Orders

In a recent paper, Vaglica et al. (2008) used brokerage data on transactions in the Spanish Stock Exchange to identify hidden orders and to characterize their statistical properties. The identification of hidden orders is done using an algorithm designed to identify segments of the inventory time series of an institution characterized by an approximately constant and statistically significant drift term. The working hypothesis is that these segments are associated with hidden orders. A hidden order is characterized by the traded volume V, the number of transactions N, and the (real) time period T needed to complete the order.[33] It is found that the distribution of these quantities scales asymptotically for large values as

$$P(V > x) \sim \frac{1}{x^2} \quad P(N > x) \sim \frac{1}{x^{1.8}} \quad P(T > x) \sim \frac{1}{x^{1.3}} \qquad \textbf{(2.106)}$$

These relations show that the size of the hidden orders is asymptotically Pareto distributed in accordance with the hidden order model described in Section 2.4.3. It should be noted that the value of the exponent for V and for N is slightly larger than the

[33] In Vaglica et al. (2008), the investigated variables are the volume and the number of trades associated with those transactions characterizing the hidden order as a buy or sell hidden order.

value 1.5 expected by the theory described in Section 2.4.3 and a more careful testing of the theory is needed. The low value of the exponents indicates that the size of hidden orders is a very heterogeneous quantity, probably reflecting the heterogeneity of market participants. To test this hypothesis, Vaglica et al. (2008) have considered the distributional properties of hidden orders of individual brokerage codes. It is found that the distribution of hidden order size of individual brokers is consistent with a lognormal distribution, whereas the pool of the hidden orders of all brokers is not consistent with a lognormal. This indicates that investor size heterogeneity is at the origin of the power-law distribution of hidden order size.

The size variables of hidden orders are clearly related to each other. If the volume V is large, we expect that the number of transactions N and the time needed to complete the orders will also be large. The relation between the size variables reflects the strategic behavior chosen by the trader to work the order. By performing a principal component analysis of the hidden orders, Vaglica et al. find that

$$N \sim V^{1.1} \quad T \sim V^{1.9} \quad N \sim T^{0.66} \tag{2.107}$$

The fraction of variance explained by the first eigenvalue is on the order of 88%, so these characterizations are reasonably sharp.

The first relation indicates that the number of transactions of a hidden order is roughly proportional to its size. This means that even if a trader needs to trade a large hidden order, she will not split the order into larger chunks. This observation is consistent with the empirical finding that it is rare that the size of market orders is larger than the volume available at the opposite best (see Section 2.8.1 and Farmer et al., 2004). The other two relations indicate that the larger the volume of the hidden order, the slower the trading rate. This result has also been verified by using other statistical hidden order detection algorithms and still needs to be properly understood. Finally, it is worth noting that the relations of Eq. 2.107 also hold approximately true when one considers hidden orders belonging to a single brokerage code. In other words, the scaling relations of Eq. 2.107 are not the effect of heterogeneity among traders.

2.11.2. Specialization of Strategies

The presence of distinct classes of institutions and their mutual interaction has been investigated in a recent work by Lillo et al. (2008b). This study clearly identifies classes of institutions that are characterized by a similar trading behavior. Specifically, the study has focused on the cross-correlation between the inventory variation of different institutions. In general it is found that the cross-correlation of the inventory variation of different institutions is often statistically significant, for both positive and negative values. Principal component analysis reveals that the first eigenvalue of the correlation matrix is associated with a factor that is strongly correlated with price return. To give an idea of the level of correlation of the activity between different institutions, in Figure 2.23 we show the contour plot of the correlation matrix of daily inventory variation of the institutions trading the stock BBVA in the Spanish Stock Exchange in 2003.

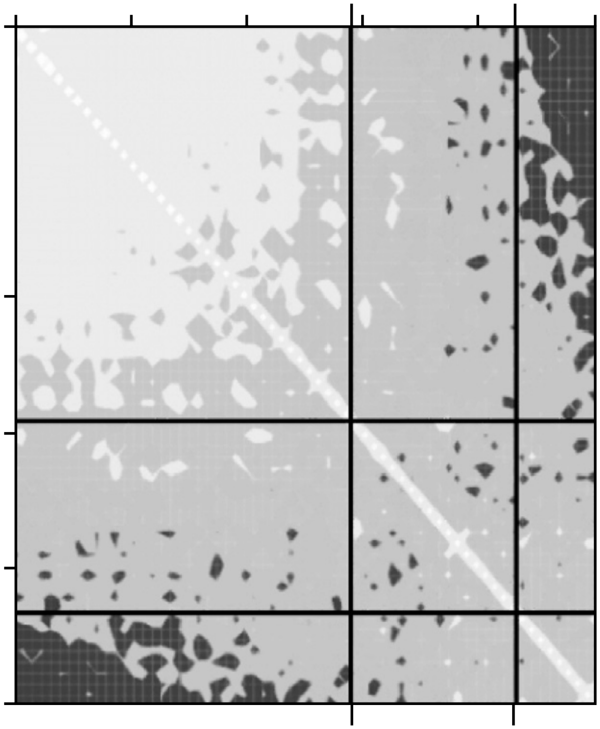

FIGURE 2.23 Contour plot of the correlation matrix of daily inventory variation of institutions trading the stock BBVA in 2003. This is plotted by sorting the firms in the rows and columns according to the strength of the correlation of their inventory variation with the return of the price of BBVA during the same period. Shades are chosen to highlight positive or negative firm daily inventory cross-correlation values above a given significance level. Specifically, *light gray* (black) indicates positive (negative) cross-correlation with a significance of 2σ, whereas *medium gray* indicates positive (negative) cross-correlation below 2σ. The *thick lines* in the matrix are obtained from the bottom panel by partitioning the firms into three groups according to the value of the correlation between returns and inventory variation (smaller than -2σ, between -2σ and 2σ, and larger than 2σ). (*Source*: Adapted from Lillo et al., 2008b).

The various shades of gray refer to different levels of correlation (see the figure caption). The institutions are sorted according to the value of the correlation of their inventory variation with the price return of BBVA. Two groups of firms are shown, one on the top left corner and the other on the bottom right corner.

The figure shows a clear block structure that makes it possible to identify communities of institutions characterized by a similar trading behavior. Specifically, the trading institutions can be partitioned into three subsets. The first (see the bottom right corner in Figure 2.23) is composed of institutions with an inventory variation positively correlated with price return—that is, these institutions buy when the price goes up and sell when the price goes down. Moreover, they are typically large institutions and have strongly autocorrelated order flow, probably because of order splitting. The second subset (see the top left corner in Figure 2.23) is composed of institutions whose inventory variation is negatively correlated with price return; these institutions buy when the price goes down and sell when the price goes up. The size of these institutions is very heterogeneous, as is the autocorrelation of their order flow. Finally, the third group is made up of uncategorized firms. As Figure 2.23 shows, the cross-correlation between the inventory variation of an institution belonging to the first group and an institution belonging to the second group is typically negative (dark areas in the top right and bottom left corners). This and other more direct evidence suggests that institutions belonging to these two groups are often trading counterparties.

2.12. CONCLUSION

In this review, we discussed market impact on two different, but overarching, levels. The first level deals with ultra high-frequency scales: that of elementary transactions (a level that in physics is called "microscopic"). It is concerned with the phenomenological description and mathematical modeling of empirical observations on order flow, impact, order book, bid–ask spread, and so on, which are of direct interest for high-frequency trading, execution, and slippage control. Results on that front are surprisingly rich and to some extent unexpected. Among the most salient points, one finds that impact of individual trades is nonlinear (concave) in volume and has a nontrivial lag dependence that can be thought of alternatively as a history-dependent impact. This is at variance with many simple models, including the famous Kyle model, where impact is assumed to be linear and permanent. The subtle temporal structure of impact can be traced back to the long memory in the fluctuations of supply and demand. The compatibility of the long memory in order flow and the absence of predictability of asset returns has profound consequences on price formation.

The second level deals with phenomena on a longer "coarse-grained" time scale and is more in line with the questions economists like to ask about markets and prices, such as: Are prices in equilibrium? What is the information content of these prices? or Why is the volatility so high? Much as in physics, where the detailed understanding of the microscopic world provides invaluable insight into macroscopic phenomena, we believe that a consistent picture of the microstructure mechanisms will help put in perspective some of these traditional questions about markets and prices. From the set of results presented previously, two concepts seem to emerge with possible far-reaching theoretical consequences:

- Because the outstanding liquidity of markets is always very small, trading is inherently an incremental process, and prices cannot be instantaneously in equilibrium and cannot instantaneouly reflect all available information. There is nearly always a substantial offset between latent offer and latent demand, which only slowly gets incorporated in prices. Only on an aggregated level does one recover an approximately linear impact with a permanent component.
- On anonymous, electronic markets, there cannot be any distinction between "informed" trades and "uninformed" trades. The average impact of all trades must

be the same, which means that impact must have a mechanical origin: If everything is otherwise held constant, the appearance of an extra buyer (seller) must on average move the price up (down).

The theory of rational expectations and efficient markets has increasingly emphasized information and belittled the role of supply and demand, in contradiction with the intuition of traders and of the layman.[34] The work we have reviewed here underlines the role of fluctuations in supply and demand, which may or may not be exogenous and may or may not be informed in a traditional sense—it does not really matter. Attempts to estimate *ex-post* the fraction of truly informed trades leads to very small numbers, at least judged on a short time basis, meaning that the concept of informed trades is not very useful to understand what is going on in markets at high frequencies. But still, prices manage to be almost perfectly unpredictable, even on very short time scales. The conclusion is that any useful notion of information must be internal to the market; trades, order flow, and cancellations *are* information, whatever the final cause of these events.

We are aware that some of these ideas go strongly against the prevailing view in market microstructure theory and entail a rather abrupt change of paradigm. We hope that this work will help clarify the issues and motivate further work to reconstruct a fully rigourous modeling framework, deeply rooted in the empirical data. Such data is now widely available and so abundant that it should be possible to raise the achievements of microstructure theory to the level of precision achieved in the physical sciences.

[34]On this point, see the lucid discussion in Lyons (2001), from which we reproduce the following excerpt: *Consider an example, one that clarifies how economist and practitioner worldviews differ. The example is the timeworn reasoning used by practitioners to account for price movements. In the case of a price increase, practitioners will assert, "there were more buyers than sellers." Like other economists, I smile when I hear this. I smile because in my mind the expression is tantamount to "price had to rise to balance demand and supply."*

APPENDIX 2.1: MECHANICAL VS. NONMECHANICAL IMPACT

As we summarized in Section 2.3.4, there are two very different views of what causes impact. The standard view is that it is essentially driven by information; the arrival of a trade signals new information, which causes market participants to update their valuations. But suppose a trade arrives that is really not based on any information. Does such a trade have a purely mechanical effect on prices? If so, what is the nature of that effect?

In Section 2.3.4 we introduced one such notion of mechanical impact, imagining a standard market clearing framework in which agents randomly alter their excess demand functions asynchronously. As each agent alters her demand function, she makes trades that generate market impact. Whether or not these are permanent depends on whether the alternations are permanent. Insofar as such alternations are permanent, the effect on prices will also be permanent.

In this appendix we examine another notion of mechanical impact for continuous double auctions. We define a mechanical impact as what happens if someone places an order in the order book if this order has no effect on any future orders. We are essentially asking the question of what happens to the price if an order is injected into the order book at random but no one pays any attention to it. We describe a method for analyzing order book data to answer this question (Farmer and Zamani, 2007). The essential result is that though there is a significant instantaneous mechanical impact due to the simple fact that such an order can consume the best quotes and move the midprice, this impact decays to zero. This decay seems to follow a power law, decaying very fast initially and very slowly later on. The reason for this decay is that orders are continually being removed from the order book, and as this happens the mechanical impact decays away. The mechanical impact as defined in this sense largely reflects the rate at which orders are flushed out of the order book.

A2.1.1. Definition of Mechanical Impact for Order Books

The following definition of mechanical impact makes the convenient simplifying assumption that the market framework is a continuous double auction. Consider the order flow $\Omega = (\omega_1, \omega_2, \ldots, \omega_t, \ldots)$ consisting of individual orders ω_t, which can be either new trading orders or cancellations of existing trading orders. Each individual order could be originated because of information relating to the value of the asset, or it could be originated "at random,"—for example due a demand for liquidity driven by events having no bearing on the asset being traded.

Under the rules of the continuous double auction, any initial limit order book and subsequent order flow generates a unique sequence of limit order books, which corresponds to a unique sequence of midprices. Auction A can be regarded as a deterministic function

$$b_{t+1} = A(b_t, \omega_{t+1})$$

that maps an order ω_t and a limit order book b_t onto a new limit order book b_{t+1}. For a given order flow $\Omega_t^{t+\tau} = \{\omega_t, \omega_{t+1}, \ldots, \omega_{t+\tau}\}$, Auction A is applied to each successive order to generate the limit order book $b_{t+\tau}$ at time $t+\tau$,

$$b_{t+\tau} = A^\tau(b_t, \Omega_t^{t+\tau})$$

The continuous double auction can thus be thought of as a deterministic dynamical system with initial condition b_0 and exogenous input Ω.

Each limit order book b_t defines a unique logarithmic midprice $p_t = p(b_t)$. The midpoint price at time $t+1$ can be written in terms of the composition of the price operator p and the auction operator A as $p_{t+1} = p \circ A(b_t, \omega_{t+1})$. Thus, any initial limit order book b_t and subsequent order flow $\Omega_t^{t+\tau}$ will generate a series of future prices $p_{t+1}, p_{t+2}, \ldots, p_{t+\tau}$, where, for example, the last price $p_{t+\tau}$ is

$$p_{t+\tau} = p \circ A^\tau(b_t, \Omega_t^{t+\tau}) \quad \textbf{(2.108)}$$

To give a precise meaning to mechanical impact, suppose we modify a particular order ω_t and replace it with a new order ω_t' while leaving the rest of the order flow unaltered. Since by assumption this modification does not affect the rest of the order flow, we can freely assume that it occurred for purely mechanical reasons. We can then compare the future stream of prices generated by the order flow $\Omega_t^{t+\tau} = \{\omega_t, \omega_{t+1}, \ldots, \omega_{t+\tau}\}$ to that generated by the altered order flow $\Omega_t'^{t+\tau} = \{\omega_t, \omega_{t+1}, \ldots, \omega_{t+\tau}\}$, for example, for time $t+\tau$, $p'_{t+\tau} = p \circ A^\tau(b_t, \Omega_t'^{t+\tau})$.

This can be used to measure the mechanical impact of any existing order ω_t by comparing the prices that are generated when ω_t is present to those that would have been generated if were were absent. We thus replace $\Omega_t^{t+\tau}$ by $\Omega_t'^{t+\tau} = \{0, \omega_{t+1}, \ldots, \omega_{t+\tau}\}$, where 0 in this case represents a null order, that is, one that does not change the order book. We can then define the *mechanical impact* $\Delta p_\tau^M(t)$ of the order ω_t as

$$\Delta p_\tau^M(t) = p \circ A^\tau(b_t, \Omega_t^{t+\tau}) - p \circ A^\tau(b_t, \Omega_{t+1}'^{t+\tau}) \quad \textbf{(2.109)}$$

The real price p contains both the informational and mechanical impact of order ω, while in the hypothetical price p' the mechanical impact is missing; that is, it contains only the informational impact. Under subtraction, only the mechanical impact remains. This isolates the part of the price impact that is "purely mechanical," in the sense that it is generated solely by the effect of placing an order in the book and observing its effect under the deterministic operation of the continuous double auction. The *information impact* can be defined as the portion of total impact that cannot be explained by mechanical impact, that is,

$$\Delta p_\tau^I = \Delta p_\tau^T - \Delta p_\tau^M$$

where Δp_τ^T is the total impact. Whatever components of the total impact not explained by mechanical impact must be due to correlations between the order ω_t and other events. With the data we have it is impossible to say whether the placement of the order ω_t causes changes in future events Ω_{t+1} or whether the properties of Ω_{t+1} are correlated with those of ω_t due to a common cause. In either case, changes in price that are not caused mechanically must be due to information—either the information contained in ω_t affecting Ω_{t+1} or external information affecting both ω_t and Ω_{t+1}. These ideas can be extended to apply to arbitrary modifications of the order stream—for example, infinitesimal modification of order ω_t, and to define mechanical generalization of elasticity (Zamani and Farmer, 2008).

A2.1.2. Empirical Results

This definition has been applied to several stocks in the London Stock Exchange and studied as a function of the order sequence number t (which as previously simply labels the temporal sequence in which the orders are received). It is clear that the mechanical impact is highly variable. In some cases there is an initial burst of mechanical impact, which dies to zero and then remains there. In some cases there are long gaps in which the impact remains at zero and then takes on nonzero values after more than 1000 transactions. In other cases there is no mechanical impact at all.

Despite the extreme variability, when an average is taken over time, a consistent pattern emerges. By definition for $\tau = 1$ the impact is entirely mechanical, since the only order that can affect the price is the reference event ω_t. After $\tau = 1$, however, the mechanical impact decays so that on average by the time of the next transaction it is roughly half its initial value; that is, it is half the total impact. In the limit as $\tau \to \infty$, for the stocks investigated in the LSE, to a good approximation for large τ the mechanical impact decays to zero as a power law $\Delta p_\tau^M(t) \sim \tau_M^\alpha$, with exponent $\alpha_M \approx 1.6$. See the example given in Figure A2.1.1. For event time the total impact and mechanical impact are by definition the same at $\tau = 1$. This is because in moving from $\tau = 0$ to $\tau = 1$ the only event that affects the price is the reference event ω_t; alterations in Ω_{t+1} cannot effect Δp_1^T. For larger values of τ the mechanical impact decreases and the informational impact increases. As the example shows, over the time scale shown here (100 orders), when measured in units of the average spread, the mechanical impact is initially about 0.17 and then decays monotonically toward zero. In contrast, the total impact increases toward what appears to be an asymptotically constant value slightly greater than 0.3. This is the source of our statement that the initial value of mechanical impact is about half the asymptotic value of the total impact.

In thinking about this, we should stress a few points. By requiring that any associated alterations of orders be considered information, we have taken a very strict definition of mechanical impact. Within our definition of "informational" there are two fundamentally different ways in which placing or removing an order can be correlated with the placement or removal of other orders. One is that placing or removal of an order causes a change in the placement or removal of another order. The alternative, however, is that the placement or removal of the two orders are caused by the same external event and are therefore correlated. From this point of view it can appear as though one order causes the other, simply because it happens to occur a bit earlier.

Though it might be surprising that the mechanical effect of order placement is completely temporary, in fact, this has a trivial explanation: Once all the orders were

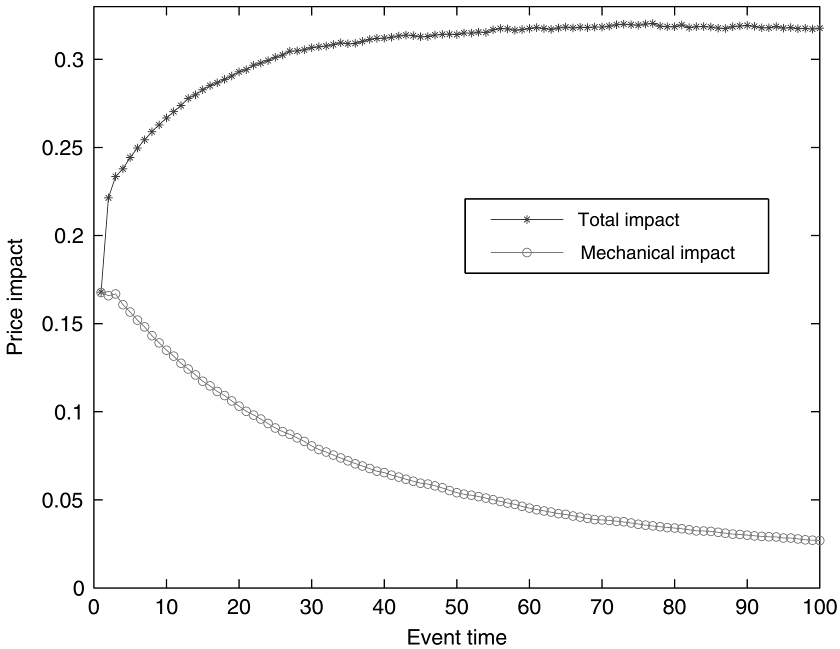

FIGURE A2.1.1 Average mechanical impact $E_t[\epsilon_t \Delta p_\tau^M(t)]$ (*squares*) and total impact $E_t[\epsilon_t \Delta p_\tau^T(t)]$ (*stars*) in units of the average spread, plotted as a function of number of the order sequence for the LSE stock AZN.

originally in the book when the reference order was placed, by definition all trace of the original order's presence is gone and so the mechanical impact is zero. Thus the power-law decay of market impact that is observed for mechanical impact is just a reflection of the rate at which orders turn over in the order book and is not related to the decay of the total impact discussed in Section 2.6.

APPENDIX 2.2: VOLUME FLUCTUATIONS

How should one take into account volume fluctuations in the formalism developed in Section 2.6.4? Since the volume of trades v is rather broadly distributed, the impact of trades could itself be highly fluctuating as well. This is not so, because large trade volumes mostly occur when a comparable volume is available at the opposite best price, in such a way that the impact of large trades is in fact quite similar to that of small trades. Mathematically, we have seen that the average impact is a slow power-law function v^ψ or even a logarithm $\log v$. As a simplifying limit, we postulate a logarithmic impact and a broad, lognormal distribution of v.

The resulting impact of the n^{th} trade $q_n = \epsilon_n \ln v_n$ is then a (zero mean) Gaussian random variable, which inherits long-range correlations from the sign process. Suppose that, as in the MRR model, only the surprise in q_n moves the price; this ensures *by* *construction* that the price returns are uncorrelated. An elegant way to write this down mathematically is to express the (correlated) Gaussian variables q_n in terms of a set of auxiliary uncorrelated Gaussian variables η_m, through:

$$q_n = \sum_{m \le n} K(n-m)\eta_m \quad E[\eta_m \eta_{m+\ell}] = \delta_{\ell,0} \tag{2.110}$$

where $K(n)$ is a certain kernel such that the q_n have the required correlations:[35]

$$C_\ell = E[q_n q_{n+\ell}] \equiv \sum_{m \ge 0} K(m+n)K(m) \tag{2.111}$$

In the case where C decays as $c_0 \ell^{-\gamma}$ with $0 < \gamma < 1$, it is easy to show that the asymptotic decay of $K(n)$ should also be a power-law $k_0 n^{-\delta}$ with $2\delta - 1 = \gamma$ and $k_0^2 = c_0 \Gamma(\delta)/\Gamma(\gamma)\Gamma(1-\delta)$. Note that $1/2 < \delta < 1$. Inverting Eq. 2.110 leads to:

$$\eta_n = \sum_{m \le n} Q(n-m) q_m \tag{2.112}$$

where Q is the matrix inverse of K such that $\sum_{m=0}^{\ell} K(\ell - m)Q(m) = \delta_{\ell,0}$. For a power-law kernel $K(n) \sim k_0 n^{-\delta}$, one obtains $Q(n) \sim (\delta - 1)\sin \pi\delta/(\pi k_0)n^{\delta-2} < 0$ for large n. Note that whenever $\delta < 1$, one can show that $\sum_{m=0}^{\infty} Q(m) \equiv 0$.

When all $q_m, m \le n-1$ are known, the corresponding $\xi_m, m \le n-1$ can be computed; the predicted value of the yet unobserved q_n is then given by:

$$E_{n-1}[q_n] = \sum_{m<n} K(n-m)\eta_m \tag{2.113}$$

and the surprise in q_n is simply:

$$q_n - E_{n-1}[q_n] = K(0)\eta_n \tag{2.114}$$

The generalization of the price equation of motion (Eq. 2.60) is therefore:

$$m_{n+1} - m_n = \xi_n + \theta K(0)\eta_n \tag{2.115}$$

which, again by construction, removes any predictability in the price returns. From this equation of motion one can derive $G_0(\ell)$ and $\mathcal{R}_\ell$.[36] From the expression of the η_n in terms of the q_n, one finds:

$$G_0(\ell) \equiv \theta K(0) \sum_{m=0}^{\ell-1} Q(m) = -\theta K(0) \sum_{m=\ell}^{\infty} Q(m) \tag{2.116}$$

[35]The following equation can be uniquely solved to extract $K(\ell)$ from C_ℓ using the so-called Levinson-Durbin algorithm for solving Toepliz systems (see, e.g., Percival, 1992).
[36]We now define G_0 as the impact of the q_n on the price.

Using the previous asymptotic estimate of the sum of matrices $Q(m)$, we finally obtain

$$G_0(\ell) \sim_{\ell \gg 1} \theta \frac{\sin(\pi\delta)K(0)}{\pi k_0} \ell^{\delta-1} \equiv \Gamma_0 \ell^{-\beta} \quad \textbf{(2.117)}$$

Identifying the exponents leads to $\beta = 1 - \delta = 1 - \gamma/2$, recovering the above equality. The quantity θ, relating surprise in order flow to price changes, measures the so-called "information content" of the trades. It can be measured from empirical data using the preceding relation between prefactors.

Finally, from Eq. 2.115, one finds the full impact function:

$$\mathcal{R}_\ell = E[(m_{n+\ell} - m_n)q_n] = \theta K(0)^2 \quad \forall \ell \quad \textbf{(2.118)}$$

that is, a completely *flat* impact function, independent of ℓ, as in the simplified MRR model decribed previously. However, if we assume with MRR that the fundamental price, rather than the midpoint, is impacted by the surprise in q_n, we find that the full impact function is again given by Eq. 2.61: $\mathcal{R}_\ell = \theta[1 - \mathcal{C}_\ell]$, which increases with ℓ.

APPENDIX 2.3: THE BID–ASK SPREAD IN THE MRR MODEL

A complementary point of view to that given in the main text is to analyze the cost of limit orders within the MRR model. The following argument is interesting because it can be, in essence, generalized to more complex cases as well. Suppose one wants to trade at a random instant in time. Compared to the initial midpoint value, the average execution cost of an infinitesimal buy limit order is given by:

$$\mathcal{C}_L = \frac{1}{2}\left(-\frac{S}{2}\right) + \frac{1}{2}\left(\mathcal{R}_1 + \mathcal{C}_L^+\right) \quad \textbf{(2.119)}$$

With probability $1/2$, the order is executed right away, $S/2$ below the midpoint; otherwise, the midpoint moves on average by a quantity $\mathcal{R}_1$, to which must be added the cost of a limit order conditioned to the last trade being a buy, $\mathcal{C}_L^+$, for which a similar equation can be obtained:

$$\mathcal{C}_L^+ = \frac{1-\rho}{2}\left(-\frac{S}{2}\right) + \frac{1+\rho}{2}\left(\mathcal{R}_1^+ + \mathcal{C}_L^{++}\right) \quad \textbf{(2.120)}$$

with obvious notations. Since the MRR model is Markovian, one has $\mathcal{R}_1^+ = \mathcal{R}_1$ and $\mathcal{C}_L^{++} = \mathcal{C}_L^+$, so that:

$$\mathcal{C}_L^+ = -\frac{S}{2} + \frac{1+\rho}{1-\rho}\mathcal{R}_1 \quad \textbf{(2.121)}$$

Plugging this last relation in Eq. 2.119, we finally find:

$$\mathcal{C}_L = -\frac{S}{2} + \frac{1}{1-\rho}\mathcal{R}_1 \quad \textbf{(2.122)}$$

Imposing that $C_L \equiv 0$, one recovers the MRR relation between the spread and the asymptotic impact (see Eq. 2.67).

References

Almgren, R., C. Thum, H. L. Hauptmann, and H. Li, Equity market impact, *Risk*, July 2005.

Ane, T., and H. Geman, Order flow, transaction clock, and normality of asset returns, *Journal of Finance*, 2000, 55, 2259–2284.

Bak, P., M. Paczuski, and M. Shubik, Price variations in a stock market with many agents. *Physica A*, 1997, 246, 430–453.

Barber, B. M., Y. -T. Lee, Y. -J. Liu, and T. Odean, Do individual day traders make money? Evidence from Taiwan, technical report, U.C. Davis, 2004.

Barclay, M. J., and J. B. Warner, Stealth trading and volatility, *Journal of Financial Economics*, 1993, 34, 281–305.

Beran, J., *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.

Bessembinder, H., Issues in assessing trade execution costs, *Journal of Financial Markets*, 2003, 6, 233–257.

Biais, B., T. Foucault, and P. Hillion, *Microstructure des marches financiers*, Presses Universitaires de France, 1997.

Black, F., Towards a fully automated exchange, *Review of Financial Analysts*, 1971, 27, 29–36.

Black, F., Noise, *Journal of Finance*, 1986, 41, 529–543.

Bollerslev, T., I. Domowitz, and J. Wang, Order flow and the bid–ask spread: An empirical probability model of screen-based trading, *Journal of Economic Dynamics and Control*, 1997, 21, 1471–1491.

Bouchaud, J. -P., and R. Cont, A Langevin approach to stock market fluctuations and crashes, *European Physics Journal B*, 1998, 6, 543–550.

Bouchaud, J. -P., Y. Gefen, M. Potters, and M. Wyart, Fluctuations and response in financial markets: The subtle nature of “random” price changes, *Quantitative Finance*, 2004, 4, 176–190.

Bouchaud, J. -P., J. Kockelkoren, and M. Potters, Random walks, liquidity molasses and critical response in financial markets, *Quantitative Finance*, 2006, 6, 115–123.

Bouchaud, J. -P., M. Mezard, and M. Potters, Statistical properties of the stock order books: Empirical results and models, *Quantitative Finance*, 2002, 2, 251–256.

Campbell, J. Y., and R. J. Shiller, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies*, 1989, 1, 195–228.

Casdagli, M., S. Eubank, J. D. Farmer, and J. Gibson, State space reconstruction in the presence of noise, *Physica D*, 1991, 51, 52–98.

Challet, D., So you are making money in financial markets. Should you tell your friends how?, technical report, 2007.

Challet, D., and R. Stinchcombe, Analyzing and modeling 1+1d markets, *Physica A*, 2001, 300, 285–299.

Chan, L. K., and J. Lakonishok, Institutional trades and intraday stock price behavior, *Journal of Financial Economics*, 1993, 33, 173–199.

Chan, L. K., and J. Lakonishok, The behavior of stock prices around institutional trades, *Journal of Finance*, 1995, 50, 1147–1174.

Chiarella, C., and G. Iori, A simulation analysis of the microstructure of double auction markets, *Quantitative Finance*, 2002, 2, 346–353.

Chordia, T., and A. Subrahmanyam, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Markets*, 2004, 72, 485–518.

Clark, P. K., Subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 1973, 41, 135–155.

Cohen, K. J., R. M. Conroy, and S. F. Maier, Order flow and the quality of the market, in Y. Amihud, T. Ho, and R. A. Schwartz (eds.), *Market Making and the Changing Structure of the Securities Industry*, pp. 93–110, Rowman & Littlefield, Lanham, 1985.

Curty, P., and M. Marsili, Phase coexistence in a forecasting game, *Journal of Statistical Mechanics*, P03013, 2006.

Cutler, D. M., J. M. Poterba, and L. H. Summers, What moves stock prices? *The Journal of Portfolio Management*, 1989, 15, 4–12.

Daniels, M. G., J. D. Farmer, L. Gillemot, G. Iori, and D. E. Smith, Quantitative model of price diffusion and market friction based on trading as a mechanistic random process, *Physical Review Letters*, 2003, 90, 108102–108104.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, Positive feedback and destabilizing rational speculation, *Journal of Finance*, 1990, 45, 379–395.

Ding, Z., C. W. J. Granger, and R. F. Engle, A long memory property of stock returns and a new model, *Journal of Empirical Finance*, 1993, 1, 83–106.

Domowitz, I., and J. Wang, Auctions as algorithms: computerized trade execution and price discovery, *Journal of Economic Dynamics and Control*, 1994, 18, 29–60.

Eisler, Z., J. -P. Bouchaud, and J. Kockelkoren, technical report, in preparation, 2008.

Eisler, Z., and J. Kertecz, Size matters, some stylized facts of the market revisited, *European Journal of Physics*, 2006, B51, 145–154.

Eliezer, D., and I. I. Kogan, Scaling laws for the market microstructure of the interdealer broker markets, technical report, www.arxiv.org/abs/cond-mat/9808240, 1998.

Embrechts, P., C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events* Springer Verlag, 1997.

Engle, R., and J. Rangel, The spline GARCH model for unconditional volatility and its global macroeconomic causes, technical report, NYU and UCSD, 2005.

Evans, M. D. D., and R. K. Lyons, Order flow and exchange rate dynamics, *Journal of Political Economy*, 2002, 110, 170–180.

Farmer, J., A. Gerig, F. Lillo, and S. Mike, Market efficiency and the long memory of supply and demand: Is price impact variable and permanent or fixed and temporary?, *Quantitative Finance*, 2006, 6, 107–112.

Farmer, J. D., Market force, ecology and evolution, *Industrial and Corporate Change*, 2002, 11, 895–953.

Farmer, J. D., L. Gillemot, F. Lillo, S. Mike, and A. Sen, What really causes large price changes?, *Quantitative Finance*, 2004, 4, 383–397.

Farmer, J. D., and F. Lillo, On the origin of power laws in financial markets, *Quantitative Finance*, 2004, 4, c7–c11.

Farmer, J. D., P. Patelli, and I. Zovko, The predictive power of zero intelligence in financial markets, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102, 2254–2259.

Farmer, J. D., and N. Zamani, Mechanical vs. informational components of price impact, *European Physical Journal B*, 2007, 55, 1899–2000.

Fisher, F. M., *Disequilibrium Foundations of Equilibrium Economics*, Cambridge University Press, 1983.

Foucault, T., O. Kadan, and E. Kandel, Limit order book as a market for liquidity, *The Review of Financial Studies*, 2005, 18, 1171–1217.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, Institutional investors and stock market volatility, *Quarterly Journal of Economics*, 2006, 121, 461–504.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, A theory of power-law distributions in financial market fluctuations, *Nature*, 2003, 423, 267–270.

Gallagher, D., and A. Looi. Trading behavior and the performance of daily institutional trades, *Accounting and Finance*, 2006, 46, 125–147.

Gerig, A., *A Theory for Market Impact: How Order Flow Affects Stock Price*, Ph.D. thesis, University of Illinois, 2007.

Gillemot, L., J. D. Farmer, and F. Lillo, There's more to volatility than volume, *Quantitative Finance*, 2006, 6, 371–384.

Glosten, L., Is the electronic limit order book inevitable?, *Journal of Finance*, 1994, 49, 1127–1161.

Glosten, L. R., and P. R. Milgrom, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics*, 1985, 14, 71–100.

Gopikrishnan, P., M. Meyer, L. Amaral, and H. E. Stanley, Inverse cubic law for the probability distribution of stock price variations, *European Physical Journal B.*, 1998, 3, 139–140.

Gopikrishnan, P., V. Plerou, X. Gabaix, and H. E. Stanley, Statistical properties of share volume traded in financial markets, *Physical Review E*, 2000, 62, R4493–R4496. Part A.

Granger, C. W. J., and R. Joyeux, An introduction to long-range time series models and fractional differencing, *Journal of Time Series Analysis*, 1980, 1, 15–30.

Grossman, S. J., *The Informational Role of Prices*. MIT Press, 1989.

Grossman, S. J., and J. E. Stiglitz, On the impossibility of informationally efficient markets, *American Economic Review*, 1980, 70, 393–408.

Gu, G. -F., W. Chen, and W. -X. Zhou, Quantifying bid–ask spreads in the Chinese stock market using limit-order book data: Intraday pattern, probability distribution, long memory, and multifractal nature, *European Physical Journal B*, 2007, 57, 81–87.

Guedj, O., and J.-P. Bouchaud, Expert's earnings forecasts: Bias, herding and gossamer information, *International Journal of Theoretical and Applied Finance*, 2005, 8, 933–946.

Handa, P., and R. A. Schwartz, Limit order trading, *Journal of Finance*, 1996, 51, 1835–61.

Harris, L., and J. Hasbrouck, Market vs. limit orders: The superdot evidence and order submission strategy, *Journal of Financial and Quantitative Financial Analysis*, 1996, 31, 213–231.

Hasbrouck, J., Measuring the information content of stock trades, *Journal of Finance*, 1991, 46, 179–187.

Hasbrouck, J., Trades, quotes, inventories, and information, *Journal of Financial Economics*, 1988, 22, 229–52.

Hasbrouck, J., *Empirical Market Microstructure: The Institutions, Economics and Econometrics of Securities Trading*, Oxford University Press, 2007.

Hausman, J., A. W. Lo, and A. C. MacKinlay, An ordered profit: Analysis of transaction stock prices, *Journal of Financial Economics*, 1992, 31, 319–379.

Hopman, C., Do supply and demand drive stock prices?, *Quantitative Finance*, 2007, 7, 37–53.

Hosking, J. R. M., Fractional differencing, *Biometrika*, 1981, 68, 165–176.

Joulin, A., A. Lefevre, D. Grunberg, and J. -P. Bouchaud, Stock price jumps: News and volume play a minor role, Wilmott *Magazine*, 2008, 46.

Keim, D. B., and A. Madhavan, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies*, 1996, 9, 1–36.

Kempf, A., and O. Korn, Market depth and order size, *Journal of Financial Markets*, 1999, 2, 29–48.

Kyle, A. S., Continuous auctions and insider trading, *Econometrica*, 1985, 53, 1315–1335.

La Spada, G., J. D. Farmer, and F. Lillo, The nontrivial random walk of stock prices, *European Journal of Physics B*, 2008, 64, 607–614.

LeBaron, B., and R. Yamamoto, Long-memory in an order-driven market, *Physica A*, 2007, 383, 85–89.

Lillo, F., Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices, *European Journal of Physics*, 2007, 55, 453–459.

Lillo, F., and J. D. Farmer, The long memory of the efficient market, *Studies in Nonlinear Dynamics & Econometrics*, 2004, 8, Article 1.

Lillo, F., and J. D. Farmer, The key role of liquidity fluctuations in determining large price fluctuations, *Fluctuations and Noise Letters*, 2005, 5, L209–L216.

Lillo, F., J. D. Farmer, and A. Gerig, A theory for aggregate market impact, technical report, Santa Fe Institute, unpublished research, 2008a.

Lillo, F., J. D. Farmer, and R. N. Mantegna, Master curve for price impact function, *Nature*, 2003b, 421, 129–130.

Lillo, F., and R. N. Mantegna, Power-law relaxation in a complex system: Omori law after a financial market crash, *Physical Review E*, 2003, 016119.

Lillo, F., S. Mike, and J. D. Farmer, Theory for long memory in supply and demand, *Physical Review E*, 2005, 7106, 066122.

Lillo, F., E. Moro, G. Vaglica, and R. Mantegna, Specialization and herding behavior of trading firms in a financial market, *New Journal of Physics*, 2008b, 10, 043019.

Lo, A. W., Long-term memory in stock market prices, *Econometrica*, 1991, 59, 1279–1313.

Lobato, I. N., and C. Velasco, Long memory in stock-market trading volume, *Journal of Business & Economic Statistics*, 2000, 18, 410–427.

Luckock, H., A steady-state model of the continuous double auction, *Quantitative Finance*, 2003, 39, 385–404.

Lux, T., The stable Paretian hypothesis and the frequency of large returns: An examination of major german stocks, *Applied Financial Economics*, 1996, 6, 463–475.

Lyons, R., *The Microstructure Approach to Foreign Exchange Rates*, MIT Press, 2001.

Madhavan, A., Market microstructure: A survey, *Journal of Financial Markets*, 2000, 3, 205–258.

Madhavan, A., M. Richardson, and M. Roomans, Why do security prices change? A transaction-level analysis of NYSE stocks, *Review of Financial Studies*, 1997, 10, 1035–1064.

Mandelbrot, B. B., The variation of certain speculative prices, *Journal of Business*, 1963, 36, 394–419.

Mandelbrot, B. B., and J. W. van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Review*, 1968, 10, 422–437.

Mandelbrot, B. B., and H. M. Taylor, On distribution of stock price differences, *Operations Research*, 1967, 15, 1057–1062.

Maslov, S., Simple model of a limit order-driven market, *Physica A*, 2000, 278, 571–578.

Mendelson, H., Market behavior in a clearing house, *Econometrica*, 1982, 50, 1505–1524.

Mike, S., and J. D. Farmer, An empirical behavioral model of liquidity and volatility, *Journal of Economic Dynamics and Control*, 2008, 32, 200–234.

Milgrom, P., and N. Stokey, Information trade and common knowledge, *Journal of Economic Theory*, 1982, 26, 17–27.

Odean, T., Do investors trade too much?, *American Economic Review*, 1999, 89, 1279–1298.

O'Hara, M., *Market Microstructure Theory*, Blackwell, 1995.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a time series, *Physical Review Letters*, 1980, 45, 712–716.

Percival, D., Simulating gaussian random processes with specified spectra, *Computing Science and Statistics*, 1992, 24, 534.

Plerou, V., P. Gopikrishnan, X. Gabaix, and H. E. Stanley, Quantifying stock price response to demand fluctuations, *Physical Review E*, 2002, 66, article no. 027104.

Plerou, V., P. Gopikrishnan, X. Gabaix, and H. E. Stanley, On the origin of power laws in financial markets, *Quantitative Finance*, 2004, 4, 11–15.

Plerou, V., P. Gopikrishnan, and H. E. Stanley, Quantifying fluctuations in market liquidity: Analysis of the bid–ask spread, *Physical Review E*, 2005, 71, 046131–9.

Ponzi, A., F. Lillo, and R. Mantegna, Market reaction to temporary liquidity crises and the permanent market impact, preprint at arXiv:physics/0608032v1.

Roll, R., Orange juice and weather, *American Economic Review*, 1984, pp. 861–880.

Rosenow, B., Fluctuations and market friction in financial trading, *International Journal of Modern Physics C*, 2002, 13, 419–425.

Ross, S., *Neoclassical Finance*. Princeton University Press, 2004.

Rosu, I., A dynamic model of the limit order book, 2008, *Review of Financial Studies*.

Sandas, P., Adverse selection and comparative market making: Empirical evidence from a limit order market, *Review of Financial Studies*, 2001, 14, 705–734.

Sebenius, J. K., and J. Geanakoplos, Don't bet on it: Contingent agreements with asymmetric information, *Journal of the American Statistical Association*, 1983, 78, 424–426.

Shiller, R. J., Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review*, 1981, 71, 421–436.

Shleifer, A., *Clarendon Lectures: Inefficient Markets*, Oxford University Press, 2000.

Slanina, F., Mean-field approximation for a limit order-driven market model, *Physical Review E*, 2001, 64, Article no. 056136, Part 2.

Smith, E., J. D. Farmer, L. Gillemot, and S. Krishnamurthy, Statistical theory of the continuous double auction, *Quantitative Finance*, 2003, 3, 481–514.

Stoll, H., Friction, *Journal of Finance*, 2000, 55, 1479.

Stoll, H. R., The supply of dealer services in securities markets, *Journal of Finance*, 1978, 33, 1133–51.

Takens, F., Detecting strange attractors in turbulence, in D. Rand and L.-S. Young (eds.), *Dynamical Systems and Turbulence*, vol. 898, pp. 366–381, Springer-Verlag, 1981.

Torre, N., *BARRA Market Impact Model Handbook*, BARRA Inc., 1997.

Vaglica, G., F. Lillo, E. Moro, and R. Mantegna, Scaling laws of strategic behavior and size heterogeneity in agent dynamics, *Physical Review E*, 2008, 77, 0036110.

Weber, P., and B. Rosenow, Order book approach to price impact, *Quantitative Finance*, 2005, 5, 357–364.

Weber, P., and B. Rosenow, Large stock price changes: Volume or liquidity?, *Quantitative Finance*, 2006, 6, 7–14.

Wiesinger, J., Z. Eisler, A. Joulin, and J. -P. Bouchaud, Dynamics of the bid–ask spread in the Glosten-Milgrom model: analytical results, technical report, 2008.

Wyart, M., J. -P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo, Relation between bid–ask spread, impact and volatility in double auction markets, *Quantitative Finance*, 8, 41–57, 2008.

Zamani, N., and J. D. Farmer, Decomposition of mechanical and informational components of orderflow, technical report, 2008.

Zawadowski, A., G. Andor, and J. Kertecz, Short-term market reaction after extreme price changes of liquid stocks, *Quantitative Finance*, 2006, 4, 283–295.

Zovko, I., and J. D. Farmer, The power of patience: A behavioral regularity in limit order placement, *Quantitative Finance*, 2002, 2, 387–392.

Zovko, I., and J. D. Farmer, Correlations and clustering in the trading of members of the London Stock Exchange, in S. Abe, T. Mie, H. Herrmann, P. Quarati, A. Rapisarda, and C. Tsallis (eds.), *Complexity, Metastability and Nonextensivity; An International Conference*, AIP Conference Proceedings, Springer, 2007.

Zumbach, G., How the trading activity scales with the company sizes in the FTSE 100, *Quantitative Finance*, 2004, 4, 441.

CHAPTER 3

Stochastic Behavioral Asset-Pricing Models and the Stylized Facts

Thomas Lux
Department of Economics, University of Kiel
and Kiel Institute for the World Economy

3.1. Introduction 162
3.2. The Stylized Facts of Financial Data 164
3.2.1. Martingales, Lack of Predictability, and Informational Efficiency 164
3.2.2. Fat Tails of Asset Returns 167
3.2.3. Volatility Clustering and Dependency in Higher Moments 173
3.2.4. Other Stylized Facts 174
3.3. The Stylized Facts as "Scaling Laws" 175
3.4. Behavioral Asset-Pricing Models with Interacting Agents 178
3.4.1. Interaction of Chartists and Fundamentalists and Nonlinear Dynamics of Asset Prices 179
3.4.2. Kirman's Model of Opinion Formation and Speculation 185
3.4.3. Beyond Local Interactions: Socioeconomic Group Dynamics in Financial Markets 191
3.4.4. Lattice Topologies of Agents' Connections 207
3.5. Conclusion 210
References 211

Note: I am grateful to Maren Brechtefeld and Claas Prelle for their careful proofreading of the manuscript: to Dietrich Stauffer and an anonymous referee for their detailed and constructive reviews of a previous version of this chapter; and to this book's editors, Klaus Reiner Schenk-Hoppé and Thorsten Hens, for their helpful comments and suggestions. Part of the present survey was completed during a visit to GREQAM, Université de la Méditerranée, Aix-Marseille II, the generous hospitality of which is gratefully acknowledged. I am also thankful for the financial support of the European Commission under STREP contract no. 516446.

Abstract

High-frequency financial data are characterized by a set of ubiquitous statistical properties that prevail with surprising uniformity. Though these "stylized facts" have been well-known for decades, attempts at their behavioral explanation have remained scarce. However, recently a new branch of simple stochastic models of interacting traders has been proposed. These models share many of the salient features of empirical data. They draw some of their inspiration from the broader current of behavioral finance. However, their design is closer in spirit to models of multiparticle interaction in physics than to traditional asset-pricing models. This reflects a basic insight in the natural sciences that similar regularities like those observed in financial markets (denoted as "scaling laws" in physics) can often be explained via the microscopic interactions of the constituent parts of a complex system. Since these emergent properties should be independent of the microscopic details of the system, this viewpoint advocates negligence of the details of the determination of individuals' market behavior and instead focuses on the study of a few plausible rules of behavior and the emergence of macroscopic statistical regularities in a market with a large ensemble of traders. This chapter reviews the philosophy of this new approach, its various implementations, and its contribution to an explanation of the stylized facts in finance.

Keywords: stylized facts, power laws, agent-based models, interacting agents

3.1. INTRODUCTION

The finance literature has a somewhat ambiguous approach toward the salient empirical features that characterize financial markets. Though they are identified as "stylized facts" in recent surveys (de Vries, 1994; Pagan, 1996), they have more often been christened as "anomalies" in the past (see Frankfurter and McGoun, 2001, who argue that the increasing [mis]use of the term *anomaly* in the finance literature is evidence of a propagandistic "effort to imply that . . . the reigning paradigm is irreplaceable . . . "; from their abstract). The difference in language is perplexing: Whereas the former notion implies an identification of robust features of the data that call for a scientific explanation, the latter rather appears to denounce the same features as a minor nuisance for the established theoretical framework. One certainly does not do injustice to a large body of theoretical research in finance by stating that it had almost entirely ignored some of the most pervasive characteristics of financial markets for quite some time. Though this does not hold for all the stylized facts, it is certainly indisputable for two important regularities that have motivated a large part of the empirical finance literature: the fat tails of asset returns and the characteristic time variation of their fluctuations. To be honest, a few attempts at explaining these features on the base of standard modeling frameworks do exist in recent literature (see Vanden, 2005),[1] but at least there has

[1]However, his results are rather supportive of an alternative approach. Studying the capacity of representative agent equilibrium models to account for volatility clustering, he concludes that "it is doubtful that there exists any representative equilibrium model . . . that is consistent with the data" (p. 374).

been no systematic theoretical approach toward their explanation within "mainstream" models.

However, it also must be emphasized that mainstream finance has not been careless about empirical results altogether; on the contrary, one of the most important empirical findings, the martingale character of prices, is at the heart of its main paradigmatic approach, the efficient market hypothesis. It appears, however, that in focusing on the explanation of this single feature, other equally universal findings have been deliberately neglected and marginalized as anomalies. The point made in this chapter is that, from a different perspective, what has been found to be strange and unexpected behavior of markets might appear as revealing characteristics that could guide the scientist toward a candidate explanation of price dynamics in financial markets. The surprising insight here is that, when presented in an appropriate format, the stylized facts so well known to econometricians and market practitioners would immediately be identified as *scaling laws* by natural scientists. Viewed from this perspective, a picture emerges that differs enormously from that of traditional finance: Scaling laws in natural science are viewed as imprints of "complex" systems composed of many interacting subunits that have to be explained as a result of their microscopic interaction. This motivates an approach toward modeling of financial markets focusing on the interaction of many actors rather than intertemporal optimization of representative investors. Models with such an emphasis have been proposed from the early 1990s both by economists dissatisfied with the representative agent methodology as well as by physicists in the evolving "econophysics" movement. To some extent the promise of the scaling approach seems to have materialized: Models with interacting agents of a certain type appear to be quite robust generators of the formerly mysterious anomalies of fat tails and clustered volatility. This explanatory power for some of the previously unexplained characteristics of financial markets might lend some credibility to this new approach.

With their focus on emergent properties of microeconomic interactions of market participants, stochastic agent-based models could provide the missing link between the more micro-oriented analysis of behavioral biases and the econometric literature on aggregate characteristics of markets. These models mostly lack a full-fledged foundation in utility maximization or alternative psychological decision mechanisms. The behavioral rules encoded in their stochastic dynamics are more germane to myopic boundedly rational behavior rather than perfectly rational utility maximization over an infinite horizon. The emphasis on the aggregate market outcome of uncoordinated activities of individual investors often provides patterns that are close in spirit to popular perceptions of financial markets and could be seen as a formalization of psychological factors and irrational components of human behavior in theories such as Kindlebeger's (1989) view on bubbles and crashes. One could argue that with its emphasis on euphoria, hysteria, and self-deception among speculators, Kindleberger's theory would defy a formalization along the lines of fully rational utility-maximizing individual behavior and would require the type of phenomenological formalization that is detailed here. In this sense, this new approach could be viewed as a continuation of a time-honored tradition that had been marginalized by its incompatibility with the basic axioms of mainstream economic theory. The formalization of these approaches provides an avenue toward empirical estimation and tests of hypotheses derived from such a framework.

Though this literature is still in its infancy, one could imagine various practical applications; the direct behavioral modeling of "sentiment" factors (see Section 3.4.3) offers new insights on the determinants and dynamics of waves of excessive optimism and pessimism and their influence on asset valuation. Widely available sentiment measures have been used in econometric studies as an exogenous variable (see Lee et al., 2002; Brown and Cliff, 2005), but the theories detailed here allow for its endogenous determination along with the unfolding price dynamics in a speculative market. In empirical applications, one could then estimate joint models of, for instance, epidemic dynamics of market sentiment together with a more conventional asset-pricing equation. For example, Alfarano et al. (2005) estimate the parameters of a simple stochastic model of two groups of interacting agents that takes the form of a stochastic volatility model, the parameters of which have a behavioral interpretation. Another important area of applications is in market design and regulatory policy: Models for which output is close to the empirical stylized facts should be a good test case for studying the effects of various trading protocols, clearing mechanisms, and regulations. Although this would require some effort at adding institutional detail to a relatively abstract theoretical setup, a certain number of studies have already scrutinized agent-based models as a means to explore the effects of various regulatory schemes (see Pellizzari and dal Forno, 2007, or Bottazzi et al., 2005).

The remainder of this chapter starts with an outline of the empirical stylized facts that have been of such utmost importance for the development of stochastic agent-based models. In Section 3.2 we discuss, in turn: the martingale property, fat tails, and clustering of volatility and we have a cursory look at other reported regularities. Section 3.3 highlights the interpretation of these stylized facts as "scaling laws" and the connotations of this view for theoretical modeling. Section 3.4 goes into detail about some representative models in the area; we start with a short exposition of sources of inspiration for these models in Section 3.4.1 in the older literature on interaction of different groups of speculators (e.g., fundamentalists vs. chartists) and then move on to models that are very explicitly based on microscopic interactions: Kirman's (1993) "ant" model and its financial interpretations are dealt with in Section 3.4.2, and the models of interacting speculators proposed by Lux and Marchesi are featured in Section 3.4.3. More complicated models with a lattice-based topological structure are considered in Section 3.4.4. Section 3.5 concludes and tries to provide an assessment of the state of this new approach vis-à-vis other approaches in the broader area of behavioral finance.

3.2. THE STYLIZED FACTS OF FINANCIAL DATA

3.2.1. Martingales, Lack of Predictability, and Informational Efficiency

The one empirical feature that has become a core ingredient of theoretical models and that a broad literature attempts to explain is the martingale property of financial prices. It can be stated simply as:

$$E[P_{t+1}|I_t] = P_t \tag{3.1}$$

where P_t denotes the prize of the asset at time t and I_t is the available information set at date t. As a consequence, ownership of the asset can be viewed as a *fair game* with expected pay-off equal to zero:

$$E[P_{t+1} - P_t | I_t] = 0 \tag{3.2}$$

and the realized price change is a random variable driven by the news arrival process that leads to a price at time $t + 1$ after new information arrives that differs from its date t conditional expectation:

$$P_{t+1} - P_t = P_{t+1} - E[P_{t+1} | I_t] = \varepsilon_t \tag{3.3}$$

with $E[\varepsilon_t] = 0$ due to the stochasticity of new information arrivals. With returns being defined as $r_{t+1} = \frac{P_{t+1} - P_t}{P_t}$ and

$$E[r_{t+1} | I_t] = \frac{E[P_{t+1} | I_t] - P_t}{P_t} \tag{3.4}$$

the randomness of price changes carries over to this quantity as well.

A glance at any financial returns series reveals that the lack of predictability of price changes, $E[\varepsilon_t] = 0$, is at least a very reasonable characterization of the data; at first view, the increments of high-frequency returns appear like random fluctuations about a mean value close to zero with no apparent asymmetry between positive and negative realizations (see the well-known examples exhibited in Figure 3.1). The random nature of price changes is explained by the *efficient market hypothesis* (EMH) as the imprint of informational efficiency, that is, all currently available information of any relevance in evaluating the asset in question is already incorporated in the market price. Therefore, only new information could lead to price changes, which then would be the immediate and unbiased reaction of the market on any new information item. It is worthwhile to note that the EMH is a theory about *market outcomes* and originally had only suggested a relatively vague concept of how this macroscopic outcome might emerge from the microscopic interaction of a diversity of agents in the marketplace. This missing behavioral underpinning has been added by the literature on market microstructure and asymmetric information (see Glosten and Milgrom, 1985; Kyle, 1985; O'Hara, 1995), which shows how the private interaction of some agents will be revealed via their trading activity and how the market over time approaches a state of complete revelation of any formerly private information. Since what is revealed of the private information of better-informed agents becomes public information, these models support the so-called *semi-strong* version of the EMH that specifies I_t as the information available to all market participants. The stronger version with I_t including even all private information is only valid asymptotically, that is, after an infinite number of trading rounds involving the better-informed agents. In these seminal contributions, the price process in the repeated trading scenario can be shown to follow a martingale.

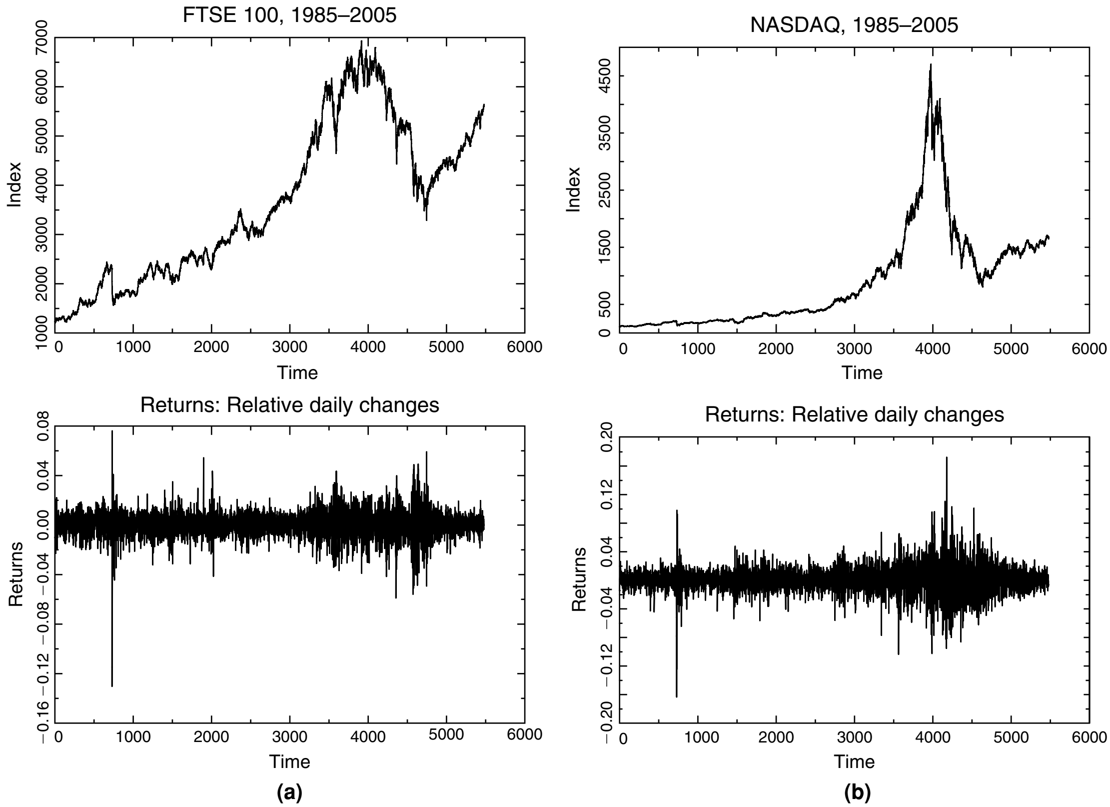

FIGURE 3.1 Two typical financial time series: the index evolution and daily increments of the (a) UK FTSE 100 index and the (b) US NASDAQ.

Traditional finance thus provides a well-established body of literature offering a plausible generic explanation of the martingale property that can be supported by microeconomic models of price formation under various institutional settings.

Of course, there are many qualifications to be made from different angles: First, the lack of predictability of price changes has been questioned in a large number of papers; variance-ratio tests try to recover long swings in stock prices, trading rules have been tested in-sample and out-of-sample for their ability to track hidden patterns in price records, and artificial intelligence and data-mining techniques have been used for the same purpose (see Taylor, 2005, for a comprehensive review). On the theoretical front it is well known that allowing for risk aversion instead of the assumed risk neutrality of early microstructure models leads to efficient markets without the martingale property (see Leroy, 1989).

We do not attempt to go into detail on any of these points in this chapter, but we simply note that markets might only be close to martingale behavior and that there might be good reasons to expect them to deviate from perfect efficiency and complete randomness of price movements.

The point we want to emphasize is rather that the traditional framework, though providing a generic explanation for one of the striking features of Figure 3.1, leaves unexplained the remaining set of similarly ubiquitous findings.

3.2.2. Fat Tails of Asset Returns

Figure 3.2 highlights the distributional properties of the returns series exhibited in Figure 3.1. A very natural benchmark for characterizing the unconditional distribution

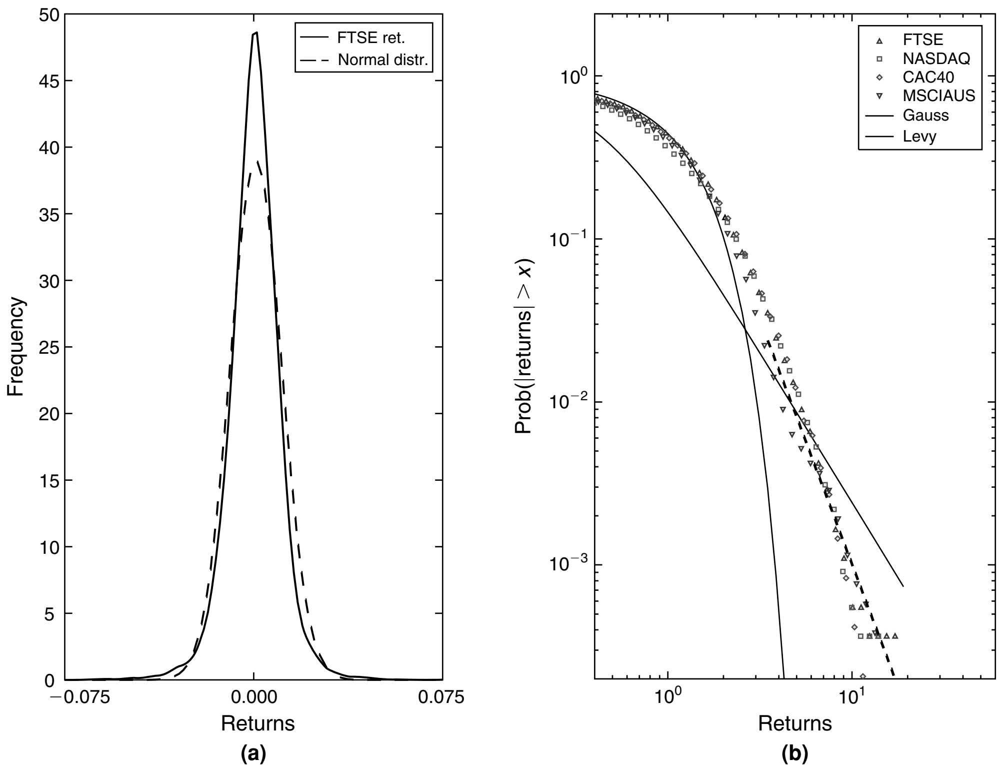

FIGURE 3.2 Distributional properties of returns: (a) Exhibits the distribution of returns of the FTSE (smoothed via a Gaussian Kernel estimator) in comparison to the Normal distribution with the same mean and standard deviation; (b) shows the empirical complement of the cumulative distribution of absolute returns for four financial indices (FTSE 100, NASDAQ, CAC 40, and MSCI Australia). Note that under the first case of hyperbolic tail behavior in Eq. 3.7, this amounts to $\text{Prob}(|\text{returns}| > x) \sim x^{-\alpha}$. In all cases, note the typical *preasymptotic* distribution: the tails are more elongated than under a Normal distribution, but thinner than under a Levy stable distribution. The broken line in (a) illustrates the decay factor of −3 of the "universal cubic law" claimed by Gopikrishnan et al. (1998). Though not an equally good fit for all indices, the inverse cubic decay is close to the empirical behavior of all financial assets.

is the Normal distribution. As has been well known at least since the early 1960s (Fama, 1963; Mandelbrot, 1963), however, the Normal distribution provides a very poor fit to financial returns. As shown in Figure 3.2, empirical distributions are, in fact, quite nicely bell shaped and symmetric but typically have more probability mass in their center and tails than the Normal distribution. Though the predominance of small fluctuations (smaller than expected under the Normal with the same standard deviation) is apparent from the histogram, the importance of fat tails can be better grasped from a comparison of empirical returns with simulated Gaussians (see Figure 3.3). As can be seen, positive and negative events exceeding, for example, five times the sample standard deviation occur quite regularly in empirical data, whereas they would have negligible probability in a Gaussian market. Table 3.1 provides some evidence that this behavior is truly

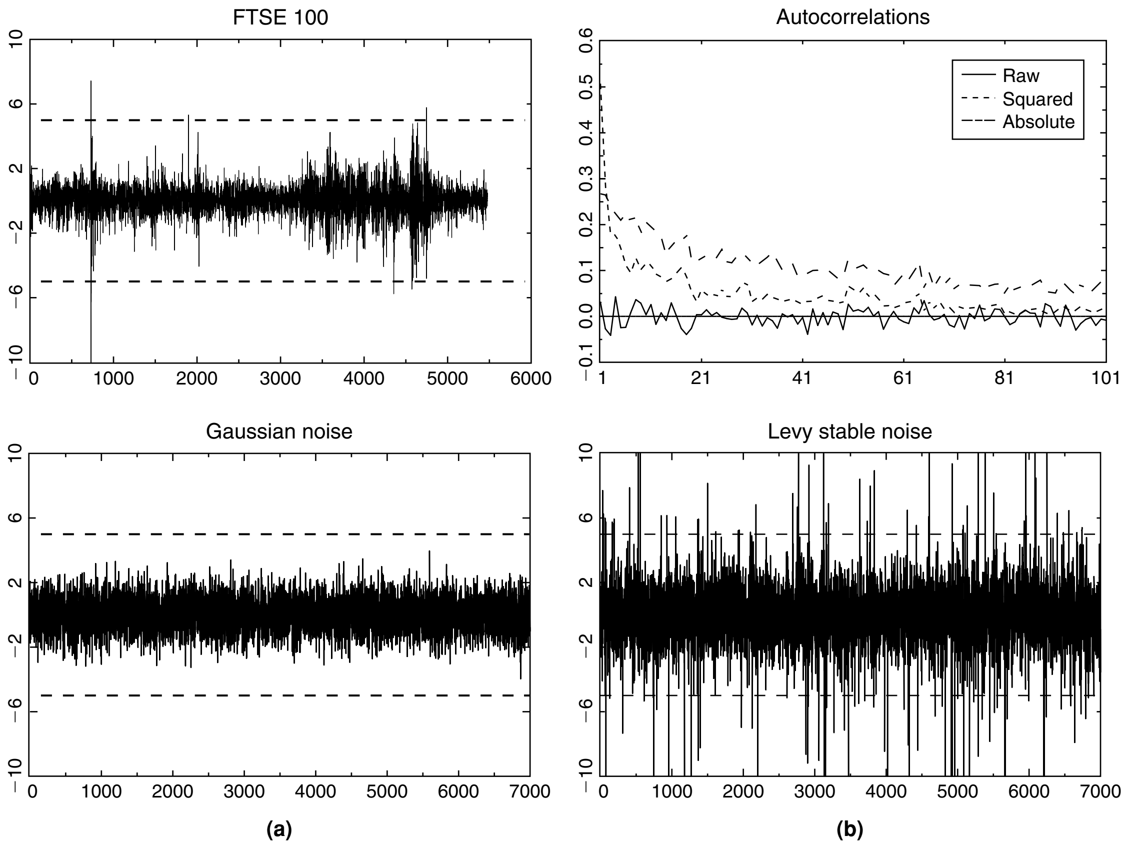

FIGURE 3.3 Returns of FTSE 100 compared to simulated Gaussian noise and Levy stable noise without temporal correlation. For better comparability, all three series have been rescaled so that their sample variances are equal to unity. The tail parameter of the Levy distribution has been set equal to 1.7, a typical estimate for stock returns. The top diagram in (b) shows the pronounced, hyperbolically decaying autocorrelations of squared and absolute returns which indicate volatility clustering and time-variation of the degree of fluctuations.

TABLE 3.1 Kurtosis Statistics and Maximum Likelihood Estimates for the Pareto Tail Index Characterizing the Extremal Law $G_{1,\alpha}$ and Tail Distribution $W_{1,\alpha}$

	$\alpha_{2.5\%}$	$\alpha_{5\%}$	$\alpha_{10\%}$	κ
FTSE 100	3.21	3.06	2.80	11.10
	(2.68, 3.75)	(2.70, 2.56)	(2.56, 3.03)	
NASDAQ	3.31	3.23	2.69	7.36
	(2.76, 3.86)	(2.85, 3.62)	(2.46, 2.91)	
CAC 40	3.64	3.17	2.87	4.72
	(2.99, 4.29)	(2.77, 3.57)	(2.62,3.13)	
MSCI Aus	3.16	3.61	3.17	46.21
	(2.62, 3.71)	(3.17, 4.05)	(2.90, 3.44)	

Note: Data are the same as in Figure 3.2, with daily sampling frequency and time horizon 1985 to 2005. The estimates are given for three different sizes of the tail region (2.5%, 5%, and 10%) with asymptotic 95% confidence intervals shown in brackets. All results are in good overall agreement with a "universal cubic law" as postulated in the pertinent literature. The tendency for a decrease of the estimated coefficient with increasing tail size is usually seen as reflecting contamination of tail data with observations from the center of the distribution. With estimated tail indices significantly below 4, the fourth moment would not exist. The expected divergence of the kurtosis statistics would lead to unstable estimates in finite samples that increase with sample size.

universal: For a number of assets, it lists the kurtosis statistics and the tail index (details follow) for various definitions of the tail region of the data. Kurtosis is defined as the standardized fourth moment:

$$\kappa = \frac{1}{N}\sum_{t=1}^{N}\left(\frac{r_t - \bar{r}}{\sigma}\right)^4 - 3 \tag{3.5}$$

with $\bar{r}$ the mean value and σ the standard deviation of the sample. The benchmark of $\kappa = 0$ characterizes the Normal distribution and separates platykurtic ($\kappa < 0$) from leptokurtic ($\kappa > 0$) distributions. Leptokurtosis (at least for unimodal distributions) has the visual appearance of higher peaks around the mean and heavier tails than the Normal, which is the kind of shape that we *always* encounter in returns.

The finding of non-Normality and leptokurtosis as universal properties of financial returns has spurred a long-lasting debate on the appropriate stochastic model for the innovations (see Eq. 3.3). Stochasticity of returns quite naturally leads to the hypothesis that aggregate returns should obey the Central Limit Law and, hence, would have to approach the Normal distribution. However, despite their aggregation over large numbers of high-frequency price changes, daily returns are apparently non-Normal.

Mandelbrot (1963) and Fama (1963) provided a solution for this conundrum, evoking the *generalized central limit law*. The basic tenet of this more general convergence theorem is that the distribution of sums of random variables converges to an appropriate member of the family of Levy stable distributions. If the second moment exists, the pertinent member is the Normal distribution (as a special parametric case of the Levy stable distributions). If the second moment does not exist, other members of this family are the limiting distributions of sums. In particular, these alternative limiting distributions are all leptokurtic and share the typical deviation of the empirical histogram from the Normal distribution. Whereas the Mandelbrot/Fama hypothesis has motivated a large literature on parameter estimation and practical application of Levy distributions, it eventually turned out that these models would largely overstate the frequency of large returns (see Figures 3.2 and 3.3 for illustrations). Much of this evidence is owed to the introduction of concepts from statistical extreme value theory in empirical finance. The key concept of extreme value theory is the so-called *tail index* that allows a classification of the extremal behavior of empirical data and distribution functions (see Beirlant, Teugels, and Vynckier, 1996). The basic result is the classification of extreme values (maxima and minima) from IID random variables with continuous distributions. Denoting by $M = \max(x_1, \ldots, x_n)$ the maximum of a sample of observations $\{x_i\}$, it can be shown that after appropriate change of location and scale, the limiting distribution of M belongs to one of only three classes of distribution functions. More formally, the distribution of the appropriately normalized maximum, $\text{Prob}[a_n M + b_n \leq x]$, converges to one of the following *extreme value distributions* (GEVs):

$$G_{1,\alpha}(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}$$

$$G_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x \leq 0 \\ 1 & x > 0 \end{cases} \tag{3.6}$$

$$G_3(x) = \exp(-e^{-x}) \quad x \in \Re$$

From this typology of extremal behavior, a similar classification of the underlying distribution's asymptotic behavior in its outer parts, that is, tails, can be inferred. Namely, denoting probabilities $\text{Prob}(x_i \leq x) \equiv W$ it follows directly from the classification of extremes in Eq. 3.6 that if the maximum of a distribution follows a GEV of type $j (j = 1, 2, 3)$, its upper tail asymptotically converges to the pertinent distribution from the following list:

$$\begin{aligned} W_{1,\alpha} &= 1 - x^{-\alpha}, x \geq 1 \\ W_{2,\alpha} &= 1 - (-x)^{\alpha}, -1 \leq x \leq 0 \\ W_3 &= 1 - \exp(-x), x \geq 0 \end{aligned} \tag{3.7}$$

These three types of tail behavior can be described as hyperbolic decline ($W_{1,\alpha}$), distributions with finite endpoints ($W_{2,\alpha}$), and exponential decline (W_3). To nest all three alternatives, one can integrate the three limit laws into a unified representation:

$$W_\gamma = 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma} \tag{3.8}$$

with $\gamma = 1/\alpha$ ($\gamma = -1/\alpha$) in the cases $W_{1,\alpha}$ and $W_{2,\alpha}$ and W_3 being covered as the limit $\gamma \to 0$ (σ is a parameter for scale adjustment). Estimation of the *tail index* α allows us to determine whether a particular distribution falls into class 1, 2, or 3. These estimates would allow us to assess whether certain distributional hypotheses are in conformity or not with the empirical behavior. For example, an empirical γ ($= 1/\alpha$) significantly above 0 would allow rejection of the Normal distribution as well as any other distribution with exponentially declining tails. The indication of hyperbolic decline would also exclude a finite endpoint as implied by $W_{2,\alpha}$ type distributions. Needless to say, the estimated α would be an extremely valuable tool in financial engineering as it could be easily used to compute the probability of large losses and gains (see Lux, 2001).

Table 3.1 shows that, with some variation depending on the selection of the tail size, empirical estimates hover within the interval of about 2 to 4. Ninety-five percent intervals from the asymptotic distribution of the pertinent maximum likelihood estimates allow us to demarcate even more sharply the set of distribution functions that would or would not be in harmony with such extremal behavior. As an important consequence, the family of Levy-stable distributions proposed by Mandelbrot (1963) and Fama (1963) would have heavier tails than the empirical records, with their α being restricted to the interval $[0, 2]$. Any empirical α significantly above 2 (as we mostly find it) would, therefore, speak against the Levy stable model (which, as a consequence, would hugely *overstate* the risks of large returns). On the positive side, an admissible candidate for the unconditional distribution would be the Student t whose degrees of freedom are equal to its tail index so that it could be tuned in a way to conform to empirical shapes of return distributions. Fergussen and Platen (2006) show that for a variety of stock indices the parameter estimates of a very general family of distributions (the generalized hyperbolic distributions) cluster in the neighborhood of a Student t with 3 d.f.

What implications does this phenomenological characterization have for theoretical models? First, from the viewpoint of the efficient market theory, price increments only have to be random. The innovations in Eq. 3.3 could, therefore, be drawn from a Student t as well as any other distribution function that meets the minimum requirement $E[\varepsilon_t] = 0$. Since ε_t reflects the news arrival process, its realm is outside economics and the EMH is agnostic as to what the joint distribution of all relevant news items might look like. However, there is a more subtle issue here: Returns over longer time intervals are aggregates of high-frequency returns, at least under continuous compounding, that is, for $r_t = \ln(P_t) - \ln(P_{t-1})$ and approximately so for $r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$. If all high-frequency returns are reflections of IID news arrival processes, their aggregates should converge toward the Normal distribution, irrespective of the underlying distribution of single high-frequency returns. This is simply a consequence of the central limit law. One might argue that at the level of daily data (the time horizon we have investigated previously), returns in liquid markets are already sums of thousands of intradaily price changes so that we should have gone through the better part of the convergence toward the Gaussian shape at this level of time aggregation. Nevertheless, as we have seen, daily returns are quite different from Normally distributed random variates.[2] With higher levels of time aggregation (e.g., monthly returns), we indeed get closer to the Gaussian, as would be expected from the central limit law.

The intriguing aspect about this phenomenology is that if we look at financial data of a particular time horizon (e.g., daily), we find a kind of *universal preasymptotic behavior* that seems to be independent of location, time, and details of the market structure. To appreciate this universality of the approximately *cubic law of asset returns* (Gopikrishnan et al., 1998), note that it appears to apply to practically all types of financial markets, for example, various developed stock markets, foreign exchange markets, precious metals, and emerging markets (Jansen and de Vries, 1991; Longin, 1996; Koedijk et al., 1990; Lux, 1996; Rockinger and Jondeau, 2003). The relevant news arrival processes might be quite different for all these markets. At least there is no *a priori* reason to assume that they should all obey a roughly identical distribution of news. Furthermore, one might argue that the velocity of efficient information processing in the trading process might have increased over time due to technical advances such as electronic trading. Nevertheless, we have no indication that the shape of the return distribution has undergone any remarkable changes over the past decades, reflecting an increase in information transmission.

It seems that the set of potential events (summarized in the distribution of returns) during a day in a financial market is always pretty much the same, irrespective of whether trading is organized via order-driven and quote-driven systems, whether shares are traded on the floor or electronically, and broadly independent of the size of the market. Despite the agnostic view of distributional properties by the EMH, we might feel somewhat uncomfortable about the apparent *universality* of the type of randomness of price changes. If the return distribution is that robust, additional factors besides new information in the trading process might be held responsible for this particular outcome of the market process. However, if this were the case, the EMH would not offer a full explanation of financial price movements and prices would *not* solely reflect new information.

One might, then, argue that the universality of distributional properties of asset returns should have its behavioral roots within the trading process and needs to the explained by the way human subjects interact in financial exchanges. Section 3.3 will further pursue this avenue by considering it from the perspective of "scaling" theory developed in the natural sciences. Before we turn to this unfamiliar approach, we expand on other ubiquitous regularities in the following subsections.

[2]This finding had actually motivated the proposal by Mandelbrot and Fama of the Levy stable distributions. According to a generalized central limit law, these are the limiting distributions of sums of random variables with *infinite* second moment (whereas with a finite second moment, we are back at the classical central limit law). Unfortunately, the data also speak against the Levy stable hypothesis.

3.2.3. Volatility Clustering and Dependency in Higher Moments

The martingale property of financial prices implies that price differences define a martingale difference process and are, thus, uncorrelated. In empirical time series, one typically finds marginally significant positive or negative autocorrelations at the first few lags for stock and currency returns, respectively. These are, however, believed to reflect the microstructural characteristics of particular markets and the way in which prices are recorded; in stock markets, small positive autocorrelations are probably due to infrequent trading for single stocks and certain common news factors of importance for the individual components of stock indices. In foreign exchange markets, bounces between the bid and ask price for currency quotes lead to negative correlation of recorded transaction prices. Since these autocorrelations, though statistically significant, could mostly not be exploited via pertinent trading strategies, they are usually not classified as strong evidence against the EMH.

However, while almost uncorrelated, asset returns are not IID stochastic processes. Another glance at Figure 3.1 reveals that although the ensemble of returns over a longer horizon leads to fairly similar distributions across different sets of data, on shorter time scales we encounter less homogeneous behavior. The comparison of empirical returns and simulated Gaussian and Levy stable data in Figure 3.3 makes the difference particularly transparent; though the latter have a very uniform degree of fluctuations, the former switch between periods of tranquility and more turbulent episodes. The returns-generating process is, thus, characterized by nonhomogeneity of its higher moments. This variability in the extent of fluctuations is actually the reason for the introduction of the concept of a martingale process in financial economics because it makes no requirement on the noise term except for $E[\varepsilon_t] = 0$.[3]

The lack of IID properties is also reflected in autocorrelations of simple transformations of returns. Considering various powers of absolute returns $\{|r_t|^\lambda\}$, one typically observes much higher and longer-lasting autocorrelations than for the raw series. Figure 3.3 illustrates this finding for the most frequent choices $\lambda = 1, 2$. As can be seen, there is strong dependence in these higher moments. Since powers of absolute returns can be interpreted as measures of volatility (because they all drop the sign and only preserve the extent of fluctuations), these results indicate a high degree of predictability of volatility (in the absence of significant predictability of the *direction* of price movements).

The volatility clustering phenomenon has been known for a long time, but models covering this feature appeared first with Engle's (1983) seminal proposal of the ARCH framework that spurred a plethora of models with nonlinear dependency in second movements (see Taylor, 2005, for an overview). Although most early literature had considered only the second moment ($\lambda = 2$), Taylor (1986) pointed out that the first absolute moment has even more pronounced dependence than the second. Ding, Engle,

[3]In contrast to the more restricted concept of a random walk, which would require a constant variance $\sigma_\varepsilon^2 = Var[\varepsilon_t]$ of the fluctuations.

and Granger (1993) discuss a whole range of positive λs and find that the highest degree of autocorrelation is typically found for $\lambda \approx 1$. Meanwhile, this hierarchy of strengths of dependency also counts as an established stylized fact (Lobato and Savin, 1998).

An important facet of the empirical findings on higher-order dependencies is the distinction between short memory and long memory in autocorrelation structures. Short-memory processes are characterized by exponentially decaying ACF functions (ARMA models as well as GARCH models are standard examples), but a long-memory process has hyperbolically decaying autocorrelations, which implies a much slower decay with long-lasting after-effects of innovations. Inspection of Figure 3.3 suggests that the autocorrelations of absolute and squared returns are examples of hyperbolic rather than exponential decline. Indeed, if one considers very long series, the autocorrelations stay significant over perplexingly long horizons: For daily S&P 500 data over the period 1928–1990, Granger and Ding (1996) report significant autocorrelations over 2500 lags—that is, more than 10 years! The decay of autocorrelations of squared and absolute returns is, in fact, indicative of a hyperbolic decline. This implies that, for example, covariances of absolute returns, would decay according to:

$$E[|r_t r_{t+\triangle t}|] \sim \triangle t^{-\beta} \qquad \textbf{(3.9)}$$

Processes with long-memory or long-term dependence have properties very different from those that only display short memory (see Beran, 1994). In particular, the variance of the sample mean decays to zero at a rate slower than n^{-1} and the spectral density diverges at the origin. Findings of long memory in the mean of certain series have motivated the development of fractional Brownian motion and autoregressive fractionally integrated processes, whereas the finding of long-term dependence in the second moment of financial data has inspired the development of pertinent extensions of GARCH type and stochastic volatility models (see Baillie, Bollerslev, and Mikkelsen, 1996).

3.2.4. Other Stylized Facts

One important recent addition to the set of time-series characteristics of financial data is long memory of trading volume. It has been known for quite some time that volume is highly contemporaneously correlated with volatility. This pronounced comovement might suggest that both series have more common characteristics. Convincing evidence for long-term dependence in volume has been presented in Lobato and Velasco (1998), although the authors also point out that volatility and volume do not share exactly the same degree of long-term dependence (i.e., have different decay parameters β).

Recent work on U.S. high-frequency stock market data has come up with the additional finding of fat tails in the unconditional distribution of transaction volume (Gopikrishnan et al., 2001) and the number of trades within a time interval. Gabaix et al. (2003) provide a theoretical framework in which they combine these findings, the power law of returns, and a Zipf's law for the size distribution of mutual funds within a choice-theoretic setting for the trading activity of large investors. However, empirical evidence for the new regularities is so far restricted to the U.S. datasets investigated by these authors.

3.3. THE STYLIZED FACTS AS "SCALING LAWS"

The neglect of almost all the prominent features of asset prices except for their martingale property by the efficient markets paradigm is not too hard to explain. If one shares the view of informational efficiency being reflected in the unpredictability of price changes, the price increments are simply one-to-one mappings of the news arrival process. As noted, neither fat tails nor long-term dependence as properties of price changes are, therefore, in contradiction to the EMH. One might, however, be aware that as a consequence from accepting the empirical facts and the validity of the traditional EMH view, one would have to concede that "news" in all times and all places seems to come with the very same underlying distribution.

Interestingly, scientists with a different background who have stumbled over one of the huge data sets from financial markets have typically arrived at very different conclusions after detecting the preceding "stylized facts." Since financial data represent the largest available records of human activity, they have indeed attracted curiosity from various other disciplines. A strong recent current is that of physicists engaging in empirical analysis and theoretical modeling of financial markets. The reaction of these researchers to the well-known stylized facts of empirical finance was entirely homogeneous and totally different from the received viewpoint recalled previously: Natural scientists saw these as imprints of a complex system with a large number of interacting microscopic entities. As Stanley et al. (1996) wrote in an influential early contribution pointing out this viewpoint:

> Statistical physics has determined that physical systems which consist of a large number of interacting particles obey universal laws that are independent of the microscopic details. This progress was mainly due to the development of scaling theory. Since economic systems also consist of a large number of interacting units, it is plausible that scaling theory can be applied to economics.

This statement basically argues that since the statistical properties of financial markets are similar to those of certain physical (or biological) systems, the explanation of these characteristics should draw on similar general principles. This assertion has (at least) three components that we can consider in turn for their empirical validity or plausibility.

1. *Financial stylized facts are analogous to the scaling laws that play a prominent role in statistical physics.* We have expended some effort in outlining how empirical finance had arrived at very parsimonious characterizations of the fat tails and clustered volatility of returns. Both the unconditional distribution of large returns (Eq. 3.7) and the conditional dependence structures of their fluctuations (Eq. 3.9) can be expressed by hyperbolic decay rates. Such hyperbolic distributional

characteristics are, however, exactly what is denoted as a power or scaling law in statistical physics. As far as the existence of these "laws" counts as well established in empirical finance, financial data in this descriptive sense share the scaling laws of various natural records.

2. *Scaling laws (stylized facts) should typically be robust (universal) and should, therefore, hold for similar phenomena independent of the microscopic details.* In finance one could interpret these apparently unimportant microscopic details as the particular institutional details of the market microstructure (floor trading vs. electronic trading, quote vs. order-driven markets, and so on). In fact, though many other empirical findings do somehow depend on the microstructure, the hyperbolic scaling laws are those features that can be found everywhere, for example, the cubic law of large returns seems to govern both stock markets as well as foreign exchange markets with their very different organization of the trading process.
3. *Scaling laws are the signature of systems with a large ensemble of interacting units that emerge from the interaction of these subunits (particles, molecules) and are only dependent on a few basic principles of interaction.* Although this is simply an observation across many categories of dynamic processes in physics and other areas of natural science (examples are turbulent flows or evolutionary processes in biology), it seems harder to accept this viewpoint for manmade systems. It appears to imply that we have to disregard the importance of individual rational choice, which is one of the basic tenets of economic theory. We would certainly not deny the rationality (or, from a behavioral perspective, the attempt toward rational behavior) of economic agents, but we could argue that the diversity of micromotives, preferences, endowments, access to information, and degrees of rationality and deliberation of these agents could be better captured by a statistical approach than by the optimization of one representative agent. It might well be that in the presence of this large ensemble of heterogenous agents, a few basic principles of interaction can be found that exert a dominating influence on the macroscopic market behavior and that prevail in more or less the same way in different institutional settings (microscopic details). It would, then, be the task of a theory motivated by the analogy between scaling in physics and finance to show that this possibility can be substantiated by sensible stochastic models of asset price dynamics. The relevant literature is reviewed in the next section.

Some words of caution on the "scaling approach" have to be added: There might well be an exaggeration of both the statistical basis and the potential implications of scaling laws as signatures of complex dynamics in the pertinent literature. As it concerns the statistical validity, detection of scaling in the natural sciences is typically based on apparent linearity in some kind of log-log plot such as Figure 3.2. With the necessity of "binning" the data (i.e., grouping it into intervals) and the violation of the independence assumption in the linear regression, this approach appears questionable from a methodological viewpoint and has often been criticized by economists (Brock, 1999; Durlauf, 2005; and Gallegatti et al., 2006): The ubiquitous declaration of statistical objects as fractal, self-similar, or scaling has also been attacked in a paper in *Science*.

The authors (Avnir et al., 1998) had surveyed all 96 articles in the *Physical Review* journals over the period 1990 to 1996 that contained some empirical scaling analysis of natural or experimental time series. They conclude that the "... scaling range of experimentally declared fractality is extremely limited, centered around 1.3 order of magnitude ..." whereas a true self-similar or fractal object in the mathematical sense would require infinitely many orders of power-law scaling. They found power-low behavior in most of these claims of studies quite questionable because of the limited range of observations.

A number of papers also point out that one could obtain "spurious" or "apparent" scaling behavior for processes without a "true" asymptotic power law. Gielens et al. (1996) show that one can always find local alternatives to fat-tailed distribution that possess thin tails (i.e., that decline exponentially), while tail index estimates would indicate a power-law behavior. The temporal scaling characteristics (long-term dependence of absolute moments) could be obtained as a spurious outcome of certain specifications of GARCH models (Crato and de Lima, 1994), stochastic volatility models (LeBaron, 2001), regime-switching processes, or even uncorrelated stochastic processes with heavy tails (Barndorff-Nielsen and Prause, 2001). However, mostly these alternatives would require a certain degree of fine-tuning of parameters in order to "fool" the pertinent statistical tests. Given enough flexibility of parameter selection, it would always be possible to design a local alternative to a process with power-law characteristics that has no "true" scaling behavior but comes arbitrarily close to scaling and could, then, not be distinguished from a generic power-law mechanism with finite samples.

An example is the Markov-switching multifractal model introduced by Calvet and Fisher (2001), which has "long memory over a finite interval" that could be made arbitrarily long by appropriate choice of the specification. This model had indeed been designed as a well-behaved stochastic process that provides a close resemblance to the statistically more cumbersome first vintage of multifractal models of asset returns with true scaling behavior (see Mandelbrot et al., 1997). The ubiquity of fat tails and long memory for financial data might, however, be viewed as support for models that have these features generically rather than apparent scaling for particular sets of parameter values.

As concerns the rigor of statistical analysis and the sample sizes of empirical data for which "scaling" has been declared, pertinent studies in finance are in a *better* position than most of the studies in natural sciences criticized by Avnir et al. (1998). First, financial econometricians routinely apply more rigorous methods than log-log plots. Most of the research on fat tails in finance is based on the theoretical concepts of extreme value theory and has adopted state-of-the-art estimators of the tail index (mostly without reference to the concept of scaling). Similarly, research on temporal dependence has also used more refined methods from stochastics (see Lux and Ausloos, 2002, for a comparison of the tools used by physicists and financial econometricians). Second, as for the sample sizes, the literature typically started out with daily recorded series but has moved on to the immense databases of intradaily, high-frequency returns. Both the findings on fat tails and long-term dependence of volatility in daily data are confirmed for intradaily records (see Abhyankar et al., 1995; Dacorogna et al., 2001; and Bollerslev and Wright, 2000).

Another concern might be the alleged relationship between power laws and "complex" interactions of heterogenous subunits. The evidence for such a relation is mainly illustrative in nature: Physics and biology offer a variety of examples in which the nonlinear interactions of elementary units result in overall system characteristics that can be described by power laws. A famous case is the leading example of self-organized criticality; dropping grains on piles of sand (or other materials such as rice) always leads to a power-law distribution of avalanches that can be explained by a stochastic process of the change of local gradients (Jensen, 1998). However, the usefulness of diffuse labels of complexity theory has been critically discussed (Horgan, 1995). It might also be noted that there exist some simple explanations for power laws: Power-law distributions could be generated via a combination of exponentials, by taking the inverse of quantities that themselves obey harmless distributions, or by splitting processes, among others (see Newman, 2004). As has been demonstrated by Granger (1980), long memory in aggregate data could result from the aggregation of heterogenous individual behavior (a principle that has recently inspired a new branch of empirical literature in political science; see Box-Steffensmeier and Smith, 1996). However, none of the simple generating mechanisms has ever been proposed as a source of power laws in financial data, and aggregation of individual behavior might not be inconsistent with the view of financial markets as a system of interacting agents. Therefore, it seems worthwhile to explore the "complexity" approach that views scaling as the consequence of phase transitions and critical phenomena.

3.4. BEHAVIORAL ASSET-PRICING MODELS WITH INTERACTING AGENTS

From the late 1980s and into the early 1990s, behavioral approaches to financial markets gained momentum. The literature on excess volatility and overreaction of asset prices to news suggested that psychological mechanisms and boundedly rational behavior might provide explanations for these and other mysterious "anomalies." At the same time surveys of trading strategies and expectation formation mechanisms of real-life traders pointed to the importance of technical trading and adaptive expectations (Allen and Taylor, 1990; Taylor and Allen, 1992). The dollar bubble of the early 1980s was believed to have been at least partially due to positive feedback trading (Frankel and Froot, 1986), and this perception gave rise to new interest in models of interacting groups of chartist and fundamentalist speculators (Beja and Goldman, 1980; Day and Huang, 1990). These models were framed as systems of difference or differential equations that contained the asset price as well as some characteristics of investors as state variables.

Some of the pertinent models assumed permanent market clearing; others used a sluggish price adjustment rule as a proxy for market-making activities in the presence of excess demand (ED). While excess demand functions of the various groups of traders either could be formulated in an *ad hoc* fashion or were derived from particular utility functions, traders were typically not assumed to be fully rational, since neither of these groups properly takes into account the effect of its own trading activity on subsequent price movements. In a sense (detailed in this section), these contributions were already motivated by the idea to explain marketwide phenomena as emergent characteristics from complex interactions, but they restricted the level of disaggregation to a small number of behavioral types with complete homogeneity within groups.

3.4.1. Interaction of Chartists and Fundamentalists and Nonlinear Dynamics of Asset Prices

The seminal paper by Beja and Goldman (1980) provides a simple example of the legacy of models of chartist–fundamentalist interaction. Beja and Goldman assume simple *ad hoc* functional forms for excess demand of fundamentalists and chartists. Fundamentalists' excess demand depends on the difference between the fundamental value P_f (assumed to be known to them) and the current market price P_t:

$$ED_f = a(P_f - P_t) \tag{3.10}$$

where a is a coefficient for the sensitivity of fundamentalists' excess demand to deviations of the price from the underlying fundamental value. Assuming an expected reversal of the market price toward P_t, together with constant risk aversion and constant expected volatility, such a function could also be derived from myopic utility maximization using a mean-variance framework or a negative exponential CARA utility function together with Normally distributed expected price changes.

This format of fundamentalists' excess demand is pretty standard in the literature and can be found in early contributions such as Baumol (1957), but there is more variation in this literature in the formulation of chartists' excess demand. The particular hypothesis employed by Beja and Goldman (1980) is that their excess demand depends on the expected price change π (i.e., expected capital gains or losses):

$$ED_c = b\pi \tag{3.11}$$

where b again captures the sensitivity of the order flow of this group to expected gains or losses. In a continuous-time framework, π is the *subjective* expectation of the infinitesimal price change $\frac{dp}{dt}$. Beja and Goldman (1980) invoke a market-maker mechanism to justify sluggish Walrasian price adjustment:

$$\frac{dP}{dt} = P'(t) = \lambda(ED_f + ED_c) = \lambda(a(P_f - P_t) + b\pi) \tag{3.12}$$

with λ the price adjustment speed. Though this is a phenomenological characterization without microeconomic motivation from the optimization problem of a market maker, one may note that it closely resembles the micro-founded price adjustment rules of the literature on price formation under asymmetric information (see, e.g., Kyle, 1985).

Given the trading strategies of the two groups of investors, price changes result endogenously from the total imbalance between demand and supply so that the chartists' expectations might be confirmed or not. In the presence of a deviation of expected from realized price movements, chartists are assumed to adaptively adjust their expectations:

$$\frac{d\pi}{dt} = \pi'(t) = c(P'(t) - \pi) \tag{3.13}$$

where c is a parameter for the speed of adaptation of expectations.

Neither group is characterized by rational expectation formation; chartists react adaptively by assumption, so they will hardly ever correctly predict price changes. Fundamentalists neglect the existence of chartists and their influence on price changes, so even if the price reverts toward its fundamental value (which might not be guaranteed), the speed of its reversal toward P_f might be different from the hypothesized adjustment coefficient of Eq. 3.10.

The model of speculative activity by Beja and Goldman (1980) boils down to a system of two differential equations (Eqs. 3.12 and 3.13). It is a typical example of a large body of literature that formalizes speculative market dynamics as a system of difference or differential equations. In most cases, these models can be expressed as dynamic systems covering the market price, plus some group characteristics that undergo changes over time in response to the asset price dynamics. It is also quite characteristic of the broader literature in its main results. The interest of the authors of this and many subsequent contributions in this vein is mainly in the *existence* and *stability* of a fundamental equilibrium in the presence of nonrational speculative activity. It is easy to see that the conditions for existence of a stationary state of the joint dynamics of P and π, $P'(t) = \pi'(t) = 0$ leads to a dynamic equilibrium $P^* = P_f, \pi^* = 0$. The only possible steady state is, therefore, obtained if both the price equals its fundamental value and chartists expect no further price changes. In this case excess demand of both groups of traders equals zero and the price remains unchanged. It is slightly more demanding to arrive at results on the stability or instability of this steady state. Applying the standard stability criteria for systems of autonomous differential equations, we find that the system converges asymptotically toward its steady state if the following necessary and sufficient condition is met:[4]

$$a\lambda + c(1 - b\lambda) > 0 \tag{3.14}$$

This condition yields the following plausible insights:

- High sensitivity of fundamentalists' (chartists') excess demand is stabilizing (destabilizing).
- Whether increased price adjustment speed is destabilizing or not depends on the relative sensitivity of the chartists' and fundamentalists' demand schedules. Higher price adjustment speed has a more stabilizing (destabilizing) tendency if $a > (<) c \cdot b$.

[4]See Beja and Goldman (1980) for details.

- The influence of the speed of expectation adjustment of chartists is ambiguous: If $1 - b\lambda > 0$ the system is always stable independent of the value of c (since with $b < \frac{1}{\lambda}$ the market maker's price adjustment succeeds in reducing chartists' excess demand over time). Conversely, if $b > \frac{1}{\lambda}$, price adjustment triggers even higher order volumes by chartists due to pronounced bandwagon effects. In this case increasing adjustment speed c in their adaptive expectation formation would have a destabilizing tendency.

The second and third items illustrate that stabilizing and destabilizing features of particular chartists strategies could be subtle and could easily change under different specifications of their strategies. It is worthwhile to note that this model provides a potential explanation of certain empirical regularities. If stability condition (3.14) is satisfied and the eigenvalues of the dynamic system are complex conjugate numbers (which happens in an open set of parameter values), the model exhibits overshooting and subsequent mean reversal in the presence of new information. As an illustration, assume that the fundamental value of the asset increases from $P_{f,1}$ to $P_{f,2}$ (see Figure 3.4). Fundamentalists knowing of the increase of the intrinsic value will start buying shares. This leads to excess demand and exerts upward pressure on market prices. The increase of the asset price is interpreted as a positive trend by chartists who subsequently also start buying shares. Due to this noninformed source of additional demand, the price will overshoot its new fundamental value and fundamentalists

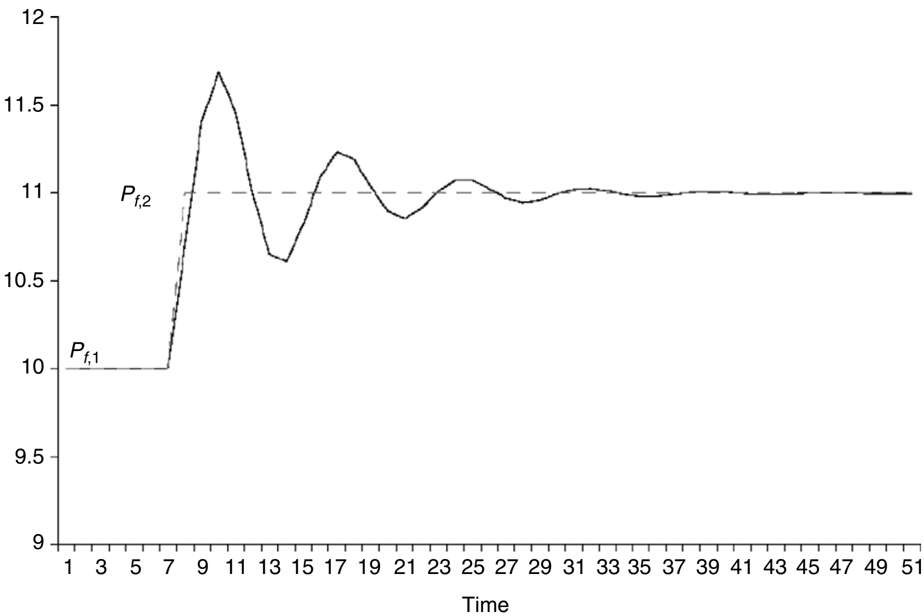

FIGURE 3.4 Overshooting and mean reversion of market prices after arrival of new information. Simulation results with $a = 0.7$, $b = 0.8$, $c = 0.9$, $\lambda = 1$, $P_{f,1} = 10$, and $P_{f,2} = 11$.

will switch from the demand to the supply side, leading to mean reversion toward $P_{f,2}$. In the following, the price will converge to its new fundamental value with damped oscillations.

If the stability criterion (3.14) is not satisfied, these oscillations would—because of strong feedback effects from chartists—display an increase rather than a decrease in amplitude. Since the model by Beja and Goldman is framed as a *linear* system of differential equations, there would be no limit to the divergence of the price from the underlying fundamental value. Of course, such a scenario is unrealistic, which essentially means that this baseline model is silent on the dynamics one would expect under local instability of the fundamental equilibrium.

Similar dynamic models such as the one proposed by Beja and Goldman with added nonlinear ingredients have been studied by a number of authors. Even prior to Beja and Goldman, Zeeman (1974) published a very similar model that assumed a nonlinear reaction function of chartists on observed price changes, which flattens out further away from the fundamental equilibrium. While Zeeman's interest is in the application of concepts from catastrophe theory (demonstrating the possibility of sudden stock market crashes), Chiarella (1992) showed that the Beja/Goldman model would generate periodic oscillations around the fundamental equilibrium in the unstable case if chartists' excess demand function gets sufficiently flat far from the equilibrium price.

Day and Huang (1990) consider a similar model formulated in discrete time whose "information traders" (equivalent to the previously discussed fundamentalists) trade more aggressively the farther the market price is from the fundamental value. With a strong reaction of chartists destabilizing the fundamental equilibrium, the assumed nonlinear reactions result in the same combination of centripetal and centrifugal forces, as in Zeeman (1974) and Chiarella (1992): Strong reaction of chartists prevents convergence to the fundamental equilibrium and generates bubble episodes of overvaluation or undervaluation of the asset. However, once the deviation of the market price from P_f becomes too large, either the chartists become more cautious or the fundamentalists step in more aggressively so that the price process does not diverge endlessly but rather reaches a turning point at which the attraction toward the fundamental value dominates over the positive feedback effect. The global dynamic is, therefore, *bounded* but not *asymptotically stable*. It does not converge to its (unique) equilibrium, but it also does not exhibit unbounded deviations from the equilibrium.

Chiarella (1992) in a nonlinear version of the described setting in continuous time ends up with a closed orbit with constant amplitude; Day and Huang get an even more exciting outcome: Depending on parameter values, the market could exhibit *chaotic* fluctuations.[5] Despite the deterministic excess demand functions and price formation rule, the price trajectories then appear similar to the realization of a stochastic process, with random switches between bear and bull markets. The difference in outcomes is mainly due to the mathematical formalization of the speculative dynamics; systems of differential equations are capable of generating chaos only if they consist of

[5]Gu (1995) analyzed market-mediating behavior of an active market maker in the framework of Day and Huang, demonstrating that it would be in the interest of this agent to churn the market rather than calm it down—that is, choose a price adjustment speed in the chaotic zone of parameter values.

at least three first-order equations (Beja and Goldman and Chiarella only have two equations), but even difference equations of the first order can generate chaotic attractors. The erratic appearance of price paths from a deterministic system and the lack of predictability of chaotic systems provided a new avenue toward an explanation of the stylized facts: Despite deterministic behavioral sources, the systematic forces of the market interactions could become "invisible" due to the apparent randomness of the chaotic dynamics.

A similar avenue is pursued within a model of the foreign exchange market by DeGrauwe et al. (1993). Assuming simple versions of moving average rules applied by chartists, they end up with a higher-order system of difference equations that yields chaotic attractors for a broad range of parameter values. Interestingly, their model also allows us to explain stylized facts of foreign exchange markets other than merely the deviation between market exchange rates and their fundamental values. In particular, they demonstrate that their chaotic process is hard to distinguish from a unit root process (martingale) by standard statistical tests and that the overall dynamics could explain the forward premium puzzle (the finding that forward rates are poor and biased predictors of subsequent exchange rate movements). Experiments with a macroeconomic news arrival process indicate that although this incoming information is incorporated into exchange rates over longer horizons, there is no one-to-one mapping between exchange rate changes and macroeconomic news in the short run. The connection between the currency movements and macroeconomic factors might, then, at times appear quite loose, explaining the so-called "disconnect" puzzle and the failure of macroeconomic models to predict exchange rates.

A closely related recent branch of literature is that of "adaptive belief systems." In contrast to the contributions reviewed earlier, this class of model allows agents to switch between various prediction functions (mostly chosen from the typical chartist and fundamentalist varieties). Thus, the fractions of agents using particular predictors become additional state variables in addition to the market price. Early work in this vein was mainly concerned with the possible bifurcation routes toward chaotic attractors in these systems (Brock and Hommes, 1997, 1998). Various extensions have considered a broad variety of prediction functions, have allowed for transaction costs, and have studied the endogenous development of wealth of agent groups as an alternative to switches due to the success or failure of their predictions (Gaunersdorfer, 2000; Chiarella, Dieci and Gardini, 2002; Chiarella and He, 2002; Brock, Hommes, and Wagener, 2005; DeGrauwe and Grimaldi, 2006).

A recent paper by Gaunersdorfer and Hommes, 2007, is concerned with a possible mechanism for volatility clustering within this framework. They demonstrate that in a scenario with coexistence of a locally stable fixed point and an additional cycle or chaotic attractor, superimposed stochastic disturbances would lead to recurrent switches between both attractors. In the vicinity of the fixed point, fluctuations will be confined to the stochastic disturbance, whereas the endogenous dynamics of the cycle or chaotic attractors will magnify the stochastic fluctuations. As a consequence, switching between both attractors will come along with the impression of volatility clustering and, therefore, could provide a possible explanation of this stylized fact. Gaunersdorfer and

Hommes show that estimation of GARCH parameters produces numbers close to those of empirical data for some parameterizations of the model.

In a related framework, He and Li (2007) point out that an appropriate combination of noise factors in an otherwise deterministic chartist-fundamentalist model (both a stochastic fundamental value and an additional noise component in aggregate excess demand are assumed) could lead to absence of autocorrelation in raw returns together with apparent hyperbolic decay of autocorrelations in squared and absolute returns.

The adaptive belief models have a close resemblance to a class of models using machine-learning tools for agents' expectation formation. The prototype of this strand of literature on artificial financial markets is the Santa Fe artificial stock market (Arthur et al., 1997; LeBaron et al., 1999) that had already been launched in the early 1990s. In this model, traders are equipped with a set of classifiers of chartist and fundamentalist types to categorize the configuration of the market and formulate expectations of future returns based on this classification. Both classifiers and forecast parameters evolve via genetic operations. The main finding of this project is that dominance of either chartist or fundamentalist components depends on the frequency of activation of the genetic operations. Under frequent activation, chartist behavior was found to dominate while, with a lower frequency of activation, fundamentalist classifiers gained in importance. Imposing "short-termism" on the artificial agents, they are apparently forced to focus on trends rather than on, for example, price-to-dividend ratios. It has also been shown that the chartist regime had higher volatility than the fundamentalist regime, and the latter exhibited excess kurtosis as well as positive correlation between volume and volatility.

Other recent artificial markets include Chen and Yeh (2002), whose traders are equipped with genetic programming tools rather than classifier systems. Simpler models with artificial agents have used genetic algorithms for parameter selection of trading strategies (Arifovic, 1996; Dawid, 1999; Szpiro, 1997; Lux and Schornstein, 2005; and Georges, 2006). Due to the inherent stochasticity of the evolutionary learning mechanism, some of these models are closer in spirit to the stochastic models discussed later in this chapter than to the deterministic approaches of the early "chaos" literature.

One concern regarding the body of literature on chartist-fundamentalist models with a deterministic structure is the lack of convincing evidence of chaotic dynamics in financial markets. There had been some hope of detecting low-dimensional deterministic chaos in financial returns in the early literature on this subject (Eckmann et al., 1988; Scheinkmann and LeBaron, 1988), but it soon turned out that the daily datasets used in these studies were too small for reliable estimation of, for example, the correlation dimension of a chaotic attractor (Ruelle, 1990). The consensus that emerged from this body of literature was that the empirical evidence for low-dimensional chaos is weak. One should also note that despite their sensitivity with respect to initial conditions, low-dimensional attractors are characterized by recurrent patterns that could probably be exploited too easily by advanced methods from the toolbox of nonlinear dynamics. Nevertheless, the literature agrees that there is strong evidence for nonlinearity in that all standard tests for IID-ness would typically reject their null hypotheses when applied to financial returns. However, this nonlinear dependency is mostly confined to higher moments (GARCH effects) and might not be uniformly present in the data.

As an interesting exercise by de Lima (1998) demonstrates, rejection of IID-ness by the popular BDS test in S&P 500 returns over the 1980s happens only if one uses data including the crash of 1987. If the series stops before this event, the data, in fact, look like white noise and would not reject the null hypothesis. These findings indicate that financial data have a structure that is even more complex than that of chaotic processes. It might, therefore, be important to allow for both deterministic and stochastic factors, the interaction of which could give rise to different behavior in different time windows.

3.4.2. Kirman's Model of Opinion Formation and Speculation

A few attempts at modeling stochastic economies of interacting agents were published decades ago (most notably Fölmer's seminal 1974 paper); a more systematic analysis of stochastic interactions only started in the 1990s and was largely confined to models of trading in financial markets. The first study that gained wider prominence within the economics literature is Kirman's model (1991, 1993) of herding through pairwise contacts. Its mechanism of contagion of opinions, which in principle could be applied to a variety of problems in economics and beyond, has also been used as an ingredient in models of interacting chartists and fundamentalists and serves to highlight the differences in results brought about by an intrinsically stochastic rather than deterministic framework.

We start with the basic stochastic interactions considered in Kirman's approach. The motivation for Kirman's model stems from experiments on information transmission among ants. If a group of foraging ants is offered two identical sources of food in the vicinity of its nest, a majority of the population will be found to exploit one of the two resources at any point in time. This concentration comes about by chemical information transmission via pheromones by which successful pioneer ants recruit followers and guide them to the same manger. The higher concentration of pheromones on one of the two paths to both food sources stimulates more and more ants to exploit the same source. However, if experiments last long enough, random switches of the preferred source are observed, and averaging over time, a bimodal distribution is found for the number of ants collecting food from one source. The switch is believed to be caused by evaporation of pheromones together with random search of ants not yet recruited for the exploitation of one resource. This combination of concentrated exploitation and random search is often viewed as an evolved optimal foraging strategy that achieves a balance between the costs and benefits of undirected search and exploitation of known resources (see Deneubourg et al., 1990). It also counts as one of the leading examples of natural optimization and together with similar findings of seemingly purposeful self-organized behavior in insect societies has motivated the new brand of "ant algorithms" in the artificial intelligence literature (Bonabeau, Dorigo, and Theraulaz, 1999).

Kirman (1993) has come up with a stochastic model of this recruitment process of foraging ants. Following the experimental setup, he assumes that ants (agents) have two alternatives at their disposal (which might be food sources, opinions, or strategies such as chartism and fundamentalism). Each individual is assumed to adhere to one of the two alternatives at any point in time. There exists a fixed number of N agents, and k denotes the number of those who are currently following alternative 1. Hence, the probability that a randomly chosen agent belongs to group 1 (2) is $\frac{k}{N}$ and $\frac{N-k}{N}$, respectively.

The state of the system then changes over time by a combination of recruitment and random changes (random search):

1. Individuals meet pairwise and exchange information on their respective strategies or opinions. From these meetings any agent might come out convinced or persuaded that the choice of the other is more preferable. This happens with a constant probability $1-\delta$ (δ denoting the probability of holding on to one's own former strategy or opinion).
2. Individuals can also change their opinion or strategy without meeting others in an autonomous fashion—say, due to idiosyncratic factors. This random change happens with a probability ϵ.

Within a small time interval (small enough to allow for at most one pairwise encounter), the number k of individuals of type 1 can, consequently, undergo the following changes:

$$k \rightarrow \begin{cases} k+1 & \text{with probability} \quad p_1 \\ k & \text{with probability} \quad 1-p_1-p_2 \\ k-1 & \text{with probability} \quad p_2 \end{cases} \tag{3.15}$$

The probabilities in Eq. 3.15 are determined by simple combinatorial considerations:

$$p_1 = \text{Prob}((k \rightarrow k+1) = \frac{N-k}{N}\left(\epsilon + (1-\delta)\frac{k}{N-1}\right) \tag{3.16}$$

$$p_2 = \text{Prob}(k \rightarrow k-1) = \frac{k}{N}\left(\epsilon + (1-\delta)\frac{N-k}{N-1}\right) \tag{3.17}$$

The resulting stochastic process converges to a limiting distribution which, for large N and small ϵ, can be approximated by the symmetric Beta distribution:

$$f(x) = \text{const} \cdot x^{\alpha-1}(1-x)^{\alpha-1} \tag{3.18}$$

where $x \equiv \frac{k}{N}$. The shape parameter $\alpha \equiv \frac{\epsilon(N-1)}{1-\delta}$ depends on the relative strength of the autonomous component ϵ and the recruitment probability is $1-\delta$. The equilibrium distribution may have a unimodal or bimodal shape; for $\epsilon > \frac{1-\delta}{N-1}$ the distribution will be unimodal with a concentration of probability mass around the mean value $\frac{k}{2}$. If we increase the herding propensity, however, the population dynamics will undergo what is denoted as a "phase transition" at $\epsilon = \frac{1-\delta}{N-1}$, with the equilibrium distribution changing from unimodal to bimodal.

Figure 3.5 illustrates the various possibilities for the distribution of opinions together with examples of stochastic simulations for the unimodal and bimodal case. Note that

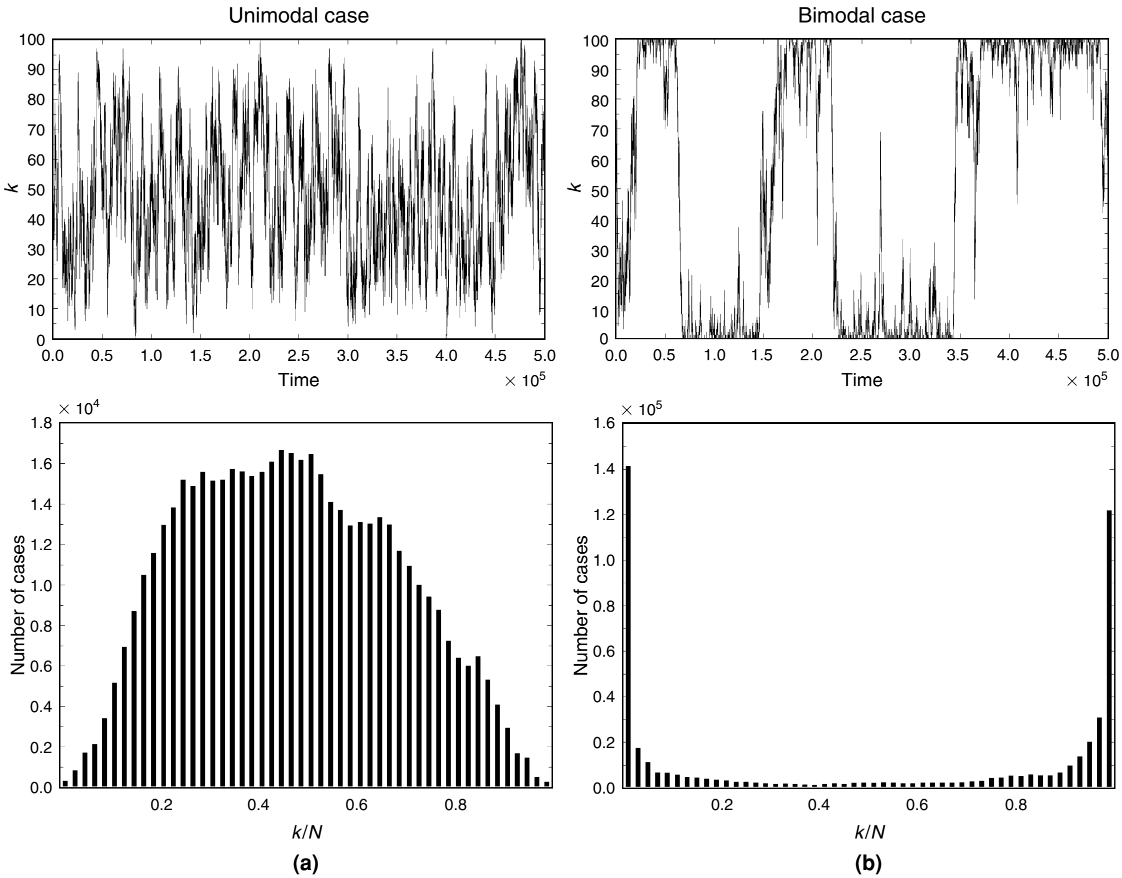

FIGURE 3.5 Two possible scenarios of Kirman's ant model: The population might fluctuate around the mean value $k/2$ with an equal number of agents in both groups (a) or it might tend toward a uniform state $k = 0$ or $k = N$ with intermittent switches between both polar cases due to the randomness of the recruitment process (b). Parameters are $\delta = 0.15, N = 100$, and $\epsilon = 0.02$ and 0.002, respectively. The top diagrams in (a) and (b) show simulations of both cases; the bottom ones exhibit the frequency of observations $\frac{k}{N}$ in these simulations.

in the bimodal case, the mean value of the equilibrium distribution is still $\frac{k}{2}$, but it is the *least probable* realization to be observed in a simulated time series. Probability mass is rather concentrated at the extreme ends, which means that a majority of agents will follow one of the two alternatives most of the time. This also means that although all agents are governed by the same conditional probabilistic laws, the emergent global configuration may be inhomogeneous with alternating phases of dominance of one or the other strategy. If we assume that the two alternatives in question are chartist and fundamentalist strategies, we would observe waves of popularity of one or the other alternative among traders due to noneconomic forces. Frankel and Froot (1986) and Liu (1996) offer some evidence for changes in popularity of both types of strategies in foreign exchange markets.) From a certain perspective, such an added noneconomic explanation for the popularity of chartist and fundamentalist trading strategies could have some appeal; in an efficient market, both alternatives would be inferior to a single buy-and-hold strategy so that some reasons outside the realm of economics would be needed to explain their perseverance and popularity among traders.[6] However, these noneconomic forces would lead to dependence between agents. The lack of independence of individuals' deviations from rational behavior would prevent applicability of a suitable law of large numbers. As a consequence, irrationality would not be washed out in the aggregate but would exert a nonnegligible influence on equilibrium prices.

Kirman (1991) had already incorporated his recruitment mechanism into a chartist-fundamentalist framework. His model is formulated as a monetary model of the foreign exchange market. This implies that the equilibrium exchange rate is determined by the *uncovered interest parity* (UIP) condition. Summarizing the fundamental macroeconomic factors influencing the domestic and foreign interest rates via a compound contemporaneous macro variable x_t, the equilibrium log exchange rate is obtained as

$$S_t = x_t + \delta E_{m,t}[S_{t+1}] \qquad \textbf{(3.19)}$$

where δ is the discount factor and $E_{m,t}[S_{t+1}]$ is the marketwide expectation of the future exchange rate entering via UIP.

The traditional monetary approach would, of course, assume rational expectation formation. Any current information on future changes of macroeconomic fundamentals (components of $x_\tau, \tau = t+1, t+2, \ldots$) would, then, be incorporated into current spot rates via their correctly predicted influence on future equilibrium exchange rates. Upon iterative solution of Eq. 3.19, the fundamental value $S_{f,t}$ would be obtained. Analogously to Beja and Goldman and the related literature on speculative models of asset price formation, rational expectations are replaced in Kirman's model by the nonrational expectations of chartists and fundamentalists. Kirman assumes the standard format of fundamentalists' expectations:

$$E_{f,t}[\Delta S_{t+1}] = a(S_{f,t} - S_{t-1}) \qquad \textbf{(3.20)}$$

with $\Delta S_{t+1} = S_{t+1} - S_t$ and a simple trend-following rule for chartists. In our representation of chartists' expectation function we slightly modify Kirman's original setup:

$$E_{c,t}[\Delta S_{t+1}] = b(S_{t-1} - S_{t-2}) \qquad \textbf{(3.21)}$$

The number of agents formulating their expectations in one or the other way is assumed to change under the influence of the stochastic recruitment process so that

[6] Kaldor (1939, p. 2) already argued that there might be a representation bias that might lead to a steady inflow of new speculators, even in the presence of the net losses of the population of existing speculators as a whole: "...even if speculation as a whole is attended by a net loss, rather than a net gain, this will not prove, even in the long run, self-corrective. For the losses of a floating population of unsuccessful speculators will be sufficient to entertain permanently a small body of successful speculators; and the existence of this body of successful speculators will be a sufficient attraction to secure a permanent supply of this floating population."

the aggregate forecast of the exchange rate, $E_m[S_{t+1}]$, is given as a weighted average whose weights change stochastically with the group occupation numbers:

$$\begin{aligned} E_{m,t}[S_{t+1}] &= E_{m,t}[\Delta S_{t+1}] + S_t \\ &= S_t + w_t E_{f,t}[\Delta S_{t+1}] + (1 - w_t) E_{c,t}[\Delta S_{t+1}] \end{aligned} \quad \textbf{(3.22)}$$

Setting $w_t = \frac{k_t}{N}$, we obtain:

$$E_{m,t}[S_{t+1}] = S_t + \frac{k_t}{N} E_{f,t}[\Delta S_{t+1}] + \frac{N - k_t}{N} E_{c,t}[\Delta S_{t+1}] \quad \textbf{(3.23)}$$

In Kirman's simulations of this model, weights are not exactly identical to group occupation numbers but are given by agents' assessment of what the majority opinion might be. For this purpose every agent is assumed to receive a noisy signal of the majority opinion. Assuming that agents follow this perceived majority, the aggregate of these signals is, then, used instead of the raw outcome from the population model, k_t. In addition, the social dynamic occurs at a faster time scale than price formation in the foreign exchange market.

In particular, in the results reported in various papers (Kirman, 1991, 1992; Kirman and Teyssière, 2002), the group occupation numbers are sampled after 10,000 pairwise encounters, to implement the weights in the unit time steps of Eq. 3.23. However, all these refinements are of minor importance. The more important insight is that we end up with a complex dynamic system in which the process of social interactions between agents exerts a crucial influence on the relatively conventional (in the light of the previous Section 3.4.1) speculative process component. To see this, plug Eqs. 3.20, 3.21, and 3.22 into 3.19:

$$S_t = x_t + w_t a(S_{f,t} - S_{t-1}) + (1 - w_t) b(S_{t-1} - S_{t-2}) \quad \textbf{(3.24)}$$

For constant fractions of chartists and fundamentalists, this is only slightly different from the model of Beja and Goldman. Despite the formulation in the tradition of monetary models of the exchange rate, the different formalization of chartists' expectations, and the discrete rather than continuous-time framework, we easily recover the stabilizing and destabilizing tendencies of both groups. In particular, we immediately see that the system would be unconditionally unstable if all traders adopted the chartist forecast rule ($w_t = 0$). In the case of complete dominance of fundamentalists, we would find stability of the fundamental equilibrium in the case $a < 1$ (as with $a > 1$, fundamentalists would overreact to a discrepancy between the current price and the fundamental value). Furthermore, for the system of two interacting groups, stability conditions can be expressed in terms of group occupation numbers. For a constant $w_t = \bar{w}$, this second-order difference equation would have an asymptotically stable equilibrium if the following conditions were satisfied:

1. $\bar{w} > 1 - \frac{1}{b}$
2. $\bar{w} < \frac{2b+1}{2b+a}$

Note that the second condition is always met if fundamentalists' reaction is not excessive ($a < 1$). The autonomous recruitment dynamics can be seen as a driving factor that sweeps the speculative dynamics back and forth between stable and unstable configurations. As can be inferred from a typical simulation, the stochastic fluctuations of the prevailing majorities lead to different characteristic phases in the market's dynamics. During phases dominated by fundamentalists, the exchange rate stays close to its fundamental value while speculative bubbles emerge if the majority turns to the chartist forecast rule. Bubbles collapse together with the stochastic switches from the chartist majority back to the fundamental majority. Kirman (1992b) shows that standard tests would mostly not reject the unit root hypothesis for simulated time series, while Kirman and Teyssière (2002), testing for long-term dependence in absolute and squared returns, find robust indication of long-term dependence with decay parameters in the range of those obtained with empirical data. Although we do not get the full set of stylized facts reviewed in Section 3.2, the sweeping through a bifurcation value (threshold for a qualitative change of the dynamics) due to superimposed stochastic forces is a more general phenomenon that also occurs in other models of interacting agents. As we will see, in a slightly different framework it appears to be a potential key mechanism generating fat tails and clustered volatility.

Alfarano, Lux, and Wagner (2008) study a continuous-time version of the "ant process." In their model, the speculative dynamic is closer to the Beja and Goldman legacy. They assume constant fractions of fundamentalists and chartists, but they have the number of buyers and sellers among chartists determined by the social interaction dynamics. Using tools from statistical physics, they derive approximate closed-form solutions for conditional and unconditional moments of returns of their asset price process. They show that leptokurtosis and volatility persistence are generic features of this model, although neither the unconditional distribution nor the autocorrelations exhibit "true" power-law decay. However, Alfarano and Lux (2007), in a closely related model, demonstrate that up to a characteristic time scale the temporal characteristics would closely resemble those of a process with "true" long memory, and the deviation from "true" asymptotic scaling behavior could only be detected for very long (simulated) time series.

Gilli and Winker (2003) estimated via a heuristic grid search method Kirman's (1991) model for the U.S.$-DEM exchange rate and obtained parameter estimates within the bimodal regime. Alfarano, Lux, and Wagner (2005) developed an approximate ML approach for a similar model that generalizes the previous framework by allowing for asymmetric autonomous transition rates between groups. This added feature leads to arbitrary asymmetries in the limiting distribution of the population configuration depending on parameter values, which translates into subtle asymmetries in the distribution of returns. Since a higher autonomous tendency toward one group leads to a certain dominance of one strategy, the empirical parameter estimates provide some evidence on the average population composition in the market under investigation. As it turns out, parameter estimates of this asymmetric ant model indicate that stock markets mostly have a higher fraction of chartists than foreign exchange markets.

3.4.3. Beyond Local Interactions: Socioeconomic Group Dynamics in Financial Markets

Pairwise interactions as they appear in the seminal ant model are just one way to formalize the interpersonal influences between economic agents. The first systematic investigation into the effects of mutual noneconomic interaction between agents in an economic setting is due to Föllmer (1974). He studied the existence and uniqueness of the equilibria in a system of markets and pointed to the existence of a *phase transition* from a unique prize vector to a "polarized" multimodal state with increasing strength of interpersonal spillovers. This phenomenon is similar to the bifurcation from unimodality to bimodality in the ant model and can be found in various broadly similar approaches.

Social Interactions: A General Framework

This section outlines another model in the chartist–fundamentalist tradition that uses a formalization of interactions that can be viewed as the opposite extreme to pairwise influences: Traders will be assumed to be influenced by the overall mood of the market, that is, an average of the influence from all their fellow traders. Such an overall influence allows us to study the macroscopic dynamics more easily via so-called mean-field approximations. As will be seen, most qualitative results of the simple model that follows are in harmony with the previously reported findings. The added advantage (besides the generalization of previous results) will be that this framework allows us to illustrate a general avenue toward the analysis of macroscopic quantities in stochastic systems of interacting agents that can be compared to typical stochastic models applied in empirical finance. Our particular framework is adopted from Lux (1995, 1998).

As in Kirman's model, a population is divided into two camps, say, optimistic and pessimistic (or bullish and bearish) individuals whose average mood can be captured by the opinion index x:

$$x = \frac{n_+ - n_-}{2N} \tag{3.25}$$

with $n_+ (n_-)$ the current number of optimists (pessimists) and $2N$ the overall number of agents. Individuals are assumed to revisit their choice of opinion from time to time and to have a tendency to switch to the majority opinion. With an overall "field" effect (i.e., all other individuals exerting the same influence on any one), the group pressure can be modeled via some feedback effect from the macroscopic configuration x on individual decisions. This feedback leads to migration of individuals between both groups under the influence of the overall "field" of the average opinion. Formally, these transitions might be specified by Poisson processes in continuous time with rates p_{+-} and p_{-+} for an individual from the $-$ group to switch to the $+$ group, and vice versa. The canonical function used for transitions in particle physics is the exponential, which motivates an *ansatz* of the following type:

$$p_{+-} = v \cdot \exp(\alpha x), p_{-+} = v \cdot \exp(-\alpha x) \tag{3.26}$$

Obviously, Eq. 3.26 supposes positive probabilities for agents to migrate between groups but hypothesizes a stronger tendency for migration following the dominating opinion: If $x > 0$ ($x < 0$), the majority of the population can be found in the + (−) group and the probability for other agents to join this group is larger than that of members of the majority to switch to the minority view. Thus we have a very direct formalization of social interaction or herding among the members of our population. Note that Eq. 3.26 includes two parameters: v, which captures the general frequency of revision of opinion within our population, and α, which parameterizes the strength of the herding effect. Because of the assumed Poisson nature of the switches of opinion, we can easily come up with probabilities for movements of agents from one group to another during a certain time interval Δt. For small time increments Δt, the simultaneous movements of two or more individuals during an interval Δt become increasingly unlikely and can be completely neglected in the limit $\Delta t \to 0$. In addition, the probability for an individual to switch from one group to the other converges to $p_{+-}\Delta t$ and $p_{-+}\Delta t$. Since these Poisson processes are assumed to be the same for all members of the optimistic and pessimistic groups, respectively, we can infer the transition rates for group occupation numbers as the limiting cases of conditional probabilities $w(n_i + 1, t + \Delta t | n_i, t)$ for $i \in \{\text{“}+\text{”}, \text{“}-\text{”}\}$:

$$\lim_{\Delta t \to 0} \frac{w(n_+ + 1, t + \Delta t | n_+, t)}{\Delta t} \equiv w(n_+ + 1 | n_+, t) = n_- p_{+-} \qquad \textbf{(3.27)}$$

$$\lim_{\Delta t \to 0} \frac{w(n_- + 1, t + \Delta t | n_-, t)}{\Delta t} \equiv w(n_- + 1 | n_-, t) = n_+ p_{-+} \qquad \textbf{(3.28)}$$

Denoting by $n = \frac{1}{2}(n_+ - n_-) = xN$ the *socioeconomic* configuration with $n \in \{-N, -N+1, \ldots, N-1, N\}$, and $x \in \{-1, -1 + \frac{1}{N}, \ldots, 0, \ldots, \frac{N-1}{N}, 1\}$, the opinion dynamics leads to a sequence of switches from n to one of the neighboring values $n \pm 1$ (or from x to $x \pm \frac{1}{N}$) in irregularly spaced time intervals. A complete description of the dynamic process is obtained via the so-called Master equation, which captures the change in time of the probabilities $\tilde{Q}(n, t)$ or $Q(x, t)$ over all candidate states n or x, respectively. This amounts to a system of differential equations for the probability flux, which in our case can be written as:

$$\begin{aligned} \frac{dQ(x;t)}{dt} &= \underbrace{w_{\downarrow}\left(x + \frac{1}{N}\right) Q\left(x + \frac{1}{N}; t\right) + w_{\uparrow}\left(x - \frac{1}{N}\right) Q\left(x - \frac{1}{N}; t\right)}_{\textit{inflow of prob. to state } x} \\ &\quad - \underbrace{(w_{\downarrow}(x) + w_{\uparrow}(x)) Q(x;t)}_{\textit{outflow of prob. from } x} \end{aligned} \qquad \textbf{(3.29)}$$

The transition rates for the changes of x by one unit $\pm\frac{1}{N}$ are identical to the rates introduced in Eq. 3.27 and 3.28 translated into the pertinent transition rates of the intensity x:

$$w_{\uparrow}(x) = n_{-}p_{+-} = (1-x)Np_{+-}, w_{\downarrow}(x) = n_{+}p_{-+} = (1+x)Np_{-+} \quad \textbf{(3.30)}$$

Note that the rates $w_{\uparrow}(x)$ and $w_{\downarrow}(x)$ are both state dependent and nonlinear due to our formalization of the individual transition rates p_{+-} and p_{-+}. Although one could in principle use the Master equation to simulate the time development of our dynamic process, it is certainly too complicated to allow an analytical solution. The major advantage of this formalization consists, however, in its use as a starting point to derive more manageable approximations. One potential avenue consists of performing a Taylor series approximation to the Master equation itself, leading to the so-called *Fokker-Planck equation*, the use of which is illustrated in Figure 3.7.

A second complementary approach is to investigate macroscopic characteristics of the dynamic process; for example, first, second, or higher moments, the implementation of which also requires the Master equation formalism. The details of both approaches have been nicely laid out in the monographs by Weidlich and Haag (1983), Aoki (1996), and Weidlich (2002). We will not go into too much detail here but simply illustrate some of the main results that can be obtained for our opinion dynamics.

We start with the first moment of the opinion index, that is, $\overline{x_t}$, whose time change characterizes the most probable development of the system conditional on an initial condition x_0 at time $t = 0$. Since the mean is defined by

$$\overline{x_t} = \sum_{x=-1}^{1} xQ(x;t) \quad \textbf{(3.31)}$$

its change in time can be exactly computed only under complete knowledge of the dynamics of the probability distribution over all states x:

$$\frac{d\overline{x_t}}{dt} = \sum_{x=-1}^{1} x\frac{dQ(x;t)}{dt} \quad \textbf{(3.32)}$$

The exact time evolution covered in Eq. 3.32 can be approximated in a Taylor series expansion around the current mean $\overline{x_t}$ to various degrees of accuracy. To first order, we obtain a self-consistent differential equation:[7]

$$\frac{d\overline{x_t}}{dt} = a_{x,1}(\overline{x_t}) \quad \textbf{(3.33)}$$

[7]Note that the first derivative vanishes because of $E[(x-\bar{x})a'_{x,1}(\bar{x})] = 0$.

while to second-order accuracy, a correction term involving the variance σ_x^2 enters the equation:

$$\frac{d\overline{x_t}}{dt} = a_{x,1}(\overline{x}) + \frac{1}{2}\sigma_x^2 a''_{x,1}(\overline{x}) \tag{3.34}$$

The function $a_{x,1}$ in Eqs. 3.33 and 3.34 is denoted as the *first-jump moment*. It gives the expected change of the system conditional on the previous realization. Evaluated at the current expectation, $\overline{x}_t$ (conditional on an initial condition), it allows us to track the mean-value dynamics of the system. In the case of an infinite population, Eq. 3.33 would be exact. For finite populations, however, the influence of higher moments has to be taken into account. Equation 3.34 includes the next higher term in the Taylor series expression of Eq. 3.32 involving the second moment. Higher-order expansions would involve higher-order moments in the additional entries on the right side of the equation. An example for the determination of the jump moment $a_{x,1}$ follows.

The dynamics of higher moments are obtained analogously. For example, the second moment $\overline{x_t^2} = \sum_x x^2 Q(x;t)$ changes over time according to

$$\frac{d}{dt}\overline{x_t^2} = \sum_x x^2 \frac{dQ(x;t)}{dt} \tag{3.35}$$

while the time change of the variance σ_x^2 is given by

$$\frac{d}{dt}\sigma_x^2 = \frac{d}{dt}(\overline{x^2} - \overline{x}^2) = \frac{d}{dt}\overline{x^2} - 2\overline{x}\frac{d}{dt}\overline{x} \tag{3.36}$$

and can be solved using Eqs. 3.34 and 3.31. Taylor series expansions of these exact equations again lead to approximations of various orders of accuracy involving non-linear functions of various moments on the right side. One immediate consequence is that any model involving a group dynamic like the one under investigation (or a broad range of alternative population processes) entails *autoregressive dependence in higher moments* as well as *cross-dependencies between moments*.

It has already been pointed out by Braglia (1990) and Ramsey (1996) that stochastic systems based on microscopic interactions (such as our illustrative example) provide a generic avenue toward interesting nontrivial dynamics in higher moments and, therefore, a potential behavioral explanation for the ubiquitous ARCH effects in financial data. Of course, it would have to be seen whether the direction and extent of autoregressive dependency in any hypothesized micro model is in qualitative and quantitative agreement with the empirical stylized facts. Interestingly, Braglia (1990) already argued that it would be natural to interpret a cross-dependence between the mean and second moment as a fads effect. Lux (1998) provides a fully worked-out analysis of the interactions between first and second moments in an asset-pricing model with nonrational speculators, along the lines of our present framework.

Let us return to our particular model of social imitation. Implementing Eq. 3.33, the exact mean-value dynamics turns out to be determined by the average change of the configuration, $a_{x,1}$:[8]

$$\frac{d\overline{x_t}}{dt} = \sum_x \sum_{x'} (x' - x) w_{x'x} Q(x,t) = \sum_x a_{x,1} Q(x,t) \tag{3.37}$$

Since possible movements within an infinitesimal time step are restricted to neighboring states, $x' - x$ can only assume values $\frac{1}{N}$ and $-\frac{1}{N}$, so that:

$$a_{x,1} = \frac{1}{N} w_{\uparrow}(x) + (-\frac{1}{N}) w_{\downarrow}(x) \tag{3.38}$$

Since $w_{\uparrow}$ and $w_{\downarrow}(x)$ are identical to $w_{\uparrow}(n)$ and $w_{\downarrow}(n)$ in Eq. 3.30, we end up with

$$\begin{aligned} a_{x,1} &= \frac{1}{N} n_- p_{+-} - \frac{1}{N} n_+ p_{-+} \\ &= (1-x) v e^{\alpha x} - (1-x) v e^{-\alpha x} \end{aligned} \tag{3.39}$$

Using the hyperbolic trigonometric functions, this can be rewritten as

$$a_{x,1} = 2v\{\tanh(\alpha x) - x\} \cosh(\alpha x) \tag{3.40}$$

The exact mean value equation is thus:

$$\frac{d\bar{x}_t}{dt} = \sum_x 2v\{\tanh(\alpha x) - x\} \cosh(\alpha x) Q(x;t) \tag{3.41}$$

Its first-order Taylor series approximation around $\bar{x}_t$ leads to the self-consistent differential equation:

$$\frac{d\bar{x}_t}{dt} = 2v\{\tanh(\alpha \bar{x}) - \bar{x}\} \cosh(\alpha \bar{x}) \tag{3.42}$$

while the second-order approximation involves a correction factor due to fluctuations around the mean:

$$\frac{d\bar{x}_t}{dt} = 2v\{\tanh(\alpha \bar{x}) - \bar{x}\} \cosh(\alpha \bar{x}) +$$

$$v\{(\alpha^2 - 2\alpha) \sinh(\alpha \bar{x}) - \bar{x}\alpha \cosh(\alpha \bar{x})\} \sigma_x^2 \tag{3.43}$$

which already reveals a rich nonlinear structure of interactions between first and second moments. Note that σ_x^2 is time-changing as well. Approximating the dynamic law (3.36) for σ_x^2 to first order, one would arrive at an equation that also depends on both the first

[8]Using $w_{x'x}$ as a shorthand notation for $\lim_{\Delta t \to 0} \frac{w(x', t+\Delta t | x, t)}{\Delta t}$.

and second moments, $\bar{x}_t$ and σ_x^2. Combining the second-order approximations of the first moment and the first-order approximation of the second moment would, thus, lead to a self-consistent system of two (highly nonlinear) first-order differential equations.

We proceed by investigating the properties of the first-order approximation, Eq. 3.42. Since $cosh(\cdot) > 0$ for all x, the condition for a steady state of the mean value dynamics is

$$\frac{d\bar{x}_t}{dt} = 0 \Leftrightarrow \bar{x^*} = \tanh(\alpha\bar{x^*}) \qquad \textbf{(3.44)}$$

Since $tanh(.)$ is bounded between -1 and 1 and its local slope at 0 is equal to 1, we arrive at the following insights concerning the equilibria of the system:

- $\alpha \leq 1$ implies existence of a stable, unique equilibrium $\overline{x_0^*} = 0$.
- $\alpha > 1$ gives rise to multiple equilibria, $\bar{x}_-^*, \bar{x}_0^*, \bar{x}_+^*$ with $\bar{x}_+^* = -\bar{x}_-^* > 0$, of which the outer ones are stable and the middle one, $\bar{x}_0^*$, is unstable (see Figure 3.6).

The bifurcation from a unique steady state to multiple steady states shows that the interaction intensity needs to surpass a certain critical value for a "polarized" state to emerge. If interaction is weak ($\alpha \leq 1$), the system would fluctuate around a balanced state with, on average and in expectation, as many optimistic as pessimistic agents. Beyond the critical value ($\alpha > 1$), however, a snowball-like process of infection would result in the emergence of either a majority of + or − agents. Note that the level of the majority $\overline{x_\pm^*}$ depends on the intensity α (see Figure 3.6). Multiplicity of equilibria of the mean-value dynamics corresponds to bimodality of the stationary distribution. The steady states $\overline{x_\pm^*}$ correspond to the two modes of the distribution, whereas the unstable steady state $\overline{x_0^*}$ is identical to the antimode, that is, the local minimum of the stationary distribution.

In the bimodal case the dynamics of the probability distribution would switch from a concentration around the initial state to a bimodal shape according to the two equally likely paths the system could take in the medium and long run. Figure 3.7 shows an example of such a transient density simulated via numerical integration of an approximation to the Master equation (the so-called Fokker-Planck equation). In contrast to the complete characterization of the stochastic process via its transient density, the mean-value dynamics would "only" indicate the most likely path leading to the nearby mode $\overline{x_+^*}$ or $\overline{x_-^*}$. This quasi-deterministic approximation would, therefore, neglect possible recurrent switches between $\overline{x_+^*}$ and $\overline{x_-^*}$ due to stochastic fluctuations.

If we interpret our social dynamics as a formalization of a fads process in our asset market, the agents could be viewed as noise traders switching between bullish and bearish disposition. How much this fads component influences the asset price would, then, depend on the intensity of interaction: With α small, no dominating majority opinion would emerge among the noise traders, and their influence would be minor because the irrational influences would cancel each other out in the aggregate. (In fact, if both the optimistic and pessimistic noise traders would have the same order volume, average

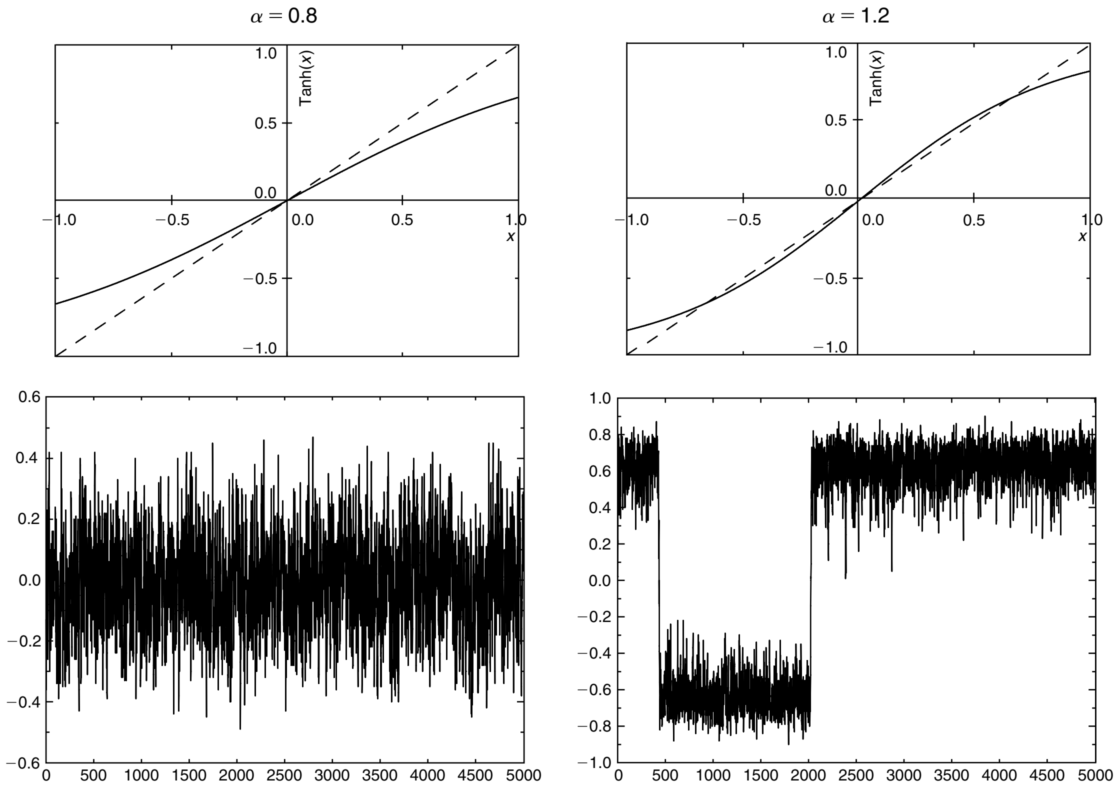

FIGURE 3.6 Two cases of the social dynamics with mean-field effect. If the interaction intensity is weak ($\alpha = 0.8$), minor fluctuations around a balanced disposition among the population occur, while a majority of + or − agents emerges with strong interactions (e.g., for $\alpha = 1.2$). The bifurcation is similar to the one in Kirman's model, but there is no tendency to totally uniform behavior.

excess demand of the irrational agents would be equal to zero.) On the contrary, if interaction is strong, one would observe more coherence among the noise traders' activities, leading to a dominance of either optimists or pessimists at any point in time. The resulting dominance of either buyers or sellers would presumably lead to price changes away from some rationally determined fundamental value. The next section concretizes these thoughts by embedding our model of social imitation into a simple asset-pricing model along the lines of Beja and Goldman (1980).

An Asset-Pricing Model with Social Interactions

In this application, we interpret the former + and − groups as bullish and bearish speculators who are influenced by herd effects together with observed price changes.

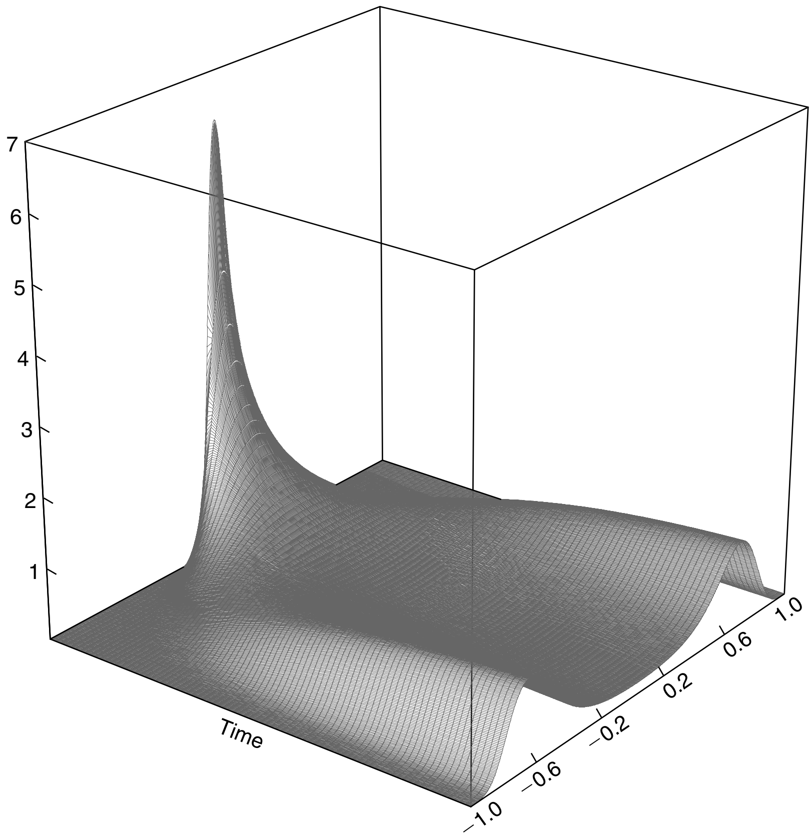

FIGURE 3.7 Example of a transient density computed via numerical integration of the Fokker-Planck equation. Parameters are $\alpha = 1.2$, $\nu = 4$, and $N = 50$. From a deterministic initial condition, $x_0 = 0$, the system is seen to converge toward the known bimodal stationary distribution over time. (The time horizon shown is $T = 2$ unit time intervals.)

Therefore, their transition rates include two terms:

$$p_{+-} = \nu \exp\left(\alpha_1 x + \frac{\alpha_2}{\nu} P'(t)\right) \tag{3.45}$$

$$p_{-+} = \nu \exp\left(-\alpha_1 x - \frac{\alpha_2}{\nu} P'(t)\right) \tag{3.46}$$

where the price change $P'(t)$ reinforces or weakens the herding tendency, depending on whether its sign is in harmony or not with a bullish (bearish) attitude.[9] Following the lines of our previous derivations, we can establish the mean value dynamics for the

[9]Division by ν of the second term is for technical reasons: An agent considers the price change during the mean time interval between switches between groups (which is ν^{-1}).

opinion index for the average bullish or bearish market sentiment (which is pretty close in its structure to some published indices of investor sentiment[10]):

$$\frac{d\bar{x}_t}{dt} = 2\nu \left\{ \tanh\left(\alpha_1 \bar{x}_t + \frac{\alpha_2}{\nu} P'(t) \right) - \bar{x}_t \right\} \cosh(\alpha_1 \bar{x}_t + \frac{\alpha_2}{\nu} P'(t)) \qquad \textbf{(3.47)}$$

To close the model, we must add a hypothesis for price adjustment. A simple possibility is Walrasian price adjustment in reaction to excess demand (ED) with a certain adjustment speed β:

$$P'(t) = \frac{dP}{dt} = \beta ED \qquad \textbf{(3.48)}$$

Following Beja and Goldman (1980), excess demand in our financial market could be decomposed into two components: excess demand by chartists (ED_c) and excess demand by fundamentalist traders (ED_f).

The chartists might be just those whom we have classified as bullish or bearish in the agent-based component of the model. If chartists have a trading volume t_c (per inidividual), this amounts to:

$$ED_c = (n_+ - n_-)t_c = 2Nxt_c = xT_c \quad \text{with} \quad T_c = 2Nt_c \qquad \textbf{(3.49)}$$

following the definition of the opinion index $x = \frac{n_+ - n_-}{2N}$. Fundamentalists, in contrast, will have their excess demand depending on the difference between the perceived fundamental value P_f and the current market price:

$$ED_f = T_f(P_f - P_t) \qquad \textbf{(3.50)}$$

with T_f the proportional trading volume of fundamentalists. Putting both components together, we arrive at the price adjustment equation:[11]

$$\frac{dP_t}{dt} = \beta(\overline{x}_t T_c + T_f(P_f - P_t)) \qquad \textbf{(3.51)}$$

Eqs. 3.47 and 3.51 formalize our interdependent dynamic system in which the group dynamics influence the price dynamics and the price development feeds back on investor sentiment.

In studying the resulting system, we might first explore the question of existence and uniqueness or multiplicity of equilibria. Steady states of the joint opinion and price

[10] In the United States, the popular sentiment data compiled by the American Association for Individual Investors (AAII) as well as those of Investors Intelligence (II) have this structure.

[11] The price equation could in principle also be formalized as a Poisson process, with transition probabilities for price changes in upward and downward direction (see Lux, 1997).

dynamics require $\frac{d\bar{x}_t}{dt} = \frac{dP_t}{dt} = 0$. Since this implies that the new second component of the herding probabilities is zero in any steady state, we arrive at the joint condition:

$$\frac{d\bar{x}_t}{dt} = \frac{dP_t}{dt} = 0 \Longrightarrow \tanh(\alpha_1 \overline{x_t}) = \overline{x_t} \quad \text{and} \quad P_t^* = \frac{T_c}{T_f}\overline{x} + P_f \tag{3.52}$$

Inspection reveals the following:

1. For $\alpha_1 \leq 1$ we have a unique equilibrium $\overline{x_0^*}$, together with $P_t^* = P_f$.
2. For $\alpha > 1$ we encounter the two majority equilibria $\overline{x_+^*}$ and $\overline{x_-^*}$ (now bullish and bearish majorities), with pertinent prices $P_\pm^* = \frac{T_c}{T_f}\overline{x_\pm^*} + P_f$.

Hence, if herding is weak (Case 1) the price converges to the fundamental value (on average); if herding is strong (Case 2), the equilibrium price comes along with an overvaluation or undervaluation of the asset compared to its fundamentals.

However, there are additional possibilities in this more complex system: Both $\overline{x_0^*}$ and the majority states $\overline{x_\pm^*}$ could be unstable (stability conditions are more involved than in the one-dimensional case). In such a scenario the market performs almost regular cycles between overvaluation and undervaluation accompanied by investor sentiment oscillating between bullish and bearish majorities (see Figure 3.8). Expanding our methodology to the 2D case, we could also characterize the fluctuations in different market phases via the variance dynamics and the time development of the covariance between P and $\overline{x}$ (see Lux, 1997).

Realistic Dynamics and the "Stylized Facts"

Of course, neither stationary bubbles nor persistent cycles are realistic scenarios for financial markets.[12] One obvious criticism is that agents maintain their potentially unprofitable strategies forever without learning from past experience. This criticism could be faced by allowing agents to adapt to their environment using some learning or artificial intelligence algorithm for the choice and adaptation of their strategies.

We reviewed some of the contributions in this vein in Section 3.4.1. Here we adopt a very simple mechanism to slightly increase the degree of smartness of our agents. Following Lux and Marchesi (1999, 2000), we allow agents to switch between the fundamentalist and chartist (or noise trader) strategy on the basis of a rough measure of their supposed profitability. As will be seen, this slight extension suffices to remove predictability of market movements to a large degree and also leads to simulated asset prices that share the ubiquitous stylized facts or scaling laws of empirical data. We will argue later that this example might also serve to reveal a general mechanism for

[12]The second part of this statement needs some modification: Note that a cycle in mean values will appear more or less blurred in single realizations of the stochastic process. This distortion might go as far as to leave no apparent trace of cyclical dynamics. Lux and Schornstein (2005) investigate a more complicated model of a foreign exchange market with agents using genetic algorithms to evolve their strategies. Simulations of this model look extremely realistic in terms of returns and their statistical properties. Nevertheless, an analysis of the mean-value dynamics reveals a clear cyclical tendency of the underlying dynamics that becomes fully visible only with a very large population of traders.

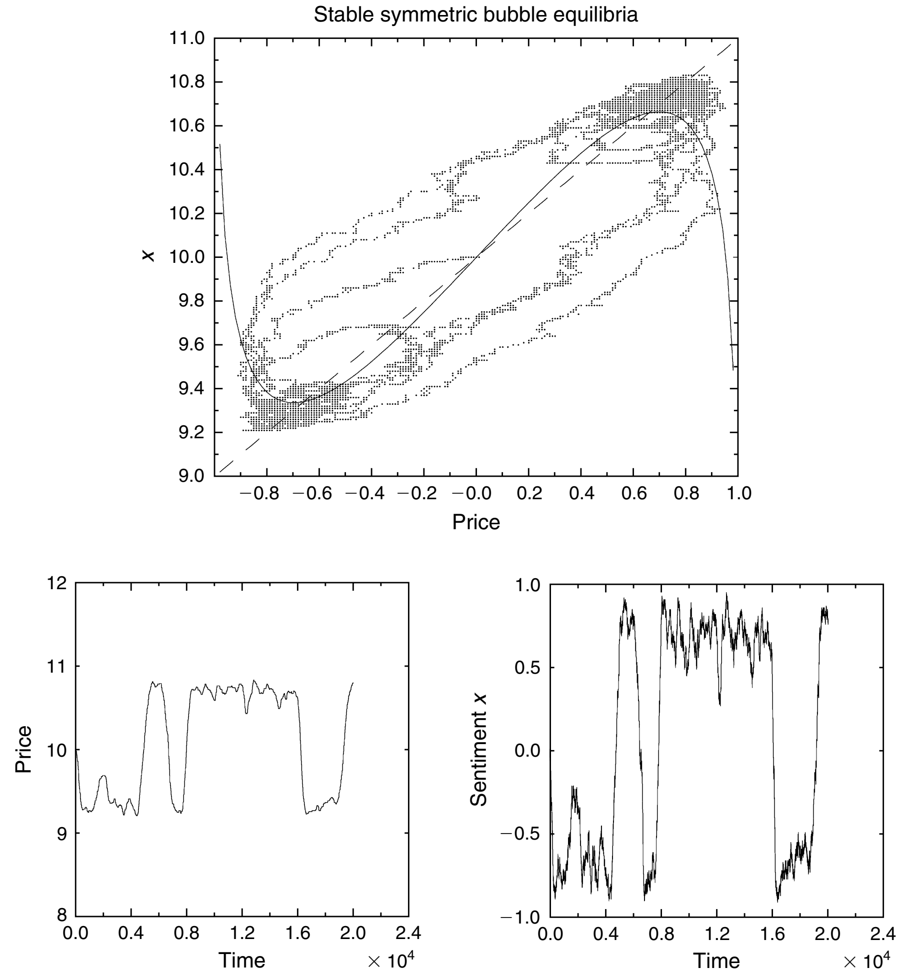

FIGURE 3.8a The case of symmetric "bubble" equilibria: the system tends toward $(\overline{x_+^*}, P_+^*)$ or $(\overline{x_-^*}, P_-^*)$ but also switches occasionally between phases with overvaluation and undervaluation. Parameters are $v = 0.5, \beta = 1, T_c = T_f = 0.5, P_f = 10, \alpha_1 = 1.2, \alpha_2 = 0.75$, and $N = 100$. The broken and solid lines demarcate the isoclines, $P'(t) = 0$ and $x'(t) = 0$, respectively. Their intersections define the equilibria of the system of differential Eqs. (3.47) and (3.48).

generating realistic behavior that could also be identified in some alternative models. The new ingredient of switches between noise traders and fundamentalists is introduced via exponential transition probabilities along the lines of Eqs. 3.45 and 3.46. Formally, four new Poisson transition rates have to be introduced for the propensity of fundamentalists to switch to the optimistic (pessimistic) noise trader camp and vice versa:

$$p_{+f} = v_2 \frac{n_+}{2N} \exp(U_{2,1})$$

$$p_{f+} = v_2 \frac{n_f}{2N} \exp(-U_{2,1})$$

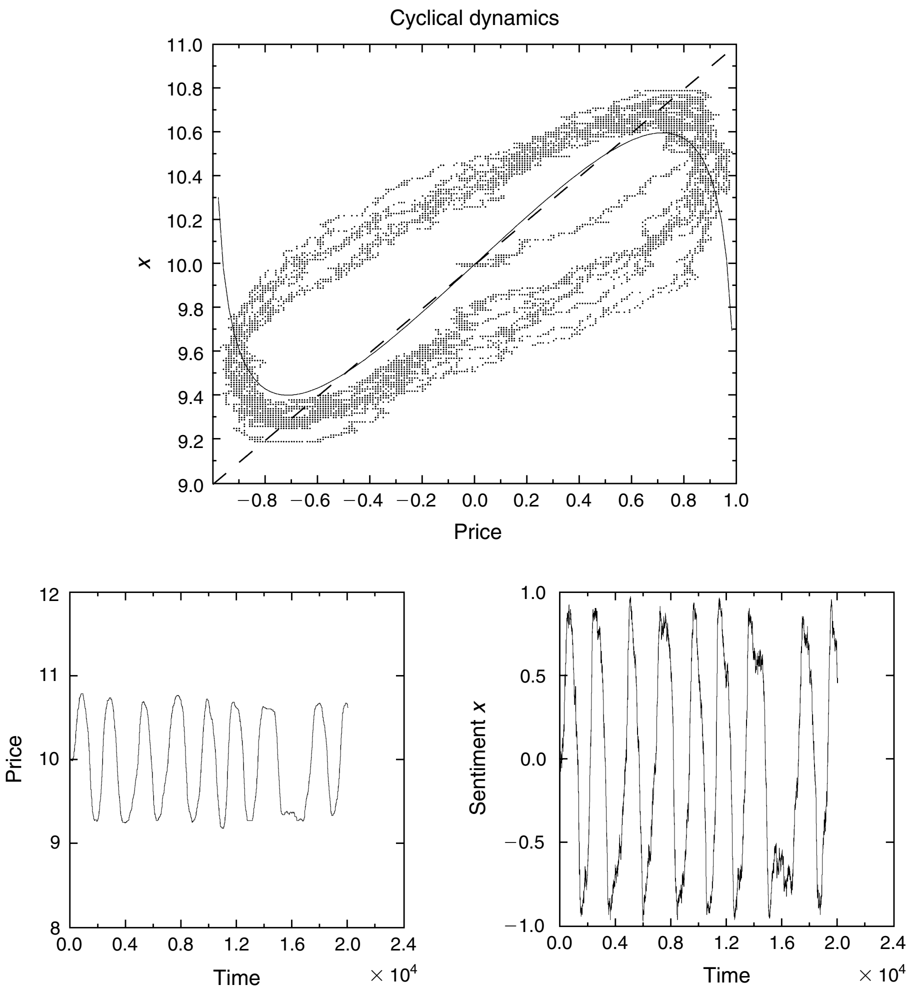

FIGURE 3.8b The case of cyclical variation between bullish and bearish phases. Parameters as before except for $\alpha_1 = 1.1$ and $\alpha_2 = 0.95$.

$$p_{-f} = v_2 \frac{n_-}{2N} \exp(U_{2,2})$$

$$p_{f-} = v_2 \frac{n_f}{2N} \exp(-U_{2,2})$$

The forcing functions $U_{2,1}$ and $U_{2,2}$ depend on the difference between the momentary profits earned by noise traders and fundamentalists, respectively. We specify these functions as

$$U_{2,1} = \alpha_3 \left\{ \frac{r + \frac{1}{v_2}\frac{dP_t}{dt}}{P_t} - R - s\left|\frac{P_f - P_t}{P_t}\right| \right\} \tag{3.53}$$

$$U_{2,2} = \alpha_3 \left\{ R - \frac{r + \frac{1}{v_2}\frac{dP_t}{dt}}{P_t} - s|\frac{P_f - P_t}{P_t}| \right\} \quad \textbf{(3.54)}$$

The first term of both functions represents the current profit of noise traders from the optimistic and pessimistic camps, respectively. The second term is the expected profit of fundamentalists after reversal to the fundamental value. Excess profits of the optimistic chartists consist of nominal dividends (r) and capital gains (dP_t/dt). Division by the actual market price (P_t) yields the revenue per unit of the asset. Subtracting the average real returns of alternative investments (or safe interest rate R) gives excess returns. Pessimistic noise traders, in contrast, leave the market so that their excess profits consist of the alternative return R minus the sum of forgone dividends plus capital gains of the pertinent stock. It is somewhat more difficult to come up with a formalization of fundamentalists' profits.

Fundamental activity is based on a perceived discrepancy between the market price and the fundamental value $P_t \neq P_f$. Profits from the pertinent traders are, however, *expected* profits only and will be materialized only if the stock price will have reverted toward its fundamental value. Because of the time needed for a reversal toward fundamental valuation and the potential uncertainty of this reversal, expected profits by fundamentalists have to be discounted by a factor $s < 1$. Otherwise, we treat fundamentalist speculation in periods of overvaluation and undervaluation symmetrically by computing the expected gain per unit of the asset as $|\frac{P_f - P_t}{P_t}|$. Note that the fundamentalists' profits did not contain dividends: This negligence is due to the assumption that they use the long-run expected asset price P_f for computing real dividends and that $r/P_f = R$; that is, (risk-adjusted) dividends are the same for alternative investments if the price is equal to its fundamental value.

This new component endogenizes the fraction of chartists and fundamentalists, which necessitates some adjustment in the $x - P$ dynamics as well. In particular, the opinion index x now refers to the numbers of optimists and pessimists within the noise trader group, whose overall population is also changing over time, $x = \frac{n_+ - n_-}{n_c}$. Furthermore, the formalization of excess demand has to take into account the changing numbers of noise traders and fundamentalists as well. Denoting by z the fraction of noise traders, $z = \frac{n_c}{2N}$, we modify Eq. 3.51 accordingly:

$$\frac{dP_t}{dt} = \beta(ED_c + ED_f) = \beta(\overline{x_t}\,\overline{z_t}\,T_c + (1 - \overline{z_t})T_f(P_f - P_t)) \quad \textbf{(3.55)}$$

Investigating the overall mean-value dynamics, the system evolution can be characterized by the time change of the expectations of x, z, and p. We restrict ourselves here to reporting the main results for the pertinent system of three differential equations. As detailed in Lux and Marchesi (2000), for this quasi-deterministic system the following characterization of its steady states can be obtained:

- There are three types of steady state:

  (i) $\overline{x_t^*} = 0, P_t^* = P_f$ with arbitrary z_t

  (ii) $\overline{x_t^*} = 0, \overline{z_t^*} = 1$ with arbitrary P_t

(iii) $\overline{z_t^*} = 0$, $P_t^* = P_f$ with arbitrary x_t

- No steady states exist with both $\overline{x_t^*} \neq 0$ and $P_t^* \neq P_f$

The second result indicates that the additional assumption of switching between strategies due to profit differentials prevents emergence of stationary bubbles. Quite obviously, such lasting situations of overvaluation or undervaluation would give rise to differences in profits between groups so that they could not persist any more.

The first part of the results indicates the types of equilibria that would be admitted under flexible strategy choice; there could be either a price equal to its fundamental value (on average) together with a balanced disposition of noise traders and an *arbitrary* composition of the overall population with respect to noise traders and fundamentalist strategy (i), there could be a dominance of noise traders ($\overline{z_t^*} = 1$) with an arbitrary price development (ii), or there could be a dominance of fundamentalists with P_t again equal to P_f on average (iii). All three categories are continua of equilibria rather than isolated fixed points, since there is one "free" variable. The more interesting of these possibilities is (i), whereas (ii) and (iii) are relatively uninteresting so-called absorbing states whose existence is difficult to avoid in a population dynamic (if one group dies out by chance, it has no way to get into existence again).

Inspecting Type i equilibria, their most interesting feature is the indeterminateness of the composition of the population (i.e., of z). After some reflection, this outcome seems quite natural: If agents are allowed to switch between strategies, then in an equilibrium none of the surviving strategies should have a higher pay-off than others. This is the case in our model almost by definition of a steady state: If there are no longer any price changes and the price is equal to its fundamental value, both the noise traders and fundamentalists would report excess profits equal to zero. Switching between subgroups would then occur unsystematically, leading to permanent changes of z along the continuum of steady states due to the stochastic elements of our process.

The set of results obtained for the mean-value dynamics of the extended model appears to indicate that the slight steps toward more rationality of agents represented in Eqs. 3.53 and 3.54 weed out the weird cyclical and bubble processes of the simpler model presented in the previous section. In any case, the interesting equilibria are characterized by a price fluctuating around its fundamental value and a balanced disposition of noise traders, that is, no more buildup of coherent optimistic or pessimistic majorities. A glance at the stability properties of these equilibria adds some additional insights.

As shown in Lux and Marchesi (2000), an equilibrium along the line $\left(\overline{x_t^*} = 0, P_t^* = P_f, z\right)$ is unstable[13] if

$$2zv_1\left(\alpha_1 + \alpha_2\frac{\beta}{v_1}zT_c - 1\right) + 2(1-z)\alpha_3\beta zT_c/P_f - \beta(1-z)T_f > 0 \qquad \textbf{(3.56)}$$

[13]Since we have a continuum rather than isolated fixed points, it is more convenient to express the stability properties in terms of conditions for instability rather than for stability.

or

$$\alpha_1 > 1 + \alpha_3 \frac{v_2 T_c R}{v_1 T_f P_f} \tag{3.57}$$

holds. The most interesting aspect of these results is that Eq. 3.56 defines a region $z \in [0, \bar{z}]$ in which the dynamics reverts to the continuum after disturbances, while for z beyond the threshold value $\bar{z}$ the dynamic becomes unstable. Since the stochastic components lead to ongoing changes of z *along* the continuum of equilibria, the system might wander from time to time from the subset of stable equilibria ($z < \bar{z}$) to that with repelling dynamics. The instability in the region $z > \bar{z}$ is due to the strong reaction on price changes in a population dominated by noise traders. Figure 3.9 shows that stronger fluctuations set in if the system is close to or surpasses the threshold $\bar{z}$. Apparently these fluctuations have the appearance of volatility clusters. They hold on for some time but die out due to inherent stabilizing tendencies that become effective out of equilibrium. Namely, strong fluctuations lead to relatively large deviations from the fundamental value, and former noise traders are induced to switch to fundamentalist behavior in large numbers.

The combination of deterministic and stochastic forces (incorporated in the stochastic formalization of agents' behavior) leads to repeated switches between turbulent and tranquil episodes. It is worthwhile to emphasize that despite a certain number of free behavioral parameters, the qualitative outcome of this process is entirely *generic; all* combinations of parameters lead to a continuum of equilibria with stochastic switching between attractive and repulsive phases.[14]

As demonstrated in Lux and Marchesi (1999) and Chen, Lux, and Marchesi (2001), the apparent proximity of simulated returns to empirical records is reflected in the agreement of many important statistics of simulated time series with empirical stylized facts. In particular, both the scaling laws of large returns and the hyperbolic decay of autocorrelations of squared and absolute returns are reproduced by the data from this artificial market, and the pertinent estimates of, for example, tail indices and decay exponents of autocorrelations of squared and absolute returns are numerically close to their typical values for empirical data. Switching between strategies also eliminates autocorrelations in raw returns to a large extent so that the apparent predictability of cyclical ups and downs of the simpler model of the previous section does not carry over to the extended framework. The lack of predictability seems plausible since the outbreak of fluctuations is triggered by the stochastic part of unsystematic population movements in the vicinity of the fundamental equilibrium. As a result, the market appears to be characterized by *speculative efficiency*. Allowing for an additional news arrival process, the market price is found to closely track the fundamental value, albeit with temporary deviations that manifest themselves in a broader leptokurtotic distribution of returns compared to changes of the fundamental value.

[14] The dynamic is also qualitatively similar in the extreme cases where $\bar{z} = 0$ because stabilizing forces out of equilibrium still prevail.

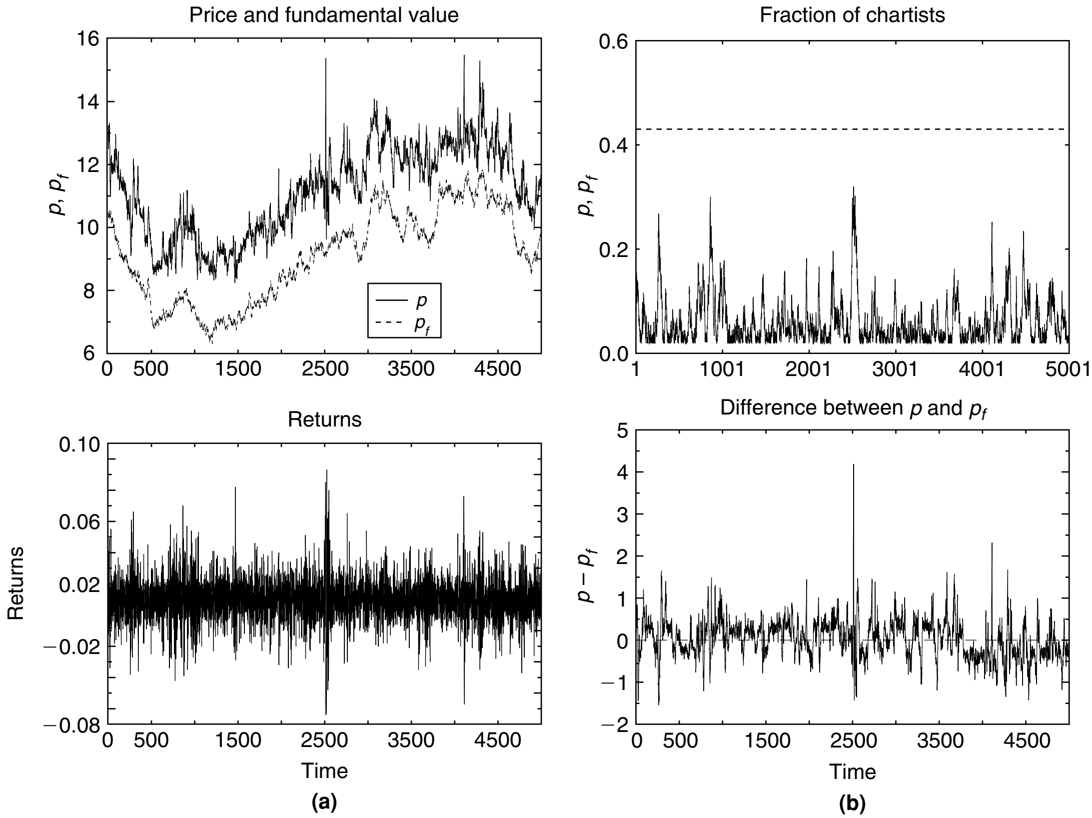

FIGURE 3.9 Example of a simulation of the model of Lux and Marchesi (1999). Parameters are $\alpha_1 = 0.8, \alpha_2 = 1, \alpha_3 = 0.5, v_1 = 1, v_2 = 0.6, T_c = T_f = 2.5, s = 0.75$, and $R = 0.0004$. The fundamental value has been shifted downward by two units to provide better visibility. As can be seen, price most closely tracks the fundamental value but shows occasional large deviations from this benchmark. The development of the fraction of noise traders or chartists in the top diagram in (a) indicates that large fluctuations of returns and large degrees of mispricing occur if many traders follow the chartist strategy. The broken line in the top diagram in (b) demarcates the theoretical bifurcation value, $\bar{z} = 0.46$ in this case.

Lux and Marchesi (2000) argue that the underlying mechanism of periodic switching between stable and unstable states due to stochastic forces constitutes a relatively general scenario to generate realistic ARCH type dynamics. In behavioral models, it seems natural that an equilibrium will be characterized by an arbitrary mixture of equally successful strategies and that the stability of such steady states will depend on the current distribution of strategies among the population. Models with similar features have been proposed by Giardina and Bouchaud (2003), with a larger set of strategies, as well as Arifovic and Gencay (2000) and Lux and Schornstein (2005). The latter have a totally different setup—a two-country general equilibrium model of the foreign exchange market with agents choosing consumption and interest strategies via genetic algorithms. Despite this very different framework, the dynamics of returns seem to be governed by a mechanism similar to that of the previously described stock market dynamics: There is a continuum of steady states with indeterminateness of investment decisions (in steady state, the revenue from domestic and foreign assets is the same), but random deviations from the steady state (brought about by the inherent randomness of genetic algorithms) destabilize this steady state and lead to the onset of fluctuations.

The original framework by Lux and Marchesi has recently been extended by Pape (2007a, b), who reformulates traders' behavior as position-based trading (in this way keeping track of their inventories) and adds both a second risky asset and a risk-free bond. As it turns out, the main mechanisms of the original model are still found to be at work in this richer setup, leading to similarly realistic simulations. It is worthwhile to point out that the combination of stochastic and deterministic forces in these models is also similar to that of Kirman's population dynamics reviewed in Section 3.4.2. in that it leads to movements back and forth across a stability threshold of the underlying deterministic benchmark system.

3.4.4. Lattice Topologies of Agents' Connections

The models reviewed in Sections 3.4.2 and 3.4.3 are among the first contributions to allow for social interactions among agents in an economic context. However, they adopt very different assumptions for the design of their social interactions: Kirman (1993) allows for pairwise interactions only (after random encounters of agents), whereas Lux (1995) and Lux and Marchesi (1999) use a mean-field approach. The latter implies that all agents influence all other agents with the same intensity or, in the language of network theory, that the social interactions are embedded in a fully connected network with equal weights of its nodes. Although we do not have reliable information on market participants' social networks, both of these alternatives might not be very realistic. Even if a certain simple topology of interactions might be acceptable for a first approach toward social influences, one could be concerned about the intensity of social interactions in relation to the size of the market (the number of agents).

As Egenter et al. (1999) show, stylized facts do vanish in the model of Lux and Marchesi (1999) if one increases the number of agents while keeping the parameters of social interactions constant. The reason is that, due to the law of large numbers, fluctuations of noise traders' moods become more and more moderate with increasing N. With the opinion index x staying close to its steady-state value $\overline{x^*} = 0$ most of the time, the emergence of an optimistic or pessimistic majority occurs less often so that the frequency of small price bubbles declines. The intrinsic chartist "information" component then loses its importance against the fundamental component in Eqs. 3.53 and 3.54 so that the profit differential works in favor of the fundamentalist strategy. As a consequence, the average fraction of noise traders gets smaller and smaller with increasing N and the distribution of returns gets closer and closer to the assumed Gaussian distribution of the news arrival process. Therefore, the "interesting dynamics" with their fat tails and clustered volatility are a *finite size* effect and do not survive in the limit $N \to \infty$. Essentially, this is a consequence of the law of large numbers as the market excess demand is an aggregate over N Poisson processes for individual traders. Obviously, the correlation between agents brought about by their social interactions is not strong enough to undo the effect of aggregation.

With more than about 5000 socially interacting agents, the model converges to returns following a pure white noise. A similar result is obtained for the very different artificial foreign exchange market with genetically generated strategies in Lux and Schornstein (2005). As it seems, the genetic operations of selection, recombination, and mutation also lead to a reduced intensity of interpersonal coupling with an increasing number of agents so that the dynamics loses its stochastic appearance with increasing numbers of market participants. Again, interesting and realistic dynamics are only obtained for markets with up to a few thousand traders. These findings are disturbing in so far as empirical stylized facts are observed in quite the same way with practically the same estimated scaling exponents for markets of all sizes. In this sense, the *universality* of the empirical records is not reproduced by the preceding stochastic models. On the other hand, the universality of non-Gaussian behavior of all known financial markets implies that there probably is strong coupling between traders in real life. With the largest markets having populations on the order of 10^6 or more market participants, the law of large numbers would imply Gaussian behavior if all these agents would act independently (or with sufficiently weak correlation). The universal non-Gaussianity, then, appears to indicate that financial markets have a typical number of *effectively independent* agents that is much smaller than their nominal number of market participants.

The challenge for models of social interactions, therefore, would be to come up with an explanation of this lack of sensitivity with respect to system size. Alfarano, Lux, and Wagner (2008) discuss this problem for a variant of Kirman's ant model. They show that if the frequency of pairwise encounters increases linearly with the number of agents, the resulting dynamic remains qualitatively the same for any number N of agents. In contrast, if the frequency of encounters is kept constant, the system converges to a Gaussian limit with increasing N. Quite similar to the experiments of Egenter et al. (1999), the relative importance of the herding component against the autonomous switching propensity declines if one does not adjust the former to the system size. The *intensity of interpersonal coupling* is only preserved in this model if the frequency of pairwise exchange increases with the number of potential partners for exchange.[15] Certainly, an ever-increasing probability of pairwise exchange is somewhat difficult to digest in its literal interpretation.

Departing from the extremes of either pairwise interactions or fully connected social systems, network topologies of agents' social interactions might be a promising avenue to explore the way the intensity of social coupling might plausibly change with system size. Alfarano and Milaković (2007) modify the ant model by replacing pair interactions with neighborhood effects within various network topologies. Increasing the number of agents but keeping the parameters of the network-generating mechanism fixed, they note

[15]Finite-size effects in alternative models of opinion formation are investigated in Toral and Tessone (2007).

that most popular network designs (regular, scale-free, and "small-world" networks) cannot overcome the N dependency within their generating mechanism, that is, without adapting crucial parameters. The only case in which the generating mechanism keeps the intensity of communication constant for varying numbers of agents by the very nature of its construction is the random network. Although these results sound somewhat disappointing, they have so far only focused on the mechanical structure of various topologies. Incentives of agents to form links could lead to changes of the connectivity with system size, which remains to be investigated. Interestingly, Alfarano and Milaković (2007) also show that allowing for a small number of independent agents who only influence others without being prone to social influences themselves (unilateral links) changes the outcome and allows for prevalence of interesting dynamics, whatever the number of herding agents (see also Schmalz, 2007).

A different type of network structure has been used in a related paper by Cont and Bouchaud (2000). Essentially, their contribution is an adaptation of the seminal percolation model from statistical physics. In this framework, agents are situated on a lattice with periodic boundary conditions. Each site of this lattice might initially be "occupied" with a certain probability p or empty with probability $1 - p$. Groups of occupied neighboring states form clusters. In Cont and Bouchaud (2000), occupied sites are traders and clusters are subsets of synchronized trading behavior (i.e., all members of a cluster are buyers or sellers or remain inactive). The type of activity of a cluster is determined via random draws. The market price is again driven by an auctioneer equation depending on excess demand over all clusters. Since the underlying formal structure has been extensively studied in physics, certain known results for the cluster size distribution can be evoked, and due to the simple link between cluster distribution and price changes, these known results also carry over to returns. In particular, both distributions will follow a power law if the probability for the connection of lattice sites is close to a critical value, the so-called percolation threshold. However, the power law is characterized by an exponent 1.5, in contrast to the empirical law with decay rate ~ 3. In the baseline version of the model, higher moments are uncorrelated so that the percolation model could not explain volatility clustering, either.

Despite (or because of) these deficits, the framework of Cont and Bouchaud has spawned a sizable literature (mostly published in physics periodicals) that tries to get its time series characteristics closer to empirical scaling laws. Interesting extensions of the original model include Stauffer et al. (1999) and Eguiluz and Zimmerman (2000), who generate autocorrelations in higher moments via sluggish changes of cluster configurations. In view of the previous discussion, it is worthwhile to note that the critical connection probability at the percolation threshold is N dependent and, therefore, has to be adjusted with system size to guarantee a power-law distribution of the clusters.

More realistic time series are obtained in some alternative lattice models: Iori (2002) considers an Ising-type model with interactions restricted to nearest neighbors, whereas Bartolozzi and Thomas (2004) propose a cellular automaton structure with a similar neighborhood structure. In both models, realistic time series seem to be a robust outcome without the need for fine-tuning certain parameter values. However, due to the complexity of these structures, it is hard to single out the key features of these models that are responsible for the interesting dynamics. It is unknown so far whether the realistic features of those models will persist for large populations of traders or not.

3.5. CONCLUSION

This chapter reviewed recent models that try to explain the characteristics of financial markets as emergent properties of interactions and dispersed activities of a large ensemble of agents populating the marketplace. This view has a certain tradition starting in the early 1990s (or even earlier if one includes contributions of the 1970s, such as Zeeman's, 1974) when chaotic processes based on simple behavioral assumptions have been proposed as an explanation of the apparent randomness of financial data. As it turned out in subsequent research, market statistics are in all likelihood more "complex" than data from low-dimensional chaotic attractors and seem to be characterized by an intricate mixture of randomness and nonlinear structure in higher moments. The most pervasive characteristics of the particular stochastic nature of financial markets are the power laws for large returns and autocorrelations of volatility. Similar systemwide features are the typical imprints of large systems of interacting subunits in the natural sciences.

Inspired by these analogies, some recent models have proposed simple structures that could reproduce the empirical findings to a high degree with statistics that are even quantitatively close to empirical ones. This appears the more remarkable since "mainstream" theory has offered hardly any hint at the generating forces behind the stylized facts, let alone models with precise numerical predictions. Offering explanations for hitherto unexplained observations is typically what characterizes a new, superior paradigm. This new view also opens the stage for entirely new avenues of research and questions that could not even have been formulated before. Among these questions, the most important task for future research might be the explanation of the *universal preasymptotic* behavior of financial markets, that is, the answer to the question of why they are not subject to the law of large numbers (as they should be, if they were populated by independent agents).

From the viewpoint of mainstream finance, it might be a perplexing experience to see some basic stylized facts explained by models that have hardly anything in common with a traditional representative-agent approach. However, what these models offer are simply those ingredients that critics of the mainstream have been emphasizing for a long time. As a prominent example, Kindleberger (1989) has stressed the importance of psychological factors and irrational behavior in explaining historical financial crises. In fact, recent microstructure literature has allowed for irrational components such as overconfidence or framing (e.g., Daniel et al., 1998; Barberis and Huang, 2001), with highly interesting results. The analysis of certain types of nonrational behavior and its consequences might explain important facets of reality, but an explanation of the overall characteristics of the market might require a different approach.

Proponents of mainstream finance have, in fact, criticized the lack of a unifying framework in the behavioral finance literature. Most notably, Fama (1998) pointed out that a variety of psychological biases could be used to explain various anomalies but that behavioral finance models were unable to explain the "big picture" and to capture the "menu of anomalies better than market efficiency" (Fama, 1998, p. 241). Though stochastic models of interacting agents have so far not focused on overreaction and other return anomalies, they appear to be able to provide generic explanations for the "deeper" anomalies of fat tails and volatility clustering. Although they are mostly not micro-based in the sense of featuring utility maximization or alternative psychological decision mechanisms, they might provide a broader macroscopic picture of emergent properties of microeconomic interaction embedding the wide spectrum of diverse deviations from perfect rationality at the micro level. Since we probably encounter a wide variety of trading motives, strategies, and degrees of (non-)rationality and (lack of) foresight among agents, a stochastic approach might be required to compensate for our ignorance of the microscopic details. This is the starting point of the preceding models. The present stochastic approach could, therefore, be seen as complementary to the focus of the previous strands of the behavioral finance literature on particular behavioral observations in that it tries to infer macroscopic regularities via a simple representation of the diverse collection of the boundedly rational behaviorial types in real markets.

References

Abhyankar, A., L. Copeland, and W. Wong, Moment condition failure in high-frequency financial data: Evidence from S&P 500, *Applied Economics Letters*, 1995, 2, 288–290.

Alfarano, S., and T. Lux, A noise trader model as a generator of apparent power laws and long memory, *Macroeconomic Dynamics*, 2007, 11, 80–101.

Alfarano, S., T. Lux, and F. Wagner, Estimation of agent-based models: The case of an asymmetric herding model, *Computational Economics*, 2005, 26, 19–49.

Alfarano, S., T. Lux, and F. Wagner, Time-variation of higher moments in financial markets with heterogeneous agents: An analytical approach, *Journal of Economic Dynamics and Control*, 2008, 32, 101–136.

Alfarano, S., and M. Milaković, Should network structure matter in agent-based finance?, Christlan-Albrechts-Universität Kiel, Economics working paper 2008-04.

Allen, H., and M. Taylor, Charts, noise and fundamentals in the London foreign exchange market, *Economic Journal*, 1990, 100, 49–59.

Aoki, M., *New Approaches to Macroeconomic Modelling: Evolutionary Stochastic Dynamics, Multiple Equilibria and Externalities as Field Effects*, Cambridge University Press, 1996.

Arifovic, J., The behavior of the exchange rate in the genetic algorithm and experimental economies, *Journal of Political Economy*, 1996, 104, 510–541.

Arifovic, J., and R. Grencay, Statistical properties of genetic learning in a model of exchange rate, *Journal of Economic Dynamics and Control*, 2000, 24, 981–1005.

Arthur, W. B., J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler, Asset pricing under endogenous expectations in an artificial stock market, *Economic Notes*, 1997, 26, 297–330.

Avnir, D., O. Biham, D. Lidar, and O. Malcai, Is the geometry of nature fractal?, *Science*, 1998, 279, 39–40.

Baillie, R. T., T. Bollerslev, and H. Mikkelsen, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 1996, 74, 3–30.

Ball, P., Culture crash, *Nature*, 2006, 441, 686–688.

Barberies, N., and M. Huang, Mental accounting, loss aversion and individual stock returns, *Journal of Finance*, 2001, 56, 1247–1295.

Barnsdorf-Nielsen, O., and K. Prause, Apparent scaling, *Finance and Stochastics*, 2001, 5, 103–113.

Bartolozzi, M., and A. Thomas, Stochastic cellular automata model for stock market dynamics, *Physical Review*, 2004, E 69, 46112–46117.

Baumol, W. J., Speculation, profitability and stability, *Review of Economics and Statistics*, 1957, 39, 263–271.

Beirland, J., P. Vynckier, and J. Teugels, *Practical Analysis of Extreme Values*, University Press, Leuven, 1996.

Beja, A., and M. B. Goldman, On the dynamic behavior of prices in disequilibrium, *Journal of Finance*, 1980, 35, 235–248.

Beran, J., *Statistics for Long-Memory Processes.* Chapman and Hall, 1994.

Bollerslev, T., and J. Wright, Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data, *Journal of Econometrics*, 2000, 98, 81–106.

Bonabeau, E., M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, 1999.

Bottazzi, G., G. Dosi, and I. Rebesco, Institutional architectures and behavior ecologies in the dynamics of financial markets, *Journal of Mathematical Economics*, 2005, 41, 197–228.

Bouchaud, J.-P., and R. Cont, Herd behavior and aggregate fluctuations in financial market, *Macroeconomic Dynamics*, 2000, 2, 170–196.

Box-Steffensmeier, J., and R. Smith, The dynamics of aggregate partisanship, *American Political Science Review*, 1996, 90, 567–580.

Braglia, M., On the variance of stock price fluctuations, *Economics Letters*, 1996, 33, 171–173.

Brock, W., Scaling in economics: A reader's guide, *Industrial and Corporate Change*, 1999, 8, 409–446.

Brock, W. A., and C. H. Hommes, A rational route to randomness, *Econometrica*, 1997, 65, 1059–1095.

Brock, W. A., and C. H. Hommes, Heterogeneous beliefs and bifurcation routes to chaos in a simple asset pricing model, *Journal of Economic Dynamics and Control*, 1998, 22, 1235–1274.

Brock, W., C. Hommes, and F. Wagener, Evolutionary dynamics in markets with many trader types, *Journal of Mathematical Economics*, 2005, 41, 7–42.

Brown, G., and M. Cliff, Investor sentiment and asset valuation, *Journal of Business*, 2005, 78, 405–440.

Calvet, L. E., and A. J. Fisher, Forecasting multifractal volatility, *Journal of Econometrics*, 2001, 105, 27–58.

Chen, S.-H., T. Lux, and M. Marchesi, Testing for non-linear structure in an artificial financial market, *Journal of Economic Behavior and Organization*, 2001, 46, 327–342.

Chen, S.-H., and C.-H. Yeh, On the emergent properties of artificial stock markets: The efficient market hypothesis and the rational expectations hypothesis, *Journal of Economic Behavior and Organization*, 2002, 49, 217–239.

Chiarella, C., The dynamics of speculative behavior, *Annals of Operations Research*, 1992, 37, 101–123.

Chiarella, C., R. Dieci, and L. Gardini, Speculative behavior and complex asset price dynamics, *Journal of Economic Behavior and Organization*, 2002, 49, 173–197.

Chiarella, C., and T. He, Heterogeneous beliefs, risk and learning in a simple asset pricing model, *Computational Economics*, 2002, 14, 95–132.

Crato, N. and P. de Lima, Long-range dependence in the conditional variance of stock returns, *Economics Letters*, 1994, 45, 281–285.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet, *An Introduction to High-Frequency Finance*, Academic Press, 2001.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam, Investor psychology and security market under- and overreaction, *Journal of Finance*, 1998, 53, 1839–1885.

Dawid, H., On the convergence of genetic learning in a double auction market, *Journal of Economic Dynamics and Control*, 1999, 23, 1545–1567.

Day, R. H., and W. Huang, Bulls, bears, and market sheep, *Journal of Economic Behavior and Organization*, 1990, 14, 299–329.

DeGrauwe, P., H. Dewachter, and M. J. Embrechts, *Exchange Rate Theory: Chaotic Models of Foreign Exchange Markets*, Blackwell, 1993.

DeGrauwe, P., and M. Grimaldi, *The Exchange Rate in a Behavioral Finance Framework*, Princeton University Press, 2006.

de Lima, P. J. F., Nonlinearities and nonstationarities in stock returns, *Journal of Business and Economic Statistics*, 1998, 16, 227–236.

Deneubourg, J.-L., S. Aron, S. Goss, and J. M. Pasteels, The self-organizing exploratory pattern of the Argentine ant, *Journal of Insect Behavior*, 1990, 3, 159–168.

de Vries, C. G., Stylized facts of nominal exchange rate returns, in F. van der Ploeg (ed.), *The Handbook of International Macroeconomics*, pp. 348–389, Blackwell, 1994.

Diebold, F., and A. Inoue, Long memory and regime switching, *Journal of Econometrics*, 2001, 105, 131–159.

Ding, Z., R. Engle, and C. Granger, A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1993, 1, 83–106.

Durlauf, S., Complexity and empirical economics, *Economic Journal*, 2005, 115, 225–243.

Eckmann, J.-P., S. Kamphorst, D. Ruelle, and J. Scheinkman Lyapunov exponents for stock markets, in P. Anderson, J. Arrow, and D. Pines (eds.), *The Economy as a Complex Evolving System*, Addison-Wesley, 1988.

Egenter, E., T. Lux, and D. Stauffer, Finite-size effects in Monte Carlo simulations of two stock market models, *Physica A*, 1999, 268, 250–256.

Eguiluz, V. M., and M. G. Zimmermann, Transmission of information and herd behavior: An application to financial markets, *Physical Review Letters*, 2000, 85, 5659–5662.

Engle., R., Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 1983, 50, 987–1008.

Fama, E., Mandelbrot and the stable Paretian hypothesis, *Journal of Business*, 1963, 36, 420–429.

Fama, E., Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics*, 1998, 49, 283–306.

Fergusson, K., and E. Platen, On the distributional characterization of daily log-returns of a world stock index, *Applied Mathematical Finance*, 2006, 13, 14–38.

Föllmer, H., Random economies with many interacting agents, *Journal of Mathematical Economics*, 1974, 1, 51–62.

Frankel, J., and K. A. Froot, Understanding the U.S. dollar in the eighties: The expectations of chartists and fundamentalists, *Economic Record*, 1986, special issue, 24–38.

Frankfurter, G., and E. McGoun, Anomalies in finance: What are they and what are they good for?, *International Review of Financial Analysis*, 2001, 10, 407–429.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, A theory of power-law distributions in financial market fluctuations, *Nature*, 2003, 423, 267–270.

Gallegatti, M., S. Keen, T. Lux, and P. Ormerod, Worrying trends in econophysics, *Physica A*, 2006, 370, 1–6.

Gaunersdorfer, A., Endogenous fluctuations in a simple asset pricing model with heterogeneous agents, *Journal of Economic Dynamics and Control*, 2000, 24, 799–831.

Gaunersdorfer, A., and C. H. Hommes, A nonlinear structural model for volatility clustering, in A. Kirman, and G. Teyssière (eds.), *Long Memory in Economics*, Springer, 2007.

Georges, C., Learning with misspecification in an artificial currency market, *Journal of Economic Behavior and Organization*, 2006, 60, 70–84.

Giardina, I., and J.-P. Bouchaud, Bubbles, crashes and intermittency in agent-based market models, *European Physical Journal B*, 2003, 31, 421–437.

Gielens, G., S. Straetmans, and C. de Vries, Fat-tail distributions and local thin-tail alternatives, *Communications in Statistics: Theory and Methods*, 1996, 25, 705–710.

Gilli, M., and P. Winker, A global optimization heuristic for estimating agent-based models, *Computational Statistics and Data Analysis*, 2003, 42, 299–312.

Glosten, L., and P. Milgrom, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Markets*, 1985, 14, 71–100.

Gopikrishnan, P., M. Meyer, L. A. N. Amaral, and H. E. Stanley, Inverse cubic law for the probability distribution of stock price variations, *European Journal of Physics B*, 1998, 3, 139–140.

Gopikrishnan P., V. Plerou, X. Gabaix, L. A. N. Amaral, and H. E. Stanley, Price fluctuations, market activity and trading volume, *Quantitative Finance*, 2001, 1, 262–270.

Granger, C., Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 1980, 14, 227–238.

Granger, C., and Z. Ding, Variations of long memory models, *Journal of Econometrics*, 1996, 73, 61–77.

Gu, M., Market mediating behavior: An economic analysis of the security exchange specialist, *Journal of Economic Behavior and Organization*, 1995, 27, 237–256.

He, X.-Z., and Y. Li, Power-law behavior, heterogeneity, and trend chasing, *Journal of Economic Dynamics and Control*, 2007, 31, 3396–3426.

Horgan, J., From complexity to perplexity, *Scientific American*, 1995, 272, 74–79.

Iori, G., A micro-simulation of traders' activity in the stock market: The role of heterogeneity of agents' interactions and trade-frictions, *Journal of Economic Behavior and Organization*, 2002, 49, 269–285.

Jansen, D., and C. de Vries, On the frequency of large stock returns: Putting booms and busts into perspective, *The Review of Economics and Statistics*, 1991, 73, 18–24.

Jensen, H., *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems,* Cambridge University Press, 1998.

Kaldor, N., Speculation and economic stability, *Review of Economic Studies*, 1939, 7, 1–27.

Kindleberger, P., *Manias, Panics and Crashes: A History of Financial Crises*, MacMillan, 1989.

Kirman, A., Epidemics of opinion and speculative bubbles in financial markets, in M. Taylor (ed.), *Money and Financial Markets*, MacMillan, 1991.

Kirman, A., Whom or what does the representative agent represent?, *Journal of Economic Perspectives*, 1992a, 6, 117–136.

Kirman, A., *The Significance of Testing for Bubbles*, manuscript, European University Institute, 1992b.

Kirman, A., Ants, rationality, and recruitment, *Quarterly Journal of Economics*, 1993, 108, 137–156.

Kirman, A., and G. Teyssière, Microeconomic models for long memory in the volatility of financial time series, *Studies in Nonlinear Dynamics and Econometrics*, 2002, 5, 137–156.

Koedijk, K., M. Schafgans, and C. de Vries, The tail index of exchange rate returns, *Journal of International Economics*, 1990, 29, 93–108.

Kyle, A., Continuous auctions and insider trading, *Econometrica*, 1985, 53, 1315–1335.

LeBaron, B., Stochastic volatility as a simple generator of power laws and long memory, *Quantitative Finance*, 2001, 1, 621–631.

LeBaron, B., W. B. Arthur, and R. Palmer, The time series properties of an artificial stock market, *Journal of Economic Dynamics and Control*, 1999, 23, 1487–1516.

Lee, W., C. Jiang, and B. Indro, Stock market volatility, excess returns, and the role of investor sentiment, *Journal of Banking and Finance*, 2002, 26, 2277–2299.

Leroy, S. F., Efficient capital markets and martingales, *Journal of Economic Literature*, 1989, 27, 1583–1621.

Liu, P., The effects of fundamentalists' and chartists' expectations on market survey, *Applied Financial Economics*, 1996, 6, 363–366.

Lobato, I., and N. E. Savin, Real and spurious long-memory properties of stock market data, *Journal of Business and Economics Statistics*, 1998, 16, 261–283.

Lobato, I., and C. Velasco, Long memory in stock market trading volume, *Journal of Business and Economic Statistics*, 1998, 18, 410–427.

Longin, F., The asymptotic distribution of extreme stock market returns, *Journal of Business*, 1996, 69, 383–408.

Lux, T., Herd behaviour, bubbles and crashes, *Economic Journal*, 1995, 105, 881–896.

Lux, T., The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks, *Applied Financial Economics*, 1996, 6, 463–475.

Lux, T., Time variation of second moments from a noise trader/infection model, *Journal of Economic Dynamics and Control*, 1997, 22, 1–38.

Lux, T., The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions, *Journal of Economic Behavior and Organization*, 1998, 33, 143–165.

Lux, T., The limiting extremal behavior of speculative returns: An analysis of intradaily data from the Frankfurt stock exchange, *Applied Financial Economics*, 2001, 11, 299–315.

Lux, T, and M. Ausloos, Market fluctuations I: Scaling, multiscaling and their possible origins, in Bunde, A., J. Kropp, and H.-J. Schellnhuber (eds.), *Theories of Disasters: Climate Disruptions, Heart Attacks and Market Crashes*, Springer, pp. 372–409, 2002.

Lux, T., and M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature*, 1999, 397, 498–500.

Lux, T., and M. Marchesi, Volatility clustering in financial markets: A microsimulation of interacting agents, *International Journal of Theoretical and Applied Finance*, 2000, 3, 675–702.

Lux, T., and S. Schornstein, Genetic algorithms as an explanation of stylized facts of foreign exchange markets, *Journal of Mathematical Economics*, 2005, 41, 169–196.

Mandelbrot, B., The variation of certain speculative prices, *Journal of Business*, 1963, 36, 394–419.

Mandelbrot, B., A. Fisher, and L. E. Calvet, A multifractal model of asset returns, *Mimeo: Cowles Foundation for Research in Economics*, 1997.

Newman, M. E. J., Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, 2005, 46, 323–351.

O'Hara, M., *Market Microstructure Theory*, Blackwell, 1995.

Pagan, A., The econometrics of financial markets, *Journal of Empirical Finance*, 1996, 3, 15–102.

Pape, B., Asset allocation and multivariate position based trading, *Journal of Economic Interaction and Coordination*, 2007a, 2, 163–193.

Pape, B., *Asset Allocation, Multivariate Position Based on Trading, and the Stylized Facts*. Ph.D. thesis, University of Vaasa, Acta Wasaensia no. 177, 2007b.

Pellizzari, P., and A. dal Forno, A comparison of different trading protocols in an agent-based market, *Journal of Economic Interaction and Coordination*, 2007, 1, 27–43.

Ramsey, J., On the existence of macro variables and macro relationships, *Journal of Economic Behavior and Organization*, 1996, 30, 275–299.

Rockinger, M., and E. Jondeau, The tail behavior of stock returns: Emerging versus mature markets, *Journal of Empirical Finance*, 2003, 10, 559–581.

Ruelle, D., Deterministic chaos: The science and the fiction, *Proceedings of the Royal Society of London*, 1990, 427 A, 465–479.

Scheinkmann, J., and B. LeBaron, Nonlinear dynamics and stock returns, *Journal of Business*, 1988, 62, 311–327.

Schmalz, M., *Directed Gossip Algorithms and Consensus Problems in Social and Economic Systems*, manuscript, University of Stuttgart, 2007.

Stanley, M. H. R., L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and H. E. Stanley, Can statistical physics contribute to the science of economics?, *Fractals*, 1996, 4, 415–425.

Stauffer, D., P. de Oliveira, and A. Bernardes, Monte-Carlo simulation of volatility clustering in a market model with herding, *International Journal of Theoretical and Applied Finance*, 1999, 2, 83–94.

Szpiro, G. G., The emergence of risk aversion, *Complexity*, 1997, 2, 31–39.

Taylor, S., *Modeling Financial Time Series*, Wiley, 1986.

Taylor, S., *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, 2005.

Taylor, M., and H. Allen, The use of technical analysis in the foreign exchange market, *Journal of International Money and Finance*, 1992, 11, 304–314.

Toral, R., and C. Tessone, Finite-size effects in the dynamics of opinion formation, *Communications in Computational Physics*, 2007, 2, 177–195.

Vanden, J., Equilibrium analysis of volatility clustering, *Journal of Empirical Finance*, 2005, 12, 374–417.

Weidlich, W., *Sociodynamics: A Systematic Approach to Mathematical Modeling in the Social Sciences*, Harwood Academic Publishers, 2002.

Weidlich, W., and G. Haag, *Concepts and Methods of a Quantitative Sociology*, Springer, 1983.

Zeeman, E. C., On the unstable behaviour of stock exchanges, *Journal of Mathematical Economics*, 1974, 1, 39–49.

CHAPTER 4

Complex Evolutionary Systems in Behavioral Finance

Cars Hommes and Florian Wagener
CeNDEF, School of Economics, University of Amsterdam

4.1. Introduction 218
4.2. An Asset-Pricing Model with Heterogeneous Beliefs 221
4.2.1. The Fundamental Benchmark with Rational Agents 222
4.2.2. Heterogeneous Beliefs 223
4.2.3. Evolutionary Dynamics 224
4.2.4. Forecasting Rules 225
4.3. Simple Examples 226
4.3.1. Costly Fundamentalists vs. Trend Followers 227
4.3.2. Fundamentalists vs. Opposite Biases 229
4.3.3. Fundamentalists vs. Trend and Bias 230
4.3.4. Efficiency 230
4.3.5. Wealth Accumulation 232
4.3.6. Extensions 236
4.4. Many Trader Types 236
4.5. Empirical Validation 241
4.5.1. The Model in Price-to-Cash Flows 242
4.5.2. Estimation of a Simple Two-Type Example 246
4.5.3. Empirical Implications 251
4.6. Laboratory Experiments 253
4.6.1. Learning to Forecast Experiments 255
4.6.2. The Price-Generating Mechanism 256
4.6.3. Benchmark Expectations Rules 257
4.6.4. Aggregate Behavior 259
4.6.5. Individual Prediction Strategies 259
4.6.6. Profitability 262
4.7. Conclusion 264

Note: We would like to thank two anonymous references and the editor Klaus Reiner Schenk-Hoppé for detailed comments on an earlier draft, which led to significant improvements. This research was supported by the Netherlands Organization for Scientific Research (NWO) and by an EU STREP-grant "Complex Markets."

Appendix 4.1: Bifurcation Theory 266
Appendix 4.2: Bifurcation Scenarios 268
References 271

Abstract

Traditional finance is built on the rationality paradigm. This chapter discusses simple models from an alternative approach in which financial markets are viewed as complex evolutionary systems. Agents are boundedly rational and base their investment decisions on market-forecasting heuristics. Prices and beliefs about future prices co-evolve over time with mutual feedback. Strategy choice is driven by evolutionary selection so that agents tend to adopt strategies that were successful in the past. Calibration of "simple complexity models" with heterogeneous expectations to real financial market data and laboratory experiments with human subjects are also discussed.

Keywords: stylized facts, power laws, agent-based models, interacting agents

4.1. INTRODUCTION

Finance is witnessing important changes, according to some even a paradigmatic shift, from the traditional, neoclassical mathematical modeling approach based on a representative, fully rational agent and perfectly efficient markets (Muth, 1961; Lucas, 1971; Fama, 1970) to a behavioral approach based on computational models where markets are viewed as complex evolving systems with many interacting, "boundedly rational" agents using simple "rule of thumb" trading strategies (e.g., Anderson et al., 1988; Brock, 1993; Arthur, 1995; Arthur et al., 1997a; Tesfatsion and Judd, 2006). Investor's psychology plays a key role in behavioral finance, and different types of psychology-based trading and behavioral modes have been identified in the literature, such as positive feedback or momentum trading, trend extrapolation, noise trading, overconfidence, overreaction, optimistic or pessimistic traders, upward- or downward-biased traders, correlated imperfect rational trades, overshooting, contrarian strategies, and so on. Some key references dealing with various aspects of investor psychology include, for example, Cutler et al. (1989), DeBondt and Thaler (1985), DeLong et al. (1990a,b), Brock and Hommes (1997, 1998), Gervais and Odean (2001), and Hong and Stein (1999, 2003), among others—see, for example, Shleifer (2000), Hirschleifer (2001), and Barberis and Thaler (2003) for extensive surveys and many more references on behavioral finance.

An important problem of a behavioral approach is that it leaves "*many degrees of freedom*." There are many ways individual agents can deviate from full rationality. Evolutionary selection based on relative performance is one plausible way to discipline the "wilderness of bounded rationality." Milton Friedman (1953) argued that nonrational agents will *not* survive evolutionary competition and will therefore be driven out of the market, thus providing support to a representative rational agent framework as a (long run) description of the economy. In the same spirit, Alchian (1950) argued that biological evolution and natural selection driven by realized profits may eliminate nonrational, nonoptimizing firms and lead to a market in which rational, profit maximizing firms dominate. Blume and Easley (1992, 2006) have, however, shown that the market selection hypothesis does not always hold and that nonrational agents may survive in the market. Brock (1993, 1997), Arthur et al. (1997b), LeBaron et al. (1999), and Farmer (2002), among others, introduced artificial stock markets, described by agent-based models with evolutionary selection between many different interacting trading strategies. They showed that the market does not generally select for the rational, fundamental strategy and that simple technical trading strategies may survive in artificial markets. Computationally oriented agent-based simulation models have been reviewed in LeBaron (2006); see also the special issue of the *Journal of Mathematical Economics* (Hens and Schenk-Hoppé, 2005) and the survey chapter of Evstigneev, Hens, and Schenk-Hoppé (2009) in this book for an overview of evolutionary finance.[1]

Stimulated by work on artificial markets, in the last decade quite a number of "simple complexity models" have been introduced. Markets are viewed as evolutionary adaptive systems with boundedly rational interacting agents, but the models are simple enough to be at least partly analytically tractable. The study of simple complexity models typically requires a well-balanced mixture of analytical and computational tools. This literature is surveyed in Hommes (2006) and Chiarella et al. (2009); see also Lux (2009), who discusses in detail how well models with interacting agents match important stylized facts such as fat tails in the returns distribution and long memory. Without repeating an extensive survey, this chapter focuses on a number of simple examples, in particular the adaptive belief systems (ABSs) of Brock and Hommes (1997, 1998). These models serve as didactic examples of nonlinear dynamic asset-pricing models with evolutionary strategy switching, and they illustrate some of the key features present in the interacting agents literature. The model also has been used to test the relevance of the theory of heterogeneous expectations empirically as well as in laboratory experiments with human subjects. Simple complexity models may also be used by practitioners or policy makers. To illustrate this point, we present an example showing how such a model can be used to evaluate how likely it is that a stock market bubble will resume.

Two important features of the ABS are that agents are *boundedly rational* and that they have *heterogeneous expectations*. An ABS is in fact a standard discounted value asset-pricing model derived from mean-variance maximization, extended to the case of *heterogeneous beliefs*. Two classes of investors that are also observed in financial practice can be distinguished: fundamentalists and technical analysts. Fundamentalists base their forecasts of future prices and returns on economic fundamentals, such as dividends, interest rates, price-earning ratios, and so on. In contrast, technical analysts are looking for patterns in past prices and base their forecasts upon extrapolation of

[1] Some other recent references are Amir et al. (2005) and Evstigneev et al. (2002, 2008).

these patterns. Fractions of these two types of traders are time varying and depend on relative performance. Strategy choice is thus based on evolutionary selection or reinforcement learning, with agents switching to more successful (i.e., profitable) rules. Asset price fluctuations are characterized by irregular switching between a stable phase when fundamentalists dominate the market and an unstable phase when trend followers dominate and asset prices deviate from benchmark fundamentals. Price deviations from the rational expectations fundamental benchmark and excess volatility are triggered by news about economic fundamentals but may be *amplified* by evolutionary selection, based on recent performance, of trend-following strategies.

There is empirical evidence that experience-based reinforcement learning plays an important role in investment decisions in real markets. For example, Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998), Rockinger (1996), and Karceski (2002) show for mutual funds data that money flows into past good performers while flowing out of past poor performers and that performance persists on a short-term basis. Pension funds are less extreme in picking good performance but are tougher on bad performers (Del Guercio and Tkac, 2002). Benartzi and Thaler (2007) have shown that heuristics and biases play a significant role in retirement savings decisions. For example, using data from Vanguard, they show that the equity allocation of new participants rose from 58% in 1992 to 74% in 2000, following a strong rise in stock prices in the late 1990s; however, it dropped back to 54% in 2002, following the extreme fall in stock prices.

Laboratory experiments with human subjects have shown that individuals often do not behave fully rationally but tend to use heuristics, possibly biased, in making economic decisions under uncertainty (Kahneman and Tversky, 1973). In a similar vein, Smith et al. (1998) have shown the occurrence of bubbles and the ease with which markets deviate from full rationality in asset-pricing laboratory experiments. These bubbles occur despite the fact that participants had sufficient information to compute the fundamental value of the asset. Laboratory experiments with human subjects provide an important tool to investigate which behavioral rules play a significant role in deviations from the rational benchmark, and they can thus help discipline the class of behavioral modes. Duffy (2008) gives a stimulating recent overview concerning the role of laboratory experiments to explain macro phenomena.

Heterogeneity in forecasting future asset prices is supported by evidence from survey data. For example, Vissing-Jorgensen (2003) reports that at the beginning of 2000, 50% of individual investors considered the stock market to be overvalued, approximately 25% believed that it was fairly valued, about 15% were unsure, and less than 10% believed that it was undervalued. This is an indication of heterogeneous beliefs among individual investors about the prospect of the stock market. Similarly, Shiller (2000) finds evidence that investors' sentiment varies over time. Both institutional and individual investors become more optimistic in response to significant increases in the recent performance of the stock market.

This chapter is organized as follows. Section 4.2 introduces the main features of adaptive belief systems, and Section 4.3 discusses a number of simple examples with two, three, and four different trader types. In Section 4.4 an analytical framework with *many* different trader types is presented. Section 4.5 discusses the empirical relevance of behavioral heterogeneity. The estimation of a simple model with fundamentalists and chartists on yearly S&P 500 data shows how the worldwide stock market bubble in the late 1990s, triggered by good news about fundamentals (a new Internet technology), may have been strongly amplified by trend-following strategies. Section 4.6 reviews some *learning to forecast* laboratory experiments with human subjects, investigating which individual forecasting rules agents may use, how these rules interact, and which aggregate outcome they cocreate. Section 4.7 concludes the chapter, sketching some challenges for future research and potential applications for financial practitioners and policy makers. An appendix contains a short mathematical overview of bifurcation theory, which plays a role in the transition to complicated price fluctuations in the simple complexity models discussed in this chapter.

4.2. AN ASSET-PRICING MODEL WITH HETEROGENEOUS BELIEFS

This section discusses the asset-pricing model with heterogeneous beliefs as introduced in Brock and Hommes (1998), using evolutionary selection of expectations as in Brock and Hommes (1997a). This simple modeling framework has been inspired by computational work at the Santa Fe Institute (SFI) and may be viewed as a simple, partly analytically tractable version of the more complicated SFI artificial stock market of Arthur et al. (1997b).

Agents can invest in either a risk-free asset or a risky asset. The risk-free asset is in perfect elastic supply and pays a fixed rate of return r; the risky asset pays an uncertain dividend. Let p_t be the price per share (ex-dividend) of the risky asset at time t, and let y_t be the stochastic dividend process of the risky asset. Wealth dynamics is given by

$$W_{t+1} = RW_t + (p_{t+1} + y_{t+1} - Rp_t)z_t \quad \textbf{(4.1)}$$

where $R = 1 + r$ is the gross rate of risk-free return and z_t denotes the number of shares of the risky asset purchased at date t. Let E_{ht} and V_{ht} denote the "beliefs" or forecasts of trader type h about conditional expectation and conditional variance.

Agents are assumed to be myopic mean-variance maximizers so that the demand z_{ht} of type h for the risky asset solves

$$\mathrm{Max}_{z_t}\{E_{ht}[W_{t+1}] - \frac{a}{2}V_{ht}[W_{t+1}]\} \quad \textbf{(4.2)}$$

where a is the risk-aversion parameter. The demand z_{ht} for risky assets by trader type h is then

$$z_{ht} = \frac{E_{ht}[p_{t+1} + y_{t+1} - Rp_t]}{aV_{ht}[p_{t+1} + y_{t+1} - Rp_t]} = \frac{E_{ht}[p_{t+1} + y_{t+1} - Rp_t]}{a\sigma^2} \quad \textbf{(4.3)}$$

where the conditional variance $V_{ht} = \sigma^2$ is assumed to be constant and equal for all types.[2] Let z^s denote the supply of outside risky shares per investor, also assumed to be constant, and let n_{ht} denote the fraction of type h at date t. Equilibrium of demand and supply yields

$$\sum_{h=1}^{H} n_{ht} \frac{E_{ht}[p_{t+1} + y_{t+1} - Rp_t]}{a\sigma^2} = z^s \tag{4.4}$$

where H is the number of different trader types.

The forecasts $E_{ht}[p_{t+1} + y_{t+1}]$ of tomorrow's prices and dividends are made *before* the equilibrium price p_t has been revealed by the market and therefore will depend on a publically available information set $I_{t-1} = \{p_{t-1}, p_{t-2}, \ldots; y_{t-1}, y_{t-2}, \ldots\}$ of past prices and dividends. Solving the heterogeneous market-clearing equation for the equilibrium price gives

$$Rp_t = \sum_{h=1}^{H} n_{ht} E_{ht}[p_{t+1} + y_{t+1}] - a\sigma^2 z^s \tag{4.5}$$

The quantity $a\sigma^2 z^s$ may be interpreted as a *risk premium* for traders to hold risky assets.

4.2.1. The Fundamental Benchmark with Rational Agents

When all agents are identical and expectations are *homogeneous*, the equilibrium pricing Eq. 4.5 reduces to

$$Rp_t = E_t[p_{t+1} + y_{t+1}] - a\sigma^2 z^s \tag{4.6}$$

where E_t is the common conditional expectation in the beginning of period t. It is well known that, assuming that a *transversality condition* $\lim_{t\to\infty}(E_t[p_{t+k}])/R^k = 0$ holds, the price of the risky asset is given by the discounted sum of expected future dividends minus the risk premium:

$$p_t^* = \sum_{k=1}^{\infty} \frac{E_t[y_{t+k}] - a\sigma^2 z^s}{R^k} \tag{4.7}$$

The price p_t^* in Eq. 4.7 is called the *fundamental rational expectations price*, or the *fundamental price* for short. It is completely determined by economic fundamentals, which are here given by the stochastic dividend process y_t. In this section we focus on the case of an independently identically distributed (IID) dividend process y_t, but the estimation of the simple two-type model discussed in Section 4.5 uses a nonstationary dividend process.[3] For the special case of an IID dividend process y_t, with constant

[2]Gaunersdorfer (2000) investigates the case with time-varying beliefs about variances and shows that the asset price dynamics are quite similar. Chiarella and He (2002, 2003) investigate the model with heterogeneous risk-aversion coefficients.

[3]Brock and Hommes (1997b) also discuss a nonstationary example, in which the dividend process follows a geometric random walk.

mean $E[y_t] = \bar{y}$, the fundamental price is constant:

$$p^* = \sum_{k=1}^{\infty} \frac{\bar{y} - a\sigma^2 z^s}{R^k} = \frac{\bar{y} - a\sigma^2 z^s}{r} \quad \textbf{(4.8)}$$

Recall that, in addition to the rational expectations fundamental solution in Eq. 4.7, so-called *rational bubble solutions* of the form $p_t = p_t^* + (1+r)^t(p_0 - p_0^*)$ also satisfy the pricing Eq. 4.6. Along these bubble solutions, traders have rational expectations (perfect foresight), but they are ruled out by the transversality condition. In a perfectly rational world, traders realize that such bubbles cannot last forever, and therefore all traders believe that the value of a risky asset is always equal to its fundamental price. Changes in asset prices are then only driven by unexpected changes in dividends and random "news" about economic fundamentals. In a heterogeneous world the situation will be quite different.

4.2.2. Heterogeneous Beliefs

It will be convenient to work with the *deviation* from the fundamental price

$$x_t = p_t - p_t^* \quad \textbf{(4.9)}$$

We make the following assumptions about the beliefs of trader type h:

B1 $V_{ht}[p_{t+1} + y_{t+1} - Rp_t] = V_t[p_{t+1} + y_{t+1} - Rp_t] = \sigma^2$, for all h, t.

B2 $E_{ht}[y_{t+1}] = E_t[y_{t+1}] = \bar{y}$, for all h, t.

B3 All beliefs $E_{ht}[p_{t+1}]$ are of the form

$$E_{ht}[p_{t+1}] = E_t[p_{t+1}^*] + E_{ht}[x_{t+1}] = p^* + f_h(x_{t-1}, \ldots, x_{t-L}) \quad \textbf{(4.10)}$$

for all h, t.

According to B1, beliefs about conditional variance are equal and constant for all types, as discussed already. Assumption B2 states that all types have correct expectations about future dividends y_{t+1} given by the conditional expectation, which is $\bar{y}$ in the case of IID dividends. According to B3, beliefs about future prices consist of two parts: a common belief about the fundamental plus a heterogeneous part f_{ht}.[4] Each forecasting rule f_h represents a *model of the market* (e.g., a technical trading rule) according to which type h believes that prices will deviate from the fundamental price.

An important and convenient consequence of the assumptions B1–B3 about traders' beliefs is that the heterogeneous agent market equilibrium Eq. 4.5 can be reformulated in *deviations* from the benchmark fundamental. In particular, substituting the price forecast

[4]The assumption that all types know the fundamental price is without loss of generality, because any forecasting rule *not* using the fundamental price can be reparameterized or reformulated for mathematical convenience in deviations from an (unknown) fundamental price p^*.

(Eq. 4.10) in the market equilibrium Eq. 4.5 and using $Rp_t^* = E_t[p_{t+1}^* + y_{t+1}] - a\sigma^2 z^s$ yields the equilibrium equation in deviations from the fundamental:

$$Rx_t = \sum_{h=1}^{H} n_{ht} E_{ht}[x_{t+1}] \equiv \sum_{h=1}^{H} n_{ht} f_{ht} \tag{4.11}$$

with $f_{ht} = f_h(x_{t-1}, \ldots, x_{t-L})$. Note that the benchmark fundamental is nested as a special case within this general setup, with all forecasting strategies $f_h \equiv 0$. Hence, the adaptive belief systems can be used in empirical and experimental testing where asset prices deviate significantly from some benchmark fundamental.

4.2.3. Evolutionary Dynamics

The evolutionary part of the model describes how beliefs are updated over time, that is, how the fractions n_{ht} of trader types evolve over time. These fractions are updated according to an *evolutionary fitness* or *performance* measure. The fitness measures of all trading strategies are publically available but subject to noise. Fitness is derived from a random utility model and given by

$$\tilde{U}_{ht} = U_{ht} + \varepsilon_{iht} \tag{4.12}$$

where U_{ht} is the *deterministic part* of the fitness measure and ε_{iht} represents an individual agent's IID error when perceiving the fitness of strategy $h = 1, \ldots H$.

To obtain analytical expressions for the probabilities or fractions, the noise term ε_{iht} is assumed to be drawn from a double exponential distribution. As the number of agents goes to infinity, the probability that an agent chooses strategy h is given by the *multinomial logit* model (or "Gibbs" probabilities):[5]

$$n_{ht} = \frac{e^{\beta U_{h,t-1}}}{\sum_{h=1}^{H} e^{\beta U_{h,t-1}}} \tag{4.13}$$

Note that the fractions n_{ht} add up to 1.

A key feature of Eq. 4.13 is that the higher the fitness of trading strategy h, the more traders will select strategy h. Hence, Eq. 4.13 represents a form of *reinforcement learning*: Agents tend to switch to strategies that have performed well in the (recent) past. The parameter β in Eq. 4.13 is called the *intensity of choice*; it measures the sensitivity of the mass of traders to selecting the optimal prediction strategy. The intensity of choice β is inversely related to the variance of the noise terms ε_{iht}. The extreme case $\beta = 0$ corresponds to noise of infinite variance so that differences in fitness cannot be observed and all fractions (4.13) will be fixed over time and equal to $1/H$.

The other extreme case $\beta = +\infty$ corresponds to the case without noise so that the deterministic part of the fitness can be observed perfectly, and in each period, *all* traders

[5]See Manski and McFadden (1981) and Anderson, de Palma, and Thisse (1993) for extensive discussion of discrete choice models and their applications in economics.

choose the optimal forecast. An increase in the intensity of choice β represents an increase in the degree of rationality with respect to evolutionary selection of trading strategies. The timing of the coupling between the market equilibrium Eq. 4.5 or 4.11 and the evolutionary selection of strategies (Eq. 4.13) is important. The market equilibrium price p_t in Eq. 4.5 depends on the fractions n_{ht}. The notation in Eq. 4.13 stresses the fact that these fractions n_{ht} depend on most recently observed *past* fitnesses $U_{h,t-1}$, which in turn depend on past prices p_{t-1} and dividends y_{t-1} in periods $t-1$ further in the past, as shown next. After the equilibrium price p_t has been revealed by the market, it will be used in evolutionary updating of beliefs and determining the new fractions $n_{h,t+1}$. These new fractions will then determine a new equilibrium price p_{t+1}, and so on. In an adaptive belief system, market equilibrium prices and fractions of different trading strategies thus coevolve over time.

A natural candidate for evolutionary fitness is (a weighted average of) *realized profits*, given by[6]

$$U_{ht} = (p_t + y_t - Rp_{t-1})\frac{E_{h,t-1}[p_t + y_t - Rp_{t-1}]}{a\sigma^2} + wU_{h,t-1} \tag{4.14}$$

where $0 \leq w \leq 1$ is a *memory* parameter measuring how fast past realized fitness is discounted for strategy selection.

Fitness can be rewritten in terms of deviations from the fundamental as

$$U_{ht} = (x_t - Rx_{t-1} + a\sigma^2 z^s + \delta_t)\left(\frac{f_{h,t-1} - Rx_{t-1} + a\sigma^2 z^s}{a\sigma^2}\right) + wU_{h,t-1} \tag{4.15}$$

where $\delta_t \equiv p_t^* + y_t - E_{t-1}[p_t^* + y_t]$ is a martingale difference sequence.

4.2.4. Forecasting Rules

To complete the model, we have to specify the class of forecasting rules. Brock and Hommes (1998) have investigated evolutionary competition between *simple linear* forecasting rules with only *one lag*, that is,

$$f_{ht} = g_h x_{t-1} + b_h \tag{4.16}$$

It can be argued that, for a forecasting rule to have any impact in real markets, it has to be simple, because it seems unlikely that enough traders will coordinate on a

[6]Note that this fitness measure does *not* take into account the risk taken at the moment of the investment decision. In fact, one could argue that the fitness measure (Eq. 4.14) does not take into account the variance term in (Eq. 4.2) capturing the investors' risk taken before obtaining that profit. On the other hand, in real markets, realized net profits or accumulated wealth may be what investors care about most, and the non-risk adjusted fitness measure (Eq. 4.14) may thus be of relevance in practice. See also DeLong et al. (1990) for a discussion of this point. Given that investors are risk averse, mean-variance maximizers maximizing their expected utility from wealth (Eq. 4.2), an alternative natural candidate for fitness, the *risk-adjusted profit* given by $\pi_{ht} = R_t z_{h,t-1} - \frac{a}{2}\sigma^2 z_{h,t-1}^2$, where $R_t = p_t + y_t - Rp_{t-1}$ and $z_{h,t-1} = E_{h,t-1}[R_t]/(a\sigma^2)$ is the demand by trader type h. Hommes (2001) shows that the risk-adjusted fitness measure is, up to a type-independent level, equivalent to minus-squared prediction errors.

complicated rule. The simple linear rule (Eq. 4.16) includes a number of important special cases. For example, when both the trend and the bias parameters $g_h = b_h = 0$, the rule reduces to the *fundamentalists* forecast, that is,

$$f_{ht} \equiv 0 \tag{4.17}$$

predicting that the deviation x from the fundamental will be 0, or equivalently, that the price will be at its fundamental value. Other important cases covered by the linear forecasting rule (Eq. 4.16) are the pure *trend followers*:

$$f_{ht} = g_h x_{t-1} \quad g_h > 0 \tag{4.18}$$

and the pure *biased belief*

$$f_{ht} = b_h \tag{4.19}$$

Notice that the simple pure bias forecast (Eq. 4.19) represents *any* positively or negatively biased forecast of next period's price that traders might have. Instead of these extremely simple habitual rule-of-thumb forecasting rules, some might prefer the rational, *perfect foresight* forecasting rule:

$$f_{ht} = x_{t+1} \tag{4.20}$$

We emphasize, however, that the perfect foresight forecasting rule (Eq. 4.20) assumes perfect knowledge of the heterogeneous market equilibrium Eq. 4.5, and in particular perfect knowledge about the beliefs of *all* other traders. Although the case with perfect foresight has much theoretical appeal, its practical relevance in a complex heterogeneous world should not be overstated, since this underlying assumption seems rather strong.[7]

4.3. SIMPLE EXAMPLES

This section presents simple but typical examples of *adaptive belief systems* (ABSs), with two, three, or four competing *linear* forecasting rules (Eq. 4.16), where the parameter g_h represents a perceived *trend* in prices and the parameter b_h represents a perceived upward or downward *bias*. The ABS with H types is given by (in deviations

[7]Brock and Hommes (1997) analyze the cobweb model with costly rational versus cheap naïve expectations and find irregular price fluctuations due to endogenous switching between free riding and costly rational forecasting. In general, however, a *temporary equilibrium* model with heterogeneous beliefs, such as the asset-pricing model, is difficult to analyze if one of the types has perfect foresight. Brock et al. (2008) discuss how a perfect foresight trader may affect the dynamics in an asset-pricing model with heterogeneous beliefs.

from the fundamental benchmark):

$$(1+r)x_t = \sum_{h=1}^{H} n_{ht}(g_h x_{t-1} + b_h) + \epsilon_t \tag{4.21}$$

$$n_{h,t} = \frac{e^{\beta U_{h,t-1}}}{\sum_{h=1}^{H} e^{\beta U_{h,t-1}}} \tag{4.22}$$

$$U_{h,t-1} = (x_{t-1} - Rx_{t-2})\left(\frac{g_h x_{t-3} + b_h - Rx_{t-2}}{a\sigma^2}\right) + wU_{h,t-2} - C_h \tag{4.23}$$

where ϵ_t is a small noise term representing, for example, a small fraction of noise traders and/or random outside supply of the risky asset.

To keep the analysis of the dynamical behavior tractable, Brock and Hommes (1998) focused on the case where the memory parameter $w = 0$, so that evolutionary fitness is given by last period's realized profit. A common feature of all examples is that, as the intensity of choice to switch prediction or trading strategies increases, the fundamental steady state becomes locally unstable and nonfundamental steady states, cycles, or even chaos arise. In the examples that follow, we encounter different bifurcation routes (i.e., transitions) to complicated dynamics. A mathematical appendix summarizes the most important *bifurcations*, that is, qualitative changes in the dynamics (e.g., when a steady state loses stability or a new cycle is created) when a model parameter changes.

4.3.1. Costly Fundamentalists vs. Trend Followers

The simplest example of an ABS only has *two* trader types, with forecasting rules

$$f_{1t} = 0 \qquad \text{fundamentalists} \tag{4.24}$$

$$f_{2t} = gx_{t-1} \quad g > 0 \qquad \text{trend followers} \tag{4.25}$$

The first type are fundamentalists predicting that the price will equal its fundamental value (or equivalently that the deviation will be zero) and the second type are pure trend followers predicting that prices will rise (or fall) by a constant rate. In this example, the fundamentalists have to pay a fixed per-period positive cost C_1 for information gathering; in all other examples discussed later, information costs are set to zero for all trader types.

For small values of the trend parameter $0 \le g < 1 + r$, the fundamental steady state is always stable. Only for sufficiently high trend parameters, $g > 1 + r$, trend followers can *destabilize* the system. For trend parameters $1 + r < g < (1 + r)^2$, the dynamic behavior of the evolutionary system depends on the intensity of choice to switch between the two trading strategies.[8] For low values of the intensity of choice,

[8]For $g > (1 + r)^2$ the system may become *globally unstable* and prices may diverge to infinity. Imposing a stabilizing force—for example, by assuming that trend followers condition their rule on deviations from the fundamental, as in Gaunersdorfer, Hommes, and Wagener (2008)—leads to a bounded system again, possibly with cycles or even chaotic fluctuations.

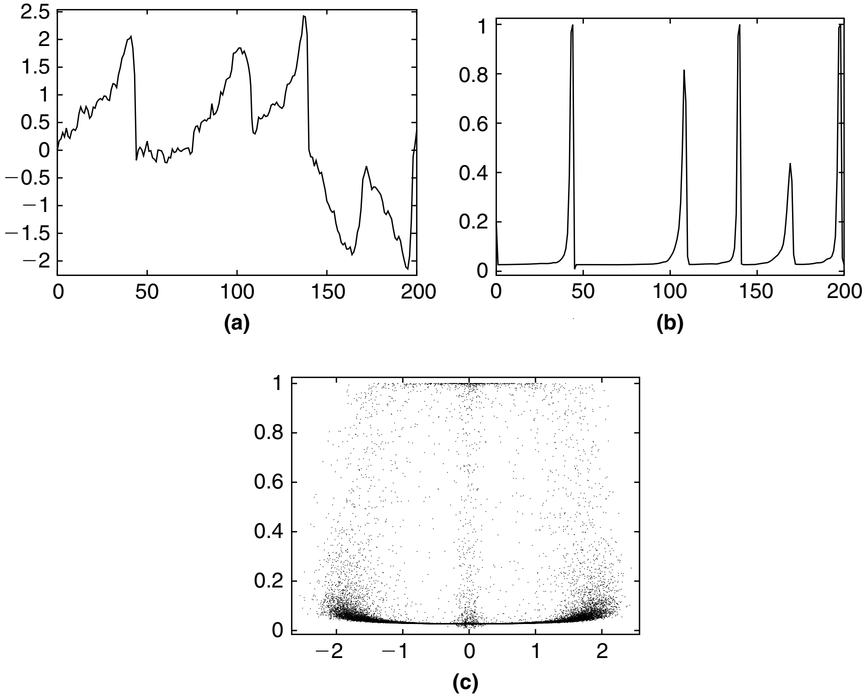

FIGURE 4.1 Time series of price deviations from fundamental (a) and fractions of fundamentalists (b) and attractor (c) in the (x_t, n_{1t})-phase space for two-type model with costly fundamentalist versus trend followers buffeted with small noise (SD = 0.1). Price dynamics are characterized by temporary bubbles when trend followers dominate the market, interrupted by sudden crashes, when fundamentalists dominate. In the presence of (small) noise, the system switches back and forth between two coexisting quasiperiodic attractors of the underlying deterministic skeleton, one with prices above and one with prices below its fundamental value. Parameters are $\beta = 3.6$, $g = 1.2$, $R = 1.1$, and $C = 1$.

the fundamental steady state will be stable. As the intensity of choice increases, the fundamental steady state becomes unstable due to a *pitchfork bifurcation* in which two additional nonfundamental steady states $-x^* < 0 < x^*$ are created.

As the intensity of choice increases further, the two nonfundamental steady states also become unstable due to a Hopf bifurcation, and limit cycles or even strange attractors can arise around each of the (unstable) nonfundamental steady states.[9] The evolutionary ABS may cycle around the positive nonfundamental steady state, cycle around the negative nonfundamental steady state, or, driven by the noise, switch back and forth between cycles around the high and the low steady states, as illustrated in Figure 4.1. The simulations use the E&F Chaos software as discussed in Diks et al. (2008).

This example shows that, in the presence of information costs and with zero memory, when the intensity of choice in evolutionary switching is high, fundamentalists *cannot*

[9]See Appendix 4.2 for a more detailed discussion of the pitchfork bifurcation and the Hopf bifurcation.

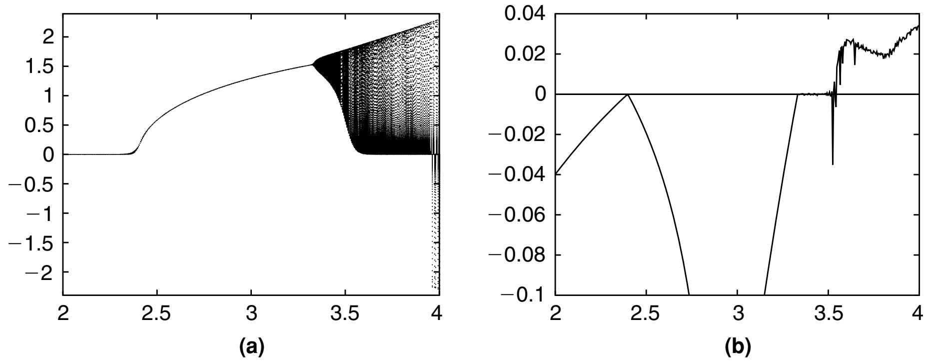

FIGURE 4.2 Bifurcation diagram (a) and largest Lyapunov exponent plot (b) as a function of the intensity of choice β for two-type model with costly fundamentalist versus trend followers. In both plots the model is buffeted with very small noise ($SD = 10^{-6}$) for the noise term ϵ_t in Eq. 4.21; to avoid that for large β values the system gets stuck in the locally unstable steady state. Parameters are $g = 1.2$, $R = 1.1$, $C = 1$, and $2 \leq \beta \leq 4$. A pitchfork bifurcation of the fundamental steady state, in which two stable nonfundamental steady states are created, occurs for $\beta \approx 2.37$. The nonfundamental steady states become unstable due to a Hopf bifurcation for $\beta \approx 3.33$, and (quasi-)periodic dynamics arise. For large values of β the largest Lyapunov exponent becomes positive, indicating chaotic price dynamics.

drive out pure trend followers and persistent deviations from the fundamental price may occur.[10]

Figure 4.2 illustrates that the asset-pricing model with costly fundamentalists versus cheap trend-following exhibits a *rational route to randomness*, that is, a bifurcation route to chaos occurs as the intensity of choice to switch strategies increases.

4.3.2. Fundamentalists vs. Opposite Biases

The second example of an ABS is an example with *three* trader types without any information costs. The forecasting rules are

$$f_{1t} = 0 \qquad \text{fundamentalists} \qquad \mathbf{(4.26)}$$

$$f_{2t} = b \qquad b > 0, \qquad \text{positive bias (optimists)} \qquad \mathbf{(4.27)}$$

$$f_{3t} = -b \qquad -b < 0 \qquad \text{negative bias (pessimists)} \qquad \mathbf{(4.28)}$$

The first type are fundamentalists as before, but there are *no* information costs for fundamentalists. The second and third types have a purely *biased* belief, expecting a constant price above and below, respectively, the fundamental price.

For low values of the intensity of choice, the fundamental steady state is stable. As the intensity of choice increases the fundamental steady becomes unstable due to a Hopf bifurcation and the dynamics of the ABS is characterized by cycles around the

[10]Brock and Hommes (1999) show that this result also holds when the memory in the fitness measure increases. In fact, an increase in the memory of the evolutionary fitness leads to bifurcation routes very similar to bifurcation routes that are due to an increase in the intensity of choice.

unstable steady state. This example shows that, even when there are *no* information costs for fundamentalists, they cannot drive out other trader types with opposite biased beliefs. In the evolutionary ABS with high intensity of choice, fundamentalists and biased traders coexist with fractions varying over time and prices fluctuating around the unstable fundamental steady state.

Moreover, Brock and Hommes (1998, p. 1259, Lemma 9) show that as the intensity of choice tends to infinity the ABS converges to a (globally) stable cycle of Period 4. Average profits along this four-cycle are equal for all three trader types. Hence, if the initial wealth is equal for all three types, in this evolutionary system in the long run, accumulated wealth will be equal for all three types. This example shows that the Friedman argument that smart fundamental traders will always drive out simple rule-of-thumb speculative traders is in general not valid.[11]

4.3.3. Fundamentalists vs. Trend and Bias

The third example of an ABS is an example with *four* trader types, with linear forecasting rules (Eq. 4.16) with parameters $g_1 = 0$, $b_1 = 0$; $g_2 = 0.9$, $b_2 = 0.2$; $g_3 = 0.9$, $b_3 = -0.2$; and $g_4 = 1 + r = 1.01$, $b_4 = 0$. The first type are fundamentalists again, without information costs, and the other three types follow a simple linear forecasting rule with one lag. The dynamical behavior is illustrated in Figures 4.3 and 4.4.

For low values of the intensity of choice, the fundamental steady state is stable. As the intensity of choice increases, as in the previous three-type example, the fundamental steady becomes unstable due to a Hopf bifurcation and a stable invariant circle around the unstable fundamental steady state arises, with periodic or quasi-periodic fluctuations. As the intensity of choice further increases, the invariant circle breaks into a strange attractor with chaotic fluctuations. In the evolutionary ABS, fundamentalists and chartists coexist with time-varying fractions and prices moving chaotically around the unstable fundamental steady state. Figure 4.4 shows that in this four-type example with fundamentalists versus trend followers and biased beliefs a rational route to randomness occurs, with positive largest Lyapunov exponents for large values of β.

This four-type example shows that, even when there are *no* information costs for fundamentalists, they cannot drive out other simple trader types and fail to stabilize price fluctuations toward its fundamental value. As in the three-type case, the opposite biases create cyclic behavior and trend extrapolation turns these cycles into unpredictable chaotic fluctuations.

4.3.4. Efficiency

What can be said about market efficiency in an ABS? The (noisy) chaotic price fluctuations are characterized by an irregular switching between phases of close-to-the-fundamental-price fluctuations, phases of "optimism" with prices following an upward

[11]This result is related to DeLong et al. (1990a,b), who show that a constant fraction of noise traders can survive in the market in the presence of fully rational traders. The ABS, however, are evolutionary models with *time-varying* fractions, driven by strategy performance.

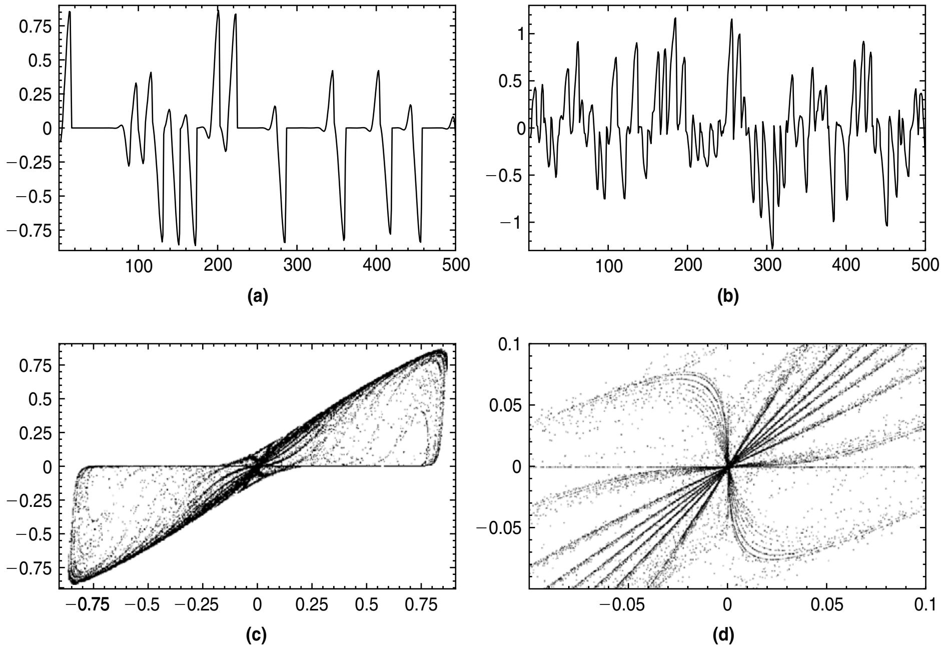

FIGURE 4.3 Chaotic (a) and noisy chaotic (b) time series of asset prices in an adaptive belief system with four trader types. Strange attractor (c) and enlargement of strange attractor (d). Belief parameters are $g_1 = 0$, $b_1 = 0$; $g_2 = 0.9$, $b_2 = 0.2$; $g_3 = 0.9$, $b_3 = -0.2$; and $g_4 = 1 + r = 1.01$, $b_4 = 0$. Other parameters are $r = 0.01$, $\beta = 90.5$, $w = 0$, and $C_h = 0$ for all $1 \leq h \leq 4$.

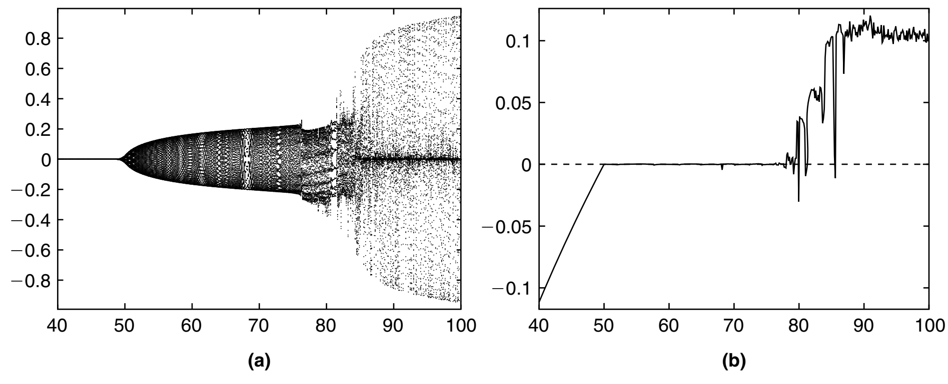

FIGURE 4.4 Bifurcation diagram (a) and largest Lyapunov exponent plot (b) for the four-type model, buffeted with very small noise ($SD = 10^{-6}$ for noise term ϵ_t in Eq. 4.21) to avoid that for large β-values the system gets stuck in the locally unstable steady state. Belief parameters are $g_1 = 0$, $b_1 = 0$; $g_2 = 0.9$, $b_2 = 0.2$; $g_3 = 0.9$, $b_3 = -0.2$; and $g_4 = 1 + r = 1.01$, $b_4 = 0$. Other parameters are $r = 0.01$, $\beta = 90.5$, $w = 0$, and $C_h = 0$ for all $1 \leq h \leq 4$. The four-type model with fundamentalists versus trend followers and biased beliefs exhibits a Hopf bifurcation at $\beta = 50$. A rational route to randomness (i.e., a bifurcation route to chaos) occurs, with positive largest Lyapunov exponents, when the intensity of choice becomes large.

trend, and phases of "pessimism," with (small) sudden market crashes, as illustrated in Figure 4.3. In fact, in the ABS, prices are characterized by evolutionary switching between the fundamental value and temporary speculative bubbles. Prices deviate persistently from their fundamental value; therefore it can be said that prices are excessively volatile. In this sense the market is inefficient. But are these deviations easy to predict? Even in the simple, stylized four-type example in the purely deterministic chaotic case, the timing and direction of the temporary bubbles seem hard to predict, but once a bubble has started, the collapse of the bubble seems predictable. In the presence of (small) noise, however, the situation is quite different, as illustrated in Figure 4.3 (top right): The timing, the direction and the collapse of the bubble all seem hard to predict.

To stress this point further, we investigate this (un)predictability by employing a so-called *nearest neighbor forecasting method* to predict the returns, at lags 1 to 20 for the purely chaotic as well as for several noisy chaotic time series, as illustrated in Figure 4.5.[12] Nearest-neighbor forecasting looks for patterns in the past that are close to the most recent pattern and then predicts the average value following all nearby past patterns. According to Takens' embedding theorem, this method yields good forecasts for deterministic chaotic systems.[13] Figure 4.5 shows that as the noise level increases, the forecasting performance of the nearest-neighbor method quickly deteriorates. Therefore, in our simple nonlinear evolutionary ABS with noise, it is hard to make good forecasts of future returns and to predict when prices will return to fundamental value. Our simple nonlinear ABS with small noise thus captures some of the intrinsic unpredictability of asset returns also present in real markets, and in terms of predictability the market is close to being efficient.

4.3.5. Wealth Accumulation

The evolutionary dynamics in an ABS are driven by realized *short-run profits*, and chartists strategies survive in a world driven by short-run profit opportunities. In this subsection, we briefly look at the accumulated wealth in an ABS. Recall that accumulated wealth for strategy type h is given by

$$W_{h,t+1} = RW_{ht} + (p_{t+1} + y_{t+1} - Rp_t)z_{ht} \tag{4.29}$$

The first term represents wealth growth due to the risk-free asset, while the last term represents wealth growth (or decay) due to investments in the risky asset. Because of market clearing, the average net inflow of wealth due to investment in the risky asset is given by

$$\sum_h n_{ht} z_{ht}(p_{t+1} + y_{t+1} - Rp_t) = z^s(p_{t+1} + y_{t+1} - Rp_t) \tag{4.30}$$

[12] We would like to thank Sebastiano Manzan for providing this figure.

[13] See Kantz and Schreiber (1997) for an extensive treatment of nonlinear time-series analysis and forecasting techniques.

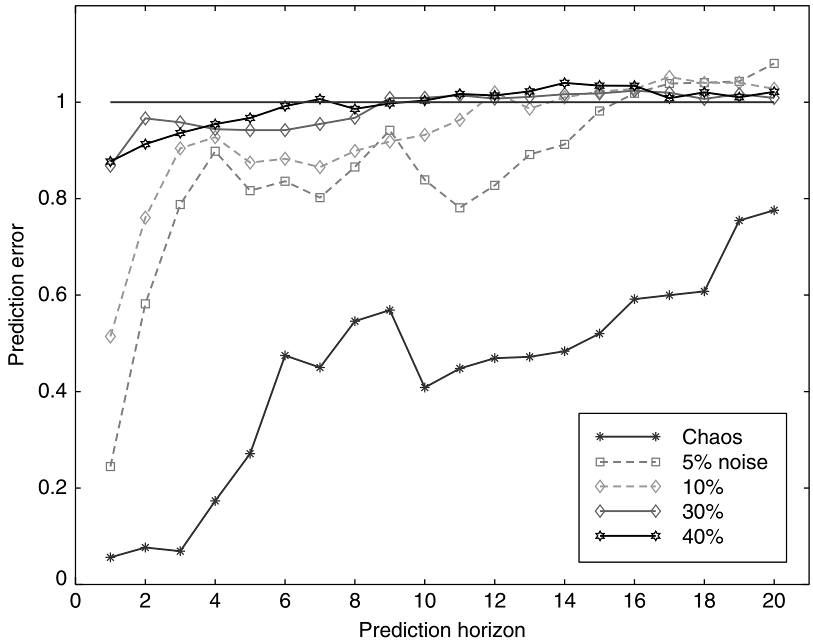

FIGURE 4.5 Forecasting errors for nearest-neighbor method applied to chaotic and noisy chaotic returns series for different noise levels in the four-type adaptive belief system. All returns series have close to 0 autocorrelations at all lags. The benchmark case of prediction by the mean 0 is represented by the horizontal line at the normalized prediction error 1. Nearest-neighbor forecasting applied to the purely deterministic chaotic series leads to much smaller forecasting errors at all prediction horizons, 1–20 (bottom graph). A noise level of, say, 10% means that the ratio of the variance of the noise ϵ_t and the variance of the deterministic price series is $1/10$. As the noise level increases, the graphs shift upward, indicating that prediction errors increase. Small dynamic noise thus quickly deteriorates forecasting performance.

This is the average risk premium required by the population of investors to hold the risky asset. In the special case $z^s = 0$ the risk premium is 0 and on average wealth of each strategy grows at the risk-free rate.

Figure 4.6 shows the development of prices and wealth of each strategy in the three-type and four-type examples of Subsections 4.3.2 and 4.3.3. Prices fluctuate around the fundamental price. For the three-type example, the wealth accumulated by each of the three strategies—fundamentalists, optimistic biased, and pessimistic biased—grows over time, at an equal rate. Recall that in the three-type example, for an infinite intensity of choice β, the system converges to a stable four-cycle with average profits equal for all three strategies. At each time t, profits of fundamentalists are always between profits of optimists and pessimists, but on average all profits are (almost) equal, and thus accumulated wealth grows at the same rate.[14]

[14]For finite intensity of choice, for example, $\beta = 3000$ as in Figure 4.6, wealth of the three-types grows at almost the same rate. For initial states chosen, as in Figure 4.6, wealth of the optimistic types slightly dominates the other two types.

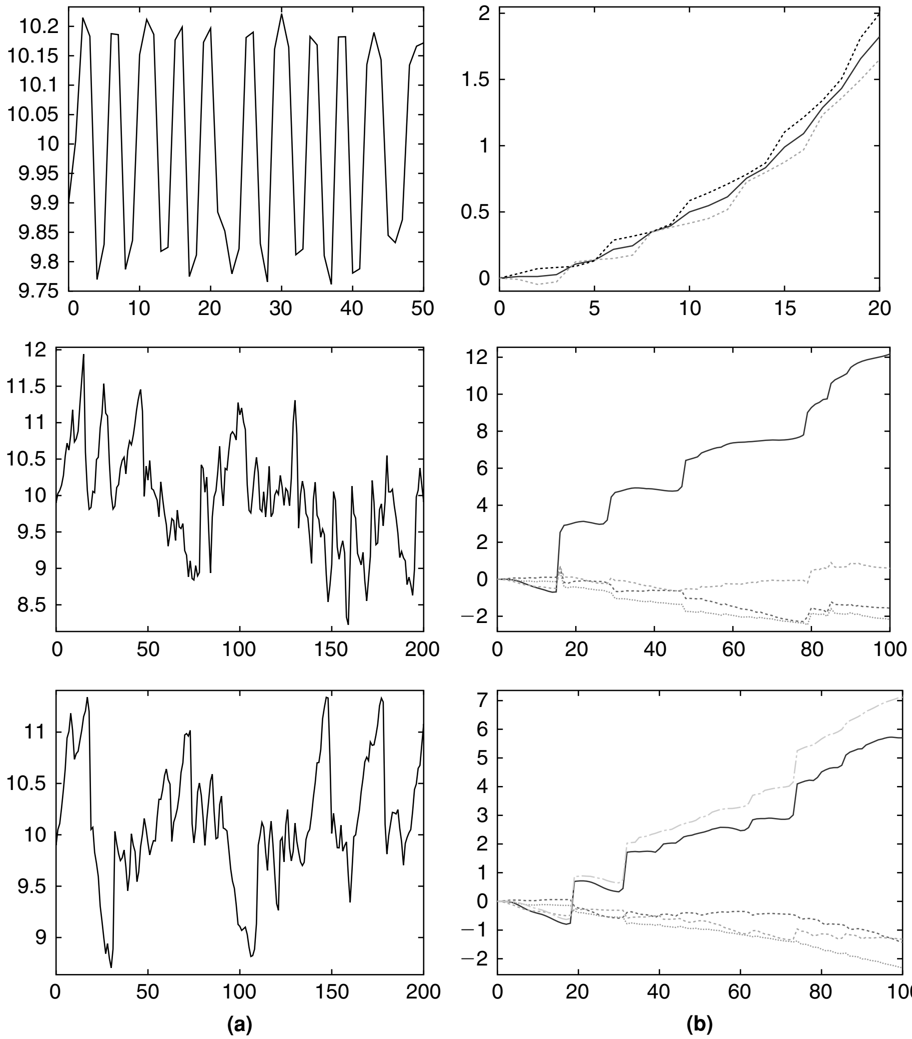

FIGURE 4.6 Time series of prices (a) and accumulated wealth (b) in three-type ABS (*top*), four-type ABS (*middle*), and five-type ABS (*bottom*). Belief parameters are $b = 0.2$ for the three-type ABS (see subsection 4.3.2) and as in Figure 4.3 in the four-type case. Other parameters are $\beta = 3000$ and $\sigma_\epsilon = 0.025$ (three-type ABS); $g_4 = (1 + r)^2 = 1.0201$, $\beta = 180$, $\bar{y} = 0.1$, $R = 1.01$, $\sigma_\epsilon = 0.2$ (four-type ABS), and $\sigma_\epsilon = 0.1$; and threshold parameter $\vartheta = 0.5$ for switching strategy (five-type ABS).

In the four-type example, trend-following strategies are profitable during temporary bubbles. Fundamentalists suffer losses during temporary bubbles, but these losses are limited. When the bubble bursts, fundamentalists make large profits while trend followers suffer from huge losses. On average, accumulated wealth of fundamentalists increases while wealth of chartists decreases, as illustrated in Figure 4.6.

The wealth in Eq. 4.29 corresponds to the accumulated wealth of a trader who always uses strategy h. How would a switching strategy perform in a heterogeneous market? Figure 4.6 (bottom panel) illustrates an example of an ABS with five strategies, where a switching strategy has been added to the four-type ABS. The fifth switching strategy is endogenous in the five-type ABS and thus affects the realized market price in the same way as the other four strategies. The switching strategy always picks the best of the other four strategies, according to last period's realized profits, conditional on how far the price deviates from the fundamental benchmark. In the simulation, when the price deviation becomes larger than a threshold parameter ($\vartheta = 0.5$), the switching strategy switches back to the fundamental strategy to avoid losses when the bubble collapses.

Figure 4.6 (bottom row) illustrates two features of the five-type ABS. First, due to the presence of the switching strategy, the amplitude of price fluctuations (bottom row, left plot) is somewhat smaller than in the four-type ABS. This is caused by the switching strategy switching back to the fundamental strategy when the price deviation exceeds the threshold. Second, the accumulated wealth of the switching strategy outperforms all other strategies, including the fundamental strategy—see Figure 4.6, bottom row, right plot. Notice that the two best strategies, the switching strategy and the fundamental strategy, also require the most information. The trend-following strategies only use publically available information on past prices.[15] The fundamental strategy uses fundamental information, whereas the switching strategy uses fundamental information as well as information about competing strategies in the market and their performance.

In the ABS evolutionary framework, agents switch strategies based on *short-run* realized profits. In the long run, a fundamental strategy often accumulates more wealth than trend-following rules. However, fundamental strategies suffer from losses during temporary bubbles when prices persistently deviate from fundamentals and may therefore suffer from "limits of arbitrage" (Shleifer and Vishny, 1997). Fundamentalists can stabilize price fluctuations but only if they are not limited by borrowing constraints or limits of arbitrage. In the long run, a simple switching strategy may accumulate more wealth than a fundamental or technical trading strategy. The fact that a simple switching strategy performs better in a heterogeneous market shows that the ABS model is *behaviorally consistent*. Agents have an incentive to keep switching strategies.

The switching strategy is very risky, however, because it requires good knowledge of the underlying fundamental and good market timing to "get off the bubble before it bursts." Interestingly, Zwart et al. (2007) provided empirical evidence, analyzing 15 emerging market currencies over the period from 1995 to 2006, that a combined strategy with time-varying weights may generate economically and statistically significant returns, after accounting for transaction costs. Their strategy is based on a combination of fundamental information on the deviation from purchasing power parity and the real interest rate differential and chartist information from moving average trading rules, with time-varying weights determined by relative performance over the past year.

[15]Recall that these strategies can be formulated without knowledge of the fundamental price; see Footnote 4.

4.3.6. Extensions

Several modifications and extensions of ABSs have been studied. In Brock and Hommes (1998) the demand for the risky asset is derived from a constant absolute risk aversion (CARA) utility function. Chiarella and He (2001) consider the case with constant relative risk aversion (CRRA) utility. This is complicated because under CARA utility, investors' relative wealth affects asset demand and realized asset price, and one has to keep track of the wealth distribution among the population of agents.[16] Anufriev and Bottazzi (2006) and Anufriev (2008) study wealth and asset price dynamics in a heterogeneous agents framework and are able to characterize the type of equilibria and their stability under fairly general behavioral assumptions. Chiarella, Dieci, and Gardini (2002, 2006) use the CRRA utility in an ABS with a market-maker price-setting rule. Chiarella and He (2003) and Hommes et al. (2005) investigate an ABS with a *market-maker* price-setting rule and find quite similar dynamical behavior as in the case of a Walrasian market-clearing price. De Fontnouvelle (2000) and Goldbaum (2005) apply strategy switching to an asset-pricing model with heterogeneous information. Chang (2007) studies how social interactions affect the dynamics of asset prices in an ABS with a Walrasian market-clearing price. DeGrauwe and Grimaldi (2005, 2006) applied the ABS framework to exchange rate modeling. Chiarella et al. (2009) discusses some of these extensions in more detail.[17]

4.4. MANY TRADER TYPES

In most heterogeneous agent models (HAMs) in the literature, the number of trader types is small—usually only two, three, or four types are considered that use simple fundamentalist or chartist strategies. Generally, analytical tractability can only be obtained at the cost of restricting a HAM to just a few types. However, Brock, Hommes, and Wagener (2005) have developed a theoretical framework to study evolutionary markets with *many* different trader types. In this subsection, we discuss their notion of *large type limit* (LTL), a simple, low-dimensional approximation of an evolutionary ABS with many trader types. The LTL can be developed in a fairly general market-clearing setting, but here we focus on its application to the asset-pricing model with heterogeneous beliefs.

Recall from Eq. 4.1 that in the asset market with H different trader types, the equilibrium price (in deviations x_t from the fundamental benchmark) is given by

$$x_t = \frac{1}{1+r}\sum_{h=1}^{H} n_{ht} f_{ht} \tag{4.31}$$

[16]In the artifical market of Levy et al. (1994), asset demand is also derived from CRRA utility.

[17]Another related stochastic model with heterogeneous agents and endogenous strategy switching similar to the ABS has been introduced in Föllmer et al. (2005). Scheinkman and Xiong (2004) review related stochastic financial models with heterogeneous beliefs and short-sale constraints. Macro models with heterogeneous expectations have been studied, for instance, in Branch and Evans (2006) and Branch and McGough (2008).

Using the *multinomial logit* probabilities (Eq. 4.13) for the fractions n_{ht}, we get

$$x_t = \frac{1}{1+r} \frac{\sum_{h=1}^{H} e^{\beta U_{h,t-1}} f_{ht}}{\sum_{h=1}^{H} e^{\beta U_{h,t-1}}} \tag{4.32}$$

It is assumed that prediction and fitness functions take the form $f_{ht} = f(\mathrm{x}, \lambda, \vartheta_h)$ and $U_{ht} = U(\mathrm{x}, \lambda, \vartheta_h)$ respectively, where $\mathrm{x} = (x_{t-1}, x_{t-2}, \ldots)$ is a vector of lagged deviations from the fundamental, λ is a structural parameter vector (e.g., containing the risk-free interest rate r, the risk-aversion parameter a, the intensity of choice β, etc.), and ϑ_h is a multidimensional variable that characterizes the belief type h.

The equilibrium Eq. 4.32 determines the evolution of the *system with H trader types*; this information is coded in the *evolution map* $\varphi_H(\mathrm{x}, \lambda, \vartheta)$:

$$\varphi_H(\mathrm{x}, \lambda, \vartheta) = \frac{1}{1+r} \frac{\sum_{h=1}^{H} e^{\beta U(\mathrm{x},\lambda,\vartheta_h)} f(\mathrm{x}, \lambda, \vartheta_h)}{\sum_{h=1}^{H} e^{\beta U(\mathrm{x},\lambda,\vartheta_h)}} \tag{4.33}$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_H)$. At the beginning of the market, a large number H of beliefs ϑ_h is sampled from a general distribution of initial beliefs. For example, all forecasting rules may be drawn from a linear class of rules with L lags,

$$f_t(\vartheta_0) = \vartheta_{00} + \vartheta_{01} x_{t-1} + \vartheta_{02} x_{t-2} + \cdots + \vartheta_{0L} x_{t-L} \tag{4.34}$$

with ϑ_{0h}, $h = 0, \ldots, L$, drawn from a multivariate normal distribution.

The evolution map φ_H in Eq. 4.33 determines the dynamical system corresponding to an *asset market with H different belief types*. When the number of trader types H is large, this dynamical system contains a large number of stochastic variables $\vartheta = (\vartheta_1, \ldots, \vartheta_H)$, where the ϑ_h are IID, with *distribution function F_μ*. At the beginning of the market, H belief types are drawn from this distribution, and they then compete against each other. The distribution function of the stochastic belief variable ϑ_h depends on a multidimensional parameter μ, called the *belief parameter*. This setup allows us to vary the population out of which the individual beliefs are sampled at the beginning of the market.

Observe that both the denominator and the numerator of the evolution map φ_H in Eq. 4.33 may be divided by the number of trader types H and thus may be seen as sample means. The evolution map ψ of the LTL is then obtained by *replacing sample means in the evolution map φ_H by population means*:

$$\begin{aligned}\psi(\mathrm{x}, \lambda, \mu) &= \frac{1}{1+r} \frac{E_\mu\left[e^{\beta U(\mathrm{x},\lambda,\vartheta_0)} f(\mathrm{x}, \lambda, \vartheta_0)\right]}{E_\mu\left[e^{\beta U(\mathrm{x},\lambda,\vartheta_0)}\right]} \\ &= \frac{1}{1+r} \frac{\int e^{\beta U(\mathrm{x},\lambda,\vartheta_0)} f(\mathrm{x}, \lambda, \vartheta_0) d\nu_\mu}{\int e^{\beta U(\mathrm{x},\lambda,\vartheta_0)} d\nu_\mu}\end{aligned} \tag{4.35}$$

where ϑ_0 is a stochastic variable, distributed in the same way as the ϑ_h, with density ν_μ. The *structural* parameter vector λ of the evolution map φ_H and of the LTL evolution map ψ coincide. However, whereas the evolution map φ_H in Eq. 4.33 of the heterogeneous agent system contains H randomly drawn multidimensional stochastic variables ϑ_h, the LTL evolution map ψ in Eq. 4.35 only contains the *belief parameter* vector μ describing the joint probability distribution. Taking an LTL thus leads to a huge reduction in stochastic belief variables.

According to the LTL theorem of Brock et al. (2005), as the number H of trader types tends to infinity, the H-type evolution map φ_H converges almost surely to the LTL map ψ. This implies that the corresponding LTL dynamical system is a good approximation of the dynamical behavior in a heterogeneous asset market when the number of belief types H is large. In particular, all *generic* and *persistent* dynamic properties will be preserved with high probability. For example, if the LTL map exhibits a bifurcation route to chaos for one of the structural parameters, then, if the number of trader types H is large, the H-type system also exhibits such a bifurcation route to chaos with high probability.

A straightforward computation using moment-generating functions shows that, for example, in the case of linear forecasting rules (Eq. 4.34) with three lags ($L = 3$), the corresponding LTL becomes a 5-D nonlinear system given by

$$\begin{aligned}(1+r)x_t &= \mu_0 + \mu_1 x_{t-1} + \mu_2 x_{t-2} + \mu_3 x_{t-3}\\ &\quad + \eta(x_{t-1} - Rx_{t-2} + a\sigma^2 z^s)\\ &\quad (\sigma_0^2 + \sigma_1^2 x_{t-1}x_{t-3} + \sigma_2^2 x_{t-2}x_{t-4} + \sigma_3^2 x_{t-3}x_{t-5})\end{aligned} \tag{4.36}$$

where $\eta = \beta/(a\sigma^2)$. The simplest special case of Eq. 4.36 that still leads to interesting dynamics is obtained when all $\vartheta_{0k} = 0$, $1 \le k \le d$, that is, when the forecasting function (Eq. 4.34) is purely biased: $f_t(\vartheta_0) = \vartheta_{00}$. The LTL then simplifies to the linear system

$$Rx_t = \mu_0 + \eta\sigma_0^2\left(x_{t-1} - Rx_{t-2} + a\sigma^2 z^s\right) \tag{4.37}$$

This simplest case already provides insight into the (in)stability of the (fundamental) steady state in an evolutionary system with many trader types. When there is no intrinsic mean bias (i.e., when the mean of the biases ϑ_{00} equals 0 (i.e., $\mu_0 = 0$) and the risk premium is zero ($z^s = 0$)), the steady state of the LTL (Eq. 4.37) coincides with the fundamental: $x^* = 0$. When the mean bias and risk premium are both positive (negative), the steady-state deviation x^* will be positive (negative) so that the steady state will be above (below) the fundamental. The natural bifurcation parameter tuning the (in)stability of the system is $\eta\sigma_0^2 = \beta\sigma_0^2/a\sigma^2$. We see that instability occurs if and only if η is increased beyond the bifurcation point $\eta_c = 1/\sigma_0^2$. Therefore this simple case already suggests mechanisms that may destabilize the evolutionary system: an increase in choice intensity β for evolutionary selection, a decrease in risk aversion a, a decrease in conditional variance of excess returns σ^2, or an increase in the diversity of purely biased beliefs σ_0^2. All these forces can push η beyond η_c, thereby triggering instability of the (fundamental) steady state.

For the LTL in Eq. 4.36, in the case of linear forecasting rules with three lags, a bifurcation route to chaos, with asset prices fluctuating around the unstable fundamental steady state, occurs when η is increased. This shows that a rational route to randomness can occur in an asset market with many different trader types, when traders

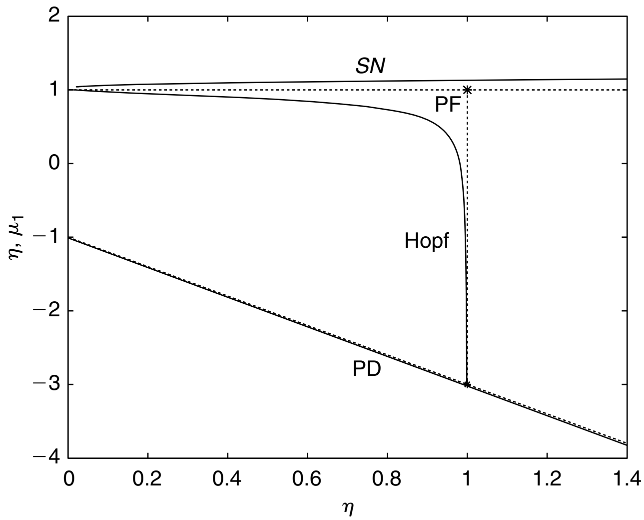

FIGURE 4.7 Bifurcation diagram in the (η, μ_1) parameter plane for LTL (Eq. 4.36), where μ_1 represents the mean of the first-order stochastic trend variable ϑ_{01} in the forecasting rule (Eq. 4.34). For $\mu_0 = a\sigma^2 = 0$, with μ_0 the mean of the constant ϑ_{00} in the forecasting rule (Eq. 4.34), the LTL is symmetric and thus nongeneric (dotted curves); when $\mu_0 \neq 0$, the LTL is nonsymmetric and generic. The diagram shows Hopf (H), period-doubling (PD), pitchfork (PF), and saddle-node (SN) bifurcation curves in the (η, μ_1) parameter plane, with other parameters fixed at $R = 1.01$, $z^s = 0$, $\mu_2 = \mu_3 = 0$, $\sigma_0 = \sigma_1 = \sigma_2 = 1$, and $\sigma_3 = 0$. Between the H and PD curves (and the PF curve when $\mu_0 = 0$), there is a unique, stable steady state. This steady state becomes unstable when crossing the H or the PD curve. Above the PF curve or the SN curve, the system has three steady states. The PF curve is nongeneric and only arises in the symmetric case with mean bias $\mu_0 = 0$. When the symmetry is broken by perturbing the mean bias to $\mu_0 = -0.1$, the PF curve "breaks" into generic Hopf and SN curves.

become increasingly sensitive to differences in fitness (i.e., an increase in the intensity of choice β) or traders become less risk averse (i.e., a decrease of the coefficient of risk-aversion a). In general, in a many-trader-type evolutionary world, fundamentalists will *not* drive out all other types and asset prices need not converge to their fundamental values.

Figure 4.7 shows a two-dimensional bifurcation diagram in the (η, μ_1) parameter plane, where μ_1 represents the mean of the first-order stochastic trend variable ϑ_{01} in the forecasting rule (Eq. 4.34). Recall that μ_0 is the mean of the constant term ϑ_{00} in the forecasting rule (Eq. 4.34); it models the "mean bias" of the trader type. When $\mu_0 = 0$ and $a\sigma^2 z_s = 0$ (expressing that the risk premium is zero), the LTL is symmetric with respect to the fundamental steady state.

In the symmetric case (dotted lines in Figure 4.7), for parameters taking values in the region enclosed by the Hopf, period-doubling (PD), and pitchfork (PF) bifurcation curves, the fundamental steady state is unique and stable.[18] As the parameters cross the PF curve, two additional nonfundamental steady states are created, one above and

[18]See Appendix 4.1 for a brief introduction to bifurcation theory, with Appendix 4.2 providing simple examples of the saddle-node, period doubling, pitchfork, and Hopf bifurcations.

one below the fundamental. Another route to instability occurs when crossing the Hopf curve, where the fundamental steady state becomes unstable and a (stable) invariant circle with periodic or quasi-periodic dynamics is created. The pitchfork bifurcation curve is *nongeneric* and occurs only in the symmetric case. When the symmetry is broken by a nonzero mean bias $\mu_0 \neq 0$, as illustrated in Figure 4.7 (bold curves) for $\mu_0 = -0.1$, the PF curve disappears and "breaks" into two generic codimension bifurcation curves: a Hopf and a saddle-node (SN) bifurcation curve. When crossing the SN curve from below, two additional steady states are created, one stable and one unstable. Notice that, as illustrated in Figure 4.7, when the perturbation is small (in the figure, $\mu_0 = -0.1$), the SN and the Hopf curves are close to the PF and the Hopf curves (dotted lines) in the symmetric case. In this sense the bifurcation diagram depends continuously on the parameters, and it is useful to consider the symmetric LTL as an "organizing" center to study bifurcation phenomena in the generic, nonsymmetric LTL.

The most relevant case from an economic viewpoint arises when the mean μ_1 of the first-order coefficient ϑ_{01} in the forecasting rule (Eq. 4.34) satisfies $0 \leq \mu_1 \leq 1$. In that case, the (fundamental) steady state loses stability in a Hopf bifurcation as η increases. Figure 4.8 illustrates the dynamical behavior of the LTL as the parameter η further

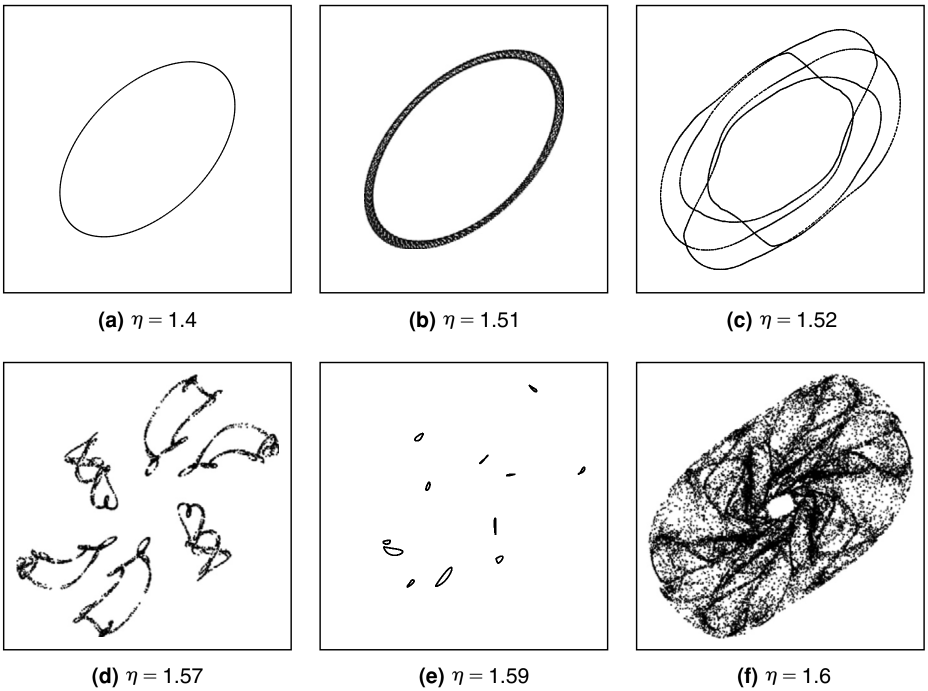

FIGURE 4.8 Attractors in the phase space for the 5-D LTL with parameters $R = 1.01$, $z^s = 0$, $\mu_0 = 0$, $\mu_1 = 0$, $\mu_2 = \mu_3 = 0$, and $\sigma_0 = \sigma_1 = \sigma_2 = \sigma_3 = 1$: (a) immediately after the Hopf bifurcation (quasi-)periodic dynamics on a stable invariant circle occurs; (b–c) after a Hopf bifurcation (quasi-)periodic dynamics on a stable invariant torus occurs; and (d–f) breaking up of the invariant torus into a strange attractor.

increases. After the Hopf bifurcation periodic and quasi-periodic dynamics on a stable invariant circle occur, and for increasing values of η, a bifurcation route to strange attractors occurs. Figure 4.8 thus presents numerical evidence of the occurrence of a rational route to randomness, that is, a bifurcation route to strange attractors as the intensity of choice to switch forecasting strategies increases. If such rational routes to randomness occur for the LTL, the LTL convergence theorem implies that in evolutionary systems with many trader types, rational routes to randomness occur with high probability.

Diks and van der Weide (2003, 2005) have generalized the notion of LTL and introduced so-called *continuous belief systems* (CBS), where the beliefs of traders are distributed according to a continuous density function. The beliefs distribution function and the equilibrium prices coevolve over time. The LTL theory discussed here as well as its extensions can be used to form a bridge between an analytical approach and the literature on evolutionary artificial market simulation models reviewed in LeBaron (2000, 2006). See also Anufriev et al. (2008) for a recent application of a LTL in a macroeconomic model with heterogeneous expectations.

4.5. EMPIRICAL VALIDATION

In this section we discuss the empirical validity of the asset pricing model with heterogeneous beliefs. There is already a large literature on heterogeneous agent models replicating many of the important stylized facts of financial time series on short time scales (say, daily or higher frequency), such as fat tails and long memory in the returns distribution and clustered volatility. Examples of HAMs able to replicate stylized facts of financial markets include, for example, Brock and LeBaron (1996), Arthur et al. (1997), Brock and Hommes (1997b), Youssefmir and Huberman (1997), LeBaron et al. (1999), Lux and Marchesi (1999, 2000), Farmer and Joshi (2002), Kirman and Teyssière (2002), Hommes (2002), Iori (2002), Cont and Bouchaud (2000), and Gaunersdorfer and Hommes (2007). The recent survey by Lux (2009) contains an extensive look at behavioral interacting agent models mimicking the stylized facts of asset returns. We have already seen examples of simple heterogeneous agent models mimicking temporary bubbles and crashes. In this section we discuss how these qualitative features match observed bubbles and crashes in real markets.

Empirical validation and estimation of HAMs on economic or financial data are still in their infancy. An early attempt has already been made by Shiller (1984) who presented a HAM with smart money traders having rational expectations versus ordinary investors (whose behavior is in fact not modeled at all). Shiller estimated the fraction of smart money investors over the period 1900 to 1983 and found considerable fluctuations of the fraction over a range between 0% and 50%. Baak (1999) and Chavas (2000) estimated HAMs on hog and beef market data and found evidence for heterogeneity of expectations. Winker and Gilli (2001) and Gilli and Winker (2003) estimated the model of Kirman (1991, 1993) with fundamentalists and chartists, using the daily DM-US$ exchange rates from 1991 to 2000. Their estimated parameter values correspond to a bimodal distribution of agents. Westerhoff and Reitz (2003) also estimated an HAM with fundamentalists and chartists to exchange rates and found considerable fluctuations of the market impact of fundamentalists. Alfarno et al. (2005) estimated an agent-based herding model where agents switch between fundamentalist and chartist strategies. Branch (2004) estimated a model with heterogeneous beliefs and time-varying fractions, using survey data on inflation expectations. In this section, we discuss the estimation of a simple two-type asset-pricing model with heterogeneous beliefs, as discussed in Section 4.2, on yearly S&P 500 data, 1871 to 2003, as done in Boswijk et al. (2007). As we will see, this simple two-type model can, for example, explain the dot-com bubble in the late 1990s and the subsequent crash starting at the end of 2000.[19]

4.5.1. The Model in Price-to-Cash Flows

In the previous sections, the dividend process of the risky asset has been assumed to be stationary. To estimate the model using yearly data of more than a century, the dividend process has to be taken growing over time and thus nonstationary. To estimate a simple two-type model, Boswijk et al. (2007) therefore reformulated the model in terms of price-to-cash flows. Recall from Eq. 4.5 that, under the assumption of zero net supply of the risky asset, the equilibrium pricing equation is

$$p_t = \frac{1}{1+r} \sum_{h=1}^{H} n_{h,t} E_{h,t}(p_{t+1} + y_{t+1}) \tag{4.38}$$

or equivalently

$$r = \sum_{h=1}^{H} n_{h,t} \frac{E_{h,t}[p_{t+1} + y_{t+1} - p_t]}{p_t} \tag{4.39}$$

In equilibrium the average required rate of return for investors to hold the risky asset equals the discount rate r. In the estimation of the model, the discount rate r has been set equal to the sum of the (risk-free) interest rate and the required risk premium on stocks. A simple, nonstationary process that fits cash flow data (dividends or earnings) well is a stochastic process with a constant growth rate. More precisely, assume that $\log y_t$ is a Gaussian random walk with drift; that is,

$$\log y_{t+1} = \mu + \log y_t + \upsilon_{t+1} \quad \upsilon_{t+1} \sim \text{i.i.d. } N(0, \sigma_\upsilon^2) \tag{4.40}$$

which implies

$$\frac{y_{t+1}}{y_t} = e^{\mu+\upsilon_{t+1}} = e^{\mu+\frac{1}{2}\sigma_\upsilon^2} e^{\upsilon_{t+1}-\frac{1}{2}\sigma_\upsilon^2} = (1+g)\varepsilon_{t+1} \tag{4.41}$$

[19]Van Norden and Schaller (1999) estimate a nonlinear time-series switching model with two regimes, an explosive and a collapsing bubble regime, with the probability of being in the explosive regime depending negatively on the relative absolute deviation of the bubble from the fundamental. Brooks and Katsaris (2005) extend this model to three regimes, adding a third dormant bubble regime where the bubble grows at the required rate of return without explosive expectations.

where $g = e^{\mu + \frac{1}{2}\sigma_\upsilon^2} - 1$ and $\varepsilon_{t+1} = e^{\upsilon_{t+1} + \frac{1}{2}\sigma_\upsilon^2}$, so that $E_t(\varepsilon_{t+1}) = 1$. As before, we assume that all types have correct beliefs on the cash flow; that is,

$$E_{h,t}[y_{t+1}] = E_t[y_{t+1}] = (1+g)y_t E_t[\varepsilon_{t+1}] = (1+g)y_t \tag{4.42}$$

Since the cash flow is an *exogenously* given stochastic process, it seems natural to assume that agents have learned the correct beliefs on next period's cash flow y_{t+1}. In particular, boundedly rational agents can learn about the constant growth rate by, for example, running a simple regression of $\log(y_t/y_{t-1})$ on a constant.

In contrast, prices are determined *endogenously* and are affected by *expectations* about next period's price. In a heterogeneous world, agreement about future prices therefore seems more unlikely than agreement about future cash flows. Therefore we assume homogeneous beliefs about future cash flow but heterogeneous beliefs about future prices.[20] The pricing Eq. 4.38 can be reformulated in terms of a price-to-cash-flow (P/Y) ratio, $\delta_t = p_t/y_t$, as[21]

$$\delta_t = \frac{1}{R^*}\left\{1 + \sum_{h=1}^{H} n_{h,t} E_{h,t}[\delta_{t+1}]\right\} \qquad R^* = \frac{1+r}{1+g} \tag{4.43}$$

In the special case, when all agents have *rational expectations*, the equilibrium pricing Eq. 4.38 simplifies to $p_t = (1/(1+r))E_t(p_{t+1} + y_{t+1})$. It is well known that, in the case of a constant discount rate r and a constant growth rate g for dividends, according to the static Gordon growth model (Gordon, 1962), the rational expectations fundamental price, p_t^*, of the risky asset is given by

$$p_t^* = \frac{1+g}{r-g} y_t \qquad r > g \tag{4.44}$$

Equivalently, in terms of price-to-cash-flow ratios, the fundamental is

$$\delta_t^* = \frac{p_t^*}{y_t} = \frac{1+g}{r-g} \equiv m \tag{4.45}$$

We refer to p_t^* as the *fundamental price* and to δ_t^* as the fundamental P/Y ratio. When all agents are rational, the pricing Eq. 4.43 in terms of the P/Y ratio, $\delta_t = p_t/y_t$, becomes

$$\delta_t = \frac{1}{R^*}\{1 + E_t[\delta_{t+1}]\} \tag{4.46}$$

[20] Barberis et al. (1998) consider a model in which agents are affected by psychological biases in forming expectations about future cash flows. In particular, agents may overreact to good news about economic fundamentals because they believe that cash flows have moved into another regime with higher growth. Their model is able to explain continuation and reversal of stock returns.

[21] In what follows we use either price-to-dividend (P/D) or price-to-earnings (P/E) ratios and use the general notation P/Y for price-to-cash flows.

In terms of the *deviation from the fundamental ratio*, $x_t = \delta_t - \delta_t^* = \delta_t - m$, this simplifies to

$$x_t = \frac{1}{R^*} E_t[x_{t+1}] \tag{4.47}$$

Under heterogeneity in expectations, the pricing Eq. 4.43 is expressed in terms of x_t as

$$x_t = \frac{1}{R^*} \sum_{h=1}^{H} n_{h,t} E_{h,t}[x_{t+1}] \tag{4.48}$$

Heterogeneous Beliefs

The expectation of belief type h about next period P/Y ratio is expressed as

$$E_{h,t}[\delta_{t+1}] = E_t[\delta_{t+1}^*] + f_h(x_{t-1}, \ldots, x_{t-L}) = m + f_h(x_{t-1}, \ldots, x_{t-L}) \tag{4.49}$$

where δ_t^* represents the fundamental price-to-cash-flow ratio P/Y, $E_t(\delta_{t+1}^*) = m$ is the rational expectation of the P/Y ratio available to all agents, x_t is the *deviation* of the P/Y ratio from its fundamental value, and $f_h(\cdot)$ represents the expected transitory deviation of the P/Y ratio from the fundamental value, depending on L past deviations. The information available to investors at time t includes present and past cash flows and past prices. In terms of deviations from the fundamental P/Y ratio, x_t, we get

$$E_{h,t}[x_{t+1}] = f_h(x_{t-1}, \ldots, x_{t-L}) \tag{4.50}$$

Note again that the rational expectations fundamental benchmark is nested in the heterogeneous agent model as a special case when $f_h \equiv 0$ for all types h. We can express Eq. 4.48 as

$$R^* x_t = \sum_{h=1}^{H} n_{h,t} f_h(x_{t-1}, \ldots, x_{t-L}) \tag{4.51}$$

From this equilibrium equation it is clear that the adjustment toward the fundamental P/Y ratio will be slow if a majority of investors has persistent beliefs about it.

Evolutionary Selection of Expectations

In addition to the empirical evidence of persistent deviations from fundamentals there is also significant evidence of time variation in the sentiment of investors. This has been documented, for example, by Shiller (1987, 2000) using survey data. In the model considered here, agents are boundedly rational and switch between different forecasting strategies according to recently realized profits. We denote by $\pi_{h,t-1}$ the realized profits of type h at the end of period $t-1$, given by (see Eq. 4.14):

$$\pi_{h,t-1} = R_{t-1} z_{h,t-2} = R_{t-1} \frac{E_{h,t-2}[R_{t-1}]}{aV_{t-2}[R_{t-1}]} \tag{4.52}$$

where $R_{t-1} = p_{t-1} + y_{t-1} - (1+r)p_{t-2}$ is the realized excess return at time $t-1$ and $z_{h,t-2}$ is the demand of the risky asset by belief type h, as given in Eq. 4.3, formed in period $t-2$.

As before, we assume that the beliefs about the conditional variance of excess returns are the same for all types and equal to fundamentalists beliefs about conditional variance, that is,

$$V_{h,t-2}[R_{t-1}] = V_{t-2}[P^*_{t-1} + y_{t-1} - (1+r)P^*_{t-2}] = y^2_{t-2}\eta^2 \tag{4.53}$$

where $\eta^2 = (1+m)^2(1+g)^2 V_{t-2}[\epsilon_{t-1}]$, with ϵ_t IID noise driving the cash flow. The fitness measure can be rewritten in terms of the deviation $x_t = \delta_t - m$ of the P/Y ratio from its fundamental value, with $m = (1+g)/(r-g)$, as

$$\pi_{h,t-1} = \frac{(1+g)^2}{a\eta^2} \left(x_{t-1} - R^* x_{t-2}\right) \left(E_{h,t-2}[x_{t-1}] - R^* x_{t-2}\right) \tag{4.54}$$

This fitness measure has a simple, intuitive explanation in terms of forecasting performance for next period's deviation from the fundamental. A positive demand $z_{h,t-2}$ may be seen as a bet that x_{t-1} would go up more than what was expected on average from $R^* x_{t-2}$. (Note that R^* is the growth rate of rational bubble solutions.) The realized fitness $\pi_{h,t-1}$ of strategy h is the realized profit from that bet and it will be positive if both the realized deviation $x_{t-1} > R^* x_{t-2}$ and the forecast of the deviation $E_{h,t-2}[x_{t-1}] > R^* x_{t-2}$. More generally, if both the realized absolute deviation $|x_{t-1}|$ and the absolute predicted deviation $|E_{h,t-2}[x_{t-1}]|$ to the fundamental value are larger than R^* times the absolute deviation $|x_{t-2}|$, strategy h generates positive realized fitness. In contrast, a strategy that wrongly predicts whether the asset price mean reverts back toward the fundamental value or moves away from it generates a negative realized fitness.

At the beginning of period t, investors compare the realized relative performances of the various strategies and withdraw capital from those that performed poorly and move it to better strategies. The fractions $n_{h,t}$ evolve according to a discrete choice model with multinomial logit probabilities, that is (see Eq. 4.13),

$$n_{h,t} = \frac{\exp[\beta \pi_{h,t-1}]}{\sum_{k=1}^{H} \exp[\beta \pi_{k,t-1}]} = \frac{1}{1 + \sum_{k \neq h} \exp[-\beta \Delta \pi^{h,k}_{t-1}]} \tag{4.55}$$

where $\beta > 0$ is the *intensity of choice* as before, and $\Delta \pi^{h,k}_{t-1} = \pi_{h,t-1} - \pi_{k,t-1}$ denotes the difference in realized profits of belief type h compared to type k.

4.5.2. Estimation of a Simple Two-Type Example

Consider the case of two types, both predicting next period's deviation by extrapolating past realizations in a linear fashion, that is,[22]

$$E_{h,t}[x_{t+1}] = f_h(x_{t-1}) = \varphi_h x_{t-1} \tag{4.56}$$

The dynamic asset-pricing model with two types can then be written as

$$R^* x_t = n_t \varphi_1 x_{t-1} + (1 - n_t)\varphi_2 x_{t-1} + \epsilon_t \tag{4.57}$$

where φ_1 and φ_2 denote the coefficients of the two belief types, n_t represents the fraction of investors that belong to the first type of traders and ϵ_t represents a disturbance term. The value of the parameter φ_h can be interpreted as follows: If it is positive and smaller than 1, investors expect the stock price to mean-revert toward the fundamental value. This type of agent represents *fundamentalists* because they expect the asset price to move back toward its fundamental value in the long run. The closer φ_h is to 1, the more persistent are the expected deviations. If the beliefs parameter φ_h is larger than 1, it implies that investors believe the deviation of the stock prices will grow over time at a constant speed. We will refer to this type of agent as a *trend follower*. Note in particular that when one group of investors believes in a strong trend, that is, $\varphi_h > R^*$, this may cause asset prices to deviate further from their fundamental value. In the case with two types with linear beliefs (Eq. 4.56), the fraction of Type 1 investors is

$$n_t = \frac{1}{1 + \exp\{-\beta^* [(\varphi_1 - \varphi_2)x_{t-3}(x_{t-1} - R^* x_{t-2})]\}} \tag{4.58}$$

where $\beta^* = \beta(1+g)^2/(a\eta^2)$.

The two-type model (Eqs. 4.57 and 4.58) has been estimated using an updated version of the dataset described in Shiller (1989), consisting of annual observations of the S&P 500 index from 1871 to 2003. Here we present the estimation results with earnings as cash flows, but using dividends as cash flows gives similar results. The valuation ratios are then the P/E ratios.[23]

Recall that according to the static Gordon growth model, the fundamental price is given by

$$p_t^* = m y_t \qquad m = \frac{1+g}{r-g} \tag{4.59}$$

The fundamental value of the asset is a multiple m of its cash flow, where m depends on the discount rate r and the cash-flow growth rate g. The multiple m can also be

[22]In the estimation of the model, higher-order lags turned out to be insignificant, so we focus on the simplest case with only one lag in the function $f_h(\cdot)$, with φ_h the parameter characterizing the strategy of type h.

[23]Since earnings data are noisy, to determine the fundamental valuation we follow the practice of Campbell and Shiller (2005) to smooth earnings by a 10-year moving average.

interpreted as the P/D or the P/E ratios implied by the present value model. Figure 4.9 shows the (log) of yearly S&P 500 data together with the fundamental benchmark as well as their P/E ratios. The figure shows a clear long-term comovement of the stock price and the fundamental value. However, the P/E ratio takes persistent swings away from the constant value predicted by the present-value model. This suggests that the fundamental value does not account completely for the dynamics of stock prices, as was suggested in the early debate on mean reversion by Summers (1986). A survey of the ongoing debate is given in Campbell and Shiller (2005). Here we use the simple

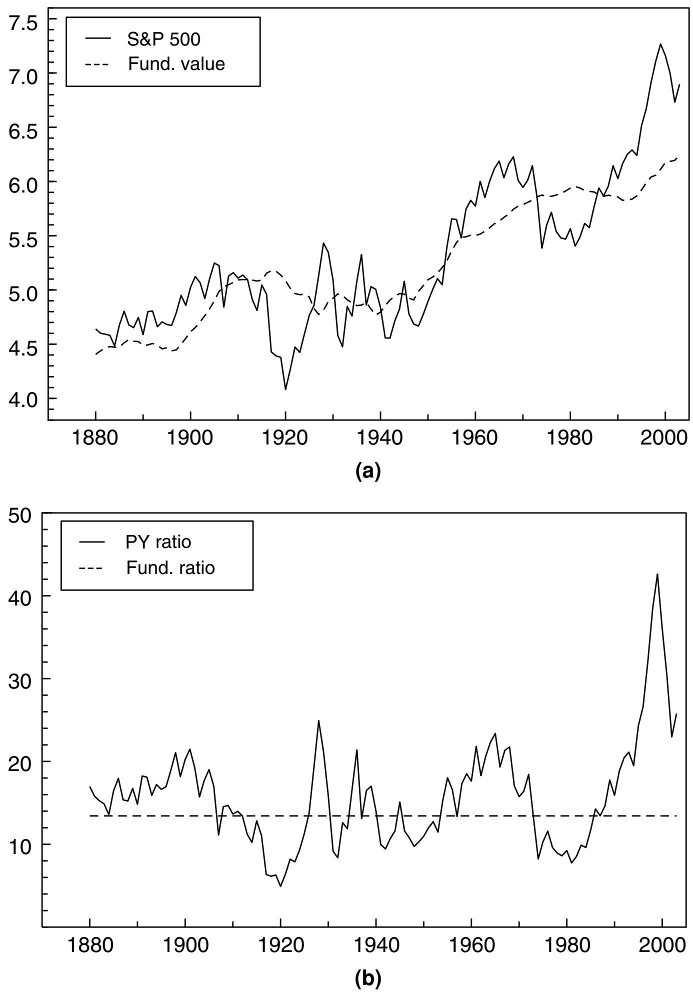

FIGURE 4.9 Yearly S&P 500, 1871–2003, and benchmark fundamental $p_t^* = my_t$, with $m = (1+g)/(1+r)$. (a) shows logs of S&P 500 and the log of the fundamental p_t^*; (b) shows the P/E-ratio of the S&P 500 around the constant fundamental benchmark $p_t^*/y_t = m$.

constant-growth Gordon model for the fundamental price and estimate the two-type model on deviations from this benchmark.[24]

Recall that $R^* = (1 + r)/(1 + g)$, where g is the constant growth rate of the cash flow and r is the discount rate equal to the risk-free interest rate plus a risk premium. We use an estimate of the risk premium—the difference between the expected return on the market portfolio of common stocks and the risk-free interest rate—to obtain R^*, as in Fama and French (2002). The risk premium safisfies

$$RP = g + y/p - i \tag{4.60}$$

where g is the growth rate of dividends, y/p denotes the average dividend yield y_t/p_{t-1} and i is the risk-free interest rate. For annual data from 1871 to 2003 of the S&P 500, the estimates are $i = 2.57\%$ and $RP = 6.56\%$ so that $r = 9.13\%$ and $R^* = 1.074$.[25] The corresponding average P/E ratio is 13.4, as illustrated in Figure 4.9.

Using yearly data of the S&P 500 index from 1871 to 2003, the parameters $(\varphi_1', \varphi_2', \beta^*)$ in the two-type model (Eqs. 4.57 and 4.58) can be estimated by nonlinear least squares. The estimation results are as follows:

$$\begin{aligned} R^* x_t &= n_t\{\underset{(0.074)}{0.80}\, x_{t-1}\} + (1 - n_t)\{\underset{(0.052)}{1.097}\, x_{t-1}\} + \widehat{\epsilon}_t \\ n_t &= \{1 + \exp[\underset{(4.93)}{-7.54}(-0.29 x_{t-3})(x_{t-1} - R^* x_{t-2})]\}^{-1} \end{aligned} \tag{4.61}$$

$$R^2 = 0.77,\ AIC = 2.23,\quad AIC_{AR(1)} = 2.29,\quad \varphi_{AR(1)} = 0.983,$$
$$Q_{LB}(4) = 0.94,\quad F^{boot}(p\text{-value}) = 10.15\ (0.011)$$

The belief coefficients are highly significant and different from each other. On the other hand, the intensity of choice β^* is not significantly different from zero. This is a common result in nonlinear switching-type regression models, where the parameter β^* in the transition function is difficult to estimate and has a large standard deviation because relatively large changes in β^* cause only a small variation of the fraction n_t.

Teräsvirta (1994) argues that this should not be worrying as long as there is significant heterogeneity in the estimated regimes. The nonlinear switching model achieves a lower value for the AIC selection criterion compared to a linear AR(1) model. This suggests that the model is capturing nonlinearity in the data. This is also confirmed by the bootstrap F-test for linearity, which strongly rejects the null hypothesis of linearity

[24]The same approach can be used for more general, time-varying fundamental processes. Manzan (2003) shows that a dynamic Gordon model for the fundamental price, where the discount rate r and/or the growth rate g are time varying, does not explain the large fluctuations in price-to-cash-flow ratios and in fact yields a fundamental price pattern close to that for the static Gordon model. Boswijk et al. (2007) also estimate a version of the model allowing for time variation in the growth rate of the cash flow and obtain similar results.

[25]These estimates are slightly different from Fama and French (2002) because, as in Shiller (1989), we use the CPI index to deflate nominal values.

in favor of the heterogeneous agent model. The residuals of the regression do not show significant evidence of autocorrelation at the 5% significance level.

The estimated coefficient of the first regime is 0.80, corresponding to a half-life of about three years. The first regime can be characterized as *fundamentalist* beliefs, expecting the asset price to move back toward its fundamental value. In contrast, the second regime has an estimated coefficient close to 1.1, implying that in this regime agents are *trend followers*, believing the deviation of the stock price to grow over time at a constant speed larger than $R^* \approx 1.074$. At times when the fraction of investors using this belief is equal or close to 1, we have explosive behavior in the P/E ratio. The sentiment of investors switches between a stable fundamentalists regime and a trend-following regime. In normal periods agents consider the deviation a temporary phenomenon and expect it to quickly revert to fundamentals. In other periods, a rapid increase in stock prices not paralleled by improvements in the fundamentals causes losses for fundamentalists and profits for trend followers. Evolutionary pressure will then cause more fundamentalists to become trend followers, thus reinforcing the trend in prices.

Figure 4.10 shows the time series of the fraction of fundamentalists and the *average market sentiment*, defined as

$$\varphi_t = \frac{n_t\varphi_1 + (1 - n_t)\varphi_2}{R^*} \tag{4.62}$$

It is clear that the fraction of fundamentalists varies considerably over time, with periods in which it is close to 0.5 and other periods in which it is close to either of the extremes 0 or 1. The series of the average market sentiment shows that there is significant time variation between periods of strong mean reversion when the market is dominated by fundamentalist and other periods in which φ_t is close to or exceeds 1 and the market is dominated by trend followers. These plots also offer an explanation of the events of the late 1990s: For six consecutive years the trend-following strategy outperformed the fundamentalist strategy and a majority of agents switched to the trend-following strategy, driving the average market sentiment beyond 1 and thus reinforcing the strong price trend. However, at the turn of the market in 2000, the fraction of fundamentalists increased again, approaching 1 and thus contributing to the reversal toward the fundamental value in subsequent years.

The estimation results show that there are two different belief strategies: one in which agents expect continuation of returns and the other in which they expect reversal. We also find that there are some years in which one type of expectation dominates the market. It is clear that the expectation of continuation of positive returns dominated the market in the late 1990s, with the average market sentiment coefficient φ_t in Eq. 4.62 larger than 1 in the late 1990s. Despite the awareness of the mispricing, in this period investors were aggressively extrapolating the continuation of the extraordinary performances realized in the previous years. Our approach endogenizes the switching of agents among beliefs. The evolutionary mechanism that relates predictor choice to their past performance is supported by the data. It also confirms previous evidence that pointed in this direction.

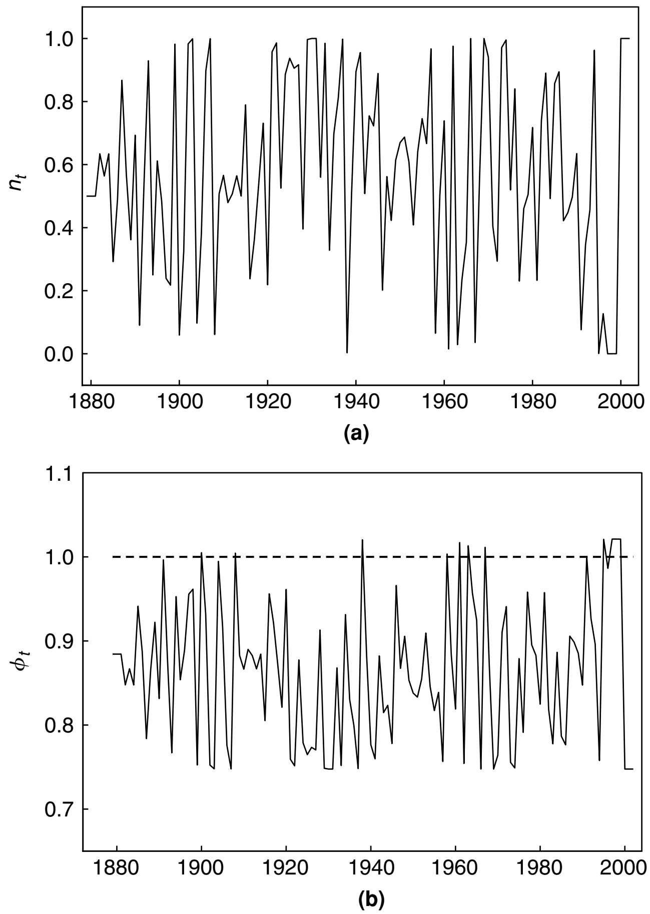

FIGURE 4.10 Estimated fraction n_t of fundamentalists (a) and market sentiment factor φ_t (b) in (Eq. 4.62) in two-type model.

Based on answers to a survey, Shiller (2000) constructed indices of "bubble expectations" and "investor confidence." In both cases, he found that the time variation in the indices is well explained by the lagged change in stock prices. Based on a different survey, Fisher and Statman (2002) found that in the late 1990s individual investors had expectations of continuation of recent stock returns while institutional investors were expecting reversals. This is an interesting approach to identifying heterogeneity of beliefs based on the type of investors rather than the type of beliefs. In the view of our model, the bubble in the 1990s was triggered by good news about economic fundamentals (a new Internet technology) and strongly reinforced by trend extrapolating behavior. The bubble was reversed by bad news about economic fundamentals (excessive growth cannot last forever and is not supported by earnings), and the crash was accelerated by switching of beliefs back to fundamentals.

4.5.3. Empirical Implications

In this subsection we discuss some empirical implications of the estimation of our nonlinear evolutionary switching model with heterogeneous beliefs. First, we investigate the response to a positive shock to fundamentals when the asset is overvalued. Second, we address the question concerning the probability that a bubble may resume by considering the evolution of the valuation ratios conditional on data until the end of 2003. These simulation experiments both show the importance of considering *nonlinear* effects in the dynamics of stock prices.

Response to a Fundamental Shock

We use the estimated parameters to investigate the response of the market valuation to good news. Assume that at the beginning of period t the cash flow increases due to a permanent increase in its growth rate. This implies that the asset has a higher fundamental valuation ratio, but what is the effect on the market valuation? We address this question both for the nonlinear switching model and a linear benchmark. The linear model may be interpreted as a model with a representative agent believing in an average mean reversion toward the fundamental.[26] Assume that at $t-1$ the fundamental valuation ratio was 15 and the good news at time t drives it to 17. Assume also that the equilibrium price at $t-1$ was 16. Figure 4.11 shows the valuation ratio dynamics in response to the good news for both the linear and the nonlinear switching models.

The figure shows the average price path over 2000 simulations of the estimated models. There is a clear difference between the linear and the nonlinear models. In the linear case, the positive shock to the fundamental value leads to an immediate increase of the price followed by mean-reversion thereafter. In contrast, for the nonlinear heterogeneous agent model, the pattern that emerges is consistent with the evidence of short-run continuation of positive returns and long-term reversal. After good news, the agents incorporate the news into their expectations and they expect that part of the previous period overvaluation will persist. One group—the trend followers—overreacts and expects a further increase of the price, while the other group—the fundamentalists—expects the price to diminish over time. The equilibrium price at time t overshoots and almost reaches 18. In the following two periods trend followers continue to buy the stock and drive its price and its valuation ratio even higher. Finally, the reversal starts and drives the ratio back to its long-run fundamental value. Initially, the aggressive investors interpret the positive news as a confirmation that the stock overvaluation was justified by forthcoming news. However, the lack of further good news convinces most investors to switch back to the mean-reverting expectations and the stock price is driven back toward its fundamental value.

[26] The linear benchmark is, for example, obtained for $\beta = 0$ or equivalently when both fractions $n_t = 1 - n_t = 0.5$ in Eq. 4.57. Hence, in the linear model there is no strategy switching between different types but rather a representative agent with a linear forecasting rule.

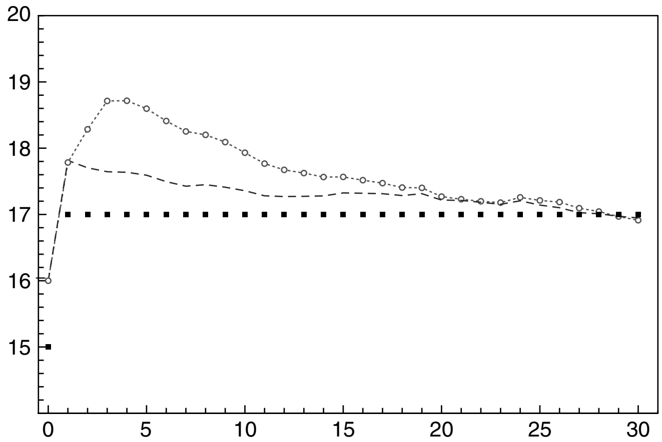

FIGURE 4.11 Average response (over 2000 simulations) to a shock to the fundamental for the linear representative agent model (dashed line) and the nonlinear two-type switching model (dotted line with circles). At period 0 there is a permanent shock to the fundamental price from 15 to 17. The simulation uses the estimated parameter values for the P/D ratio, with a representative agent average belief parameter $\varphi = 0.968$ and heterogeneous agent parameters $\varphi_1 = 0.762$ and $\varphi_2 = 1.135$ for the two types. The nonlinear heterogeneous agent model exhibits short-run continuation of positive returns and long-term reversal.

Will the Bubble Resume?

As a model forecasting exercise, we simulate the evolution of the valuation ratios using the estimated heterogeneous agent model. We then obtain the predicted evolution of the valuation ratio conditional on the value realized at the end of 2003. Innovations are obtained by reshuffling the estimated residuals. Instead of focusing our attention only on the mean or the median of the distribution, we consider the quantiles corresponding to 10%, 30%, 50%, 70%, and 90% probability over 2000 replications of the estimated model in Eq. 4.61 for the P/E ratio. In addition to the quantiles predicted by our nonlinear model, we also plot those predicted by the linear representative agent model. Figure 4.12 shows the 1- to 5-year periods ahead quantiles of the estimated model's predictive distribution.

The linear model—Figure 4.12(a)—predicts that the valuation ratio reverts toward the mean at all quantiles considered. In contrast, the nonlinear switching model predicts that there is a significant probability that the ratio may increase again as a result of the activation of the trend-following regime. The 70% and 90% quantiles clearly show that the PD ratio may increase again to levels close to 35. Stated differently, our heterogeneous agent model predicts that with probability over 30% the PD ratio may increase to more than 30. Note, however, that the median predicts that the ratio should decrease as implied by the linear mean-reverting model.

Another implication of our model is that if the first (mean-reverting) regime dominates the beliefs of investors, it will enforce a much faster adjustment than predicted by the linear model. This is clear from the bottom quantiles of the distributions. These

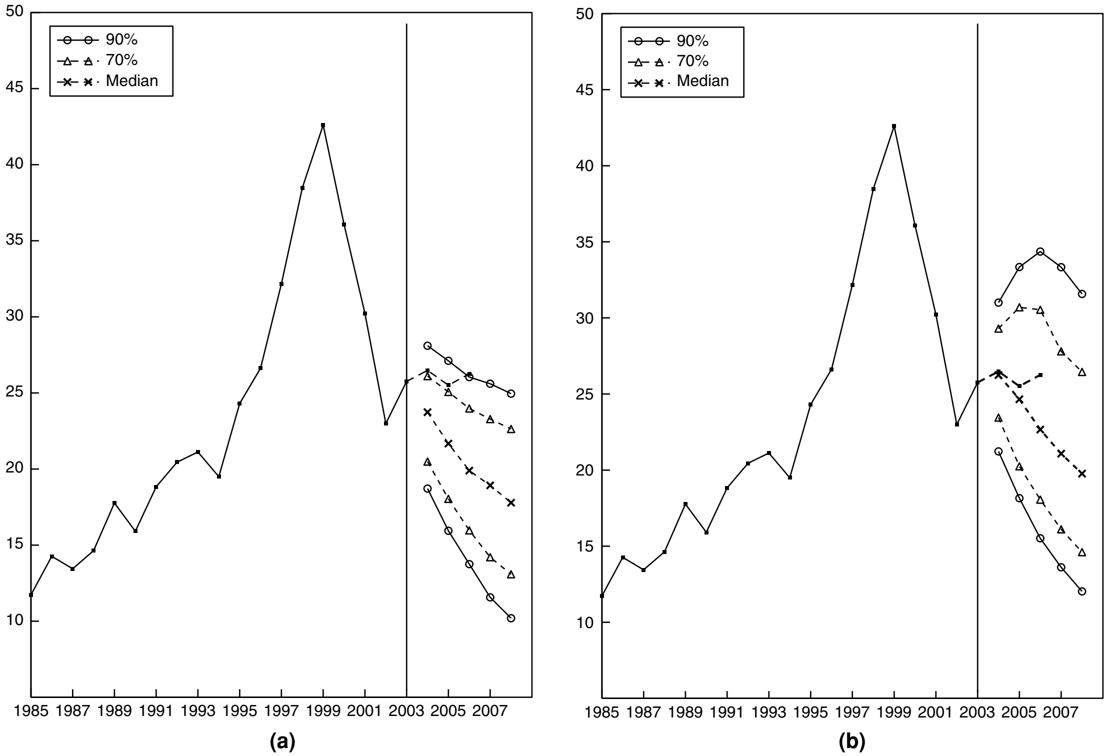

FIGURE 4.12 Prediction of P/E ratios five years ahead, based on linear representative agent model (a) and nonlinear two-type switching model (b). The estimated belief parameters are $\varphi = 0.983$ for the representative agent and $\varphi_1 = 0.80$ and $\varphi_2 = 1.097$ for the two-type switching model. The quantiles corresponding to 10, 30, 50, 70, and 90% probability over 2000 replications are shown. The realized P/E ratios, 2004–2006, are also indicated and show that the realized P/E ratio for 2006 falls outside the 10% quantile of the linear model.

simulations show that predictions from a linear, representative agent model versus a nonlinear, heterogeneous agent model are quite different. In particular, extreme events with large deviations from the benchmark fundamental valuation are much more likely in a nonlinear world.

4.6. LABORATORY EXPERIMENTS

Asset-pricing models with heterogeneous beliefs exhibit interesting dynamics characterized by temporary bubbles and crashes, triggered by news about fundamentals and reinforced by self-fulfilling expectations and trend-following investment strategies. The previous section focused on the empirical relevance of such models; this section confronts the model with data from laboratory experiments with human subjects. Laboratory experiments are well suited to discipline the class of behavioral modes (or heuristics) boundedly rational subjects may use in economic decision making. Here, we discuss a number of "learning to forecast experiments" in which subjects must forecast the price of an asset for which the realized market price is an aggregation of individual expectations.

In real markets, it is hard to obtain detailed information about investors' individual expectations. One approach is to collect survey data on individual expectations, as done for example by Turnovsky (1970) on expectations about the Consumer Price Index and the unemployment rate during the post-Korean War period. Frankel and Froot (1987a,b, 1990a,b), Allen and Taylor (1990), Ito (1990), and Taylor and Allen (1992) use a survey on exchange rate expectations and conclude that financial practitioners use different forecasting and trading strategies. A consistent finding from survey data is that at short horizons, investors tend to use extrapolative chartists' trading rules, whereas at longer horizons investors tend to use mean-reverting fundamentalists' trading rules. Shiller (1987, 1990, 2000) analyzes surveys on expectations about stock market prices and real estate prices and finds evidence for time variation in investors' sentiment; see also Vissing-Jorgensen (2003).

Laboratory experiments with human subjects provide an alternative, complementary approach to study the interaction of individual expectations and the resulting aggregate outcomes. An important advantage of the experimental approach is that the experimenter has full control over the underlying economic fundamentals. Surprisingly little experimental work has focused on expectation formation in markets. Williams (1987) considers expectation formation in an experimental double auction market that varies from period to period by small shifts in the market-clearing price. Participants predict the mean contract price for four or five consecutive periods. The participant with the lowest forecast error earns $1.00.

In Smith, Suchanek, and Williams (1998), expectations and the occurrence of speculative bubbles are studied in an experimental asset market. In a series of papers, Marimon, Spear, and Sunder (1993) and Marimon and Sunder (1993, 1994, 1995) studied expectation formation in inflationary overlapping generations' economies. Marimon, Spear, and Sunder (1993) find experimental evidence for expectationally driven cycles and coordination of beliefs on a sunspot two-cycle equilibrium but only after agents have been exposed to exogenous shocks of a similar kind. Marimon and Sunder (1995) present experimental evidence that a "simple" rule, such as a constant growth of the money supply, can help coordinate agents' beliefs and help stabilize the economy. Duffy (2006, 2008) gives stimulating surveys of laboratory experiments in various macro settings and how individual and aggregate behavior could be explained by agent-based models.

However, most of these papers cannot be viewed as pure experimental testing of the expectations hypothesis, everything else being constant, because in the experiments dynamic market equilibrium is affected not only by expectations feedback but also by other types of human decisions, such as trading behavior. A number of laboratory experiments have focused on expectation formation exclusively. Schmalensee (1976) presented subjects with historical data on wheat prices and asked them to predict the mean wheat price for the next five periods. In Dwyer et al. (1993) and Hey (1994), subjects had to predict a time series generated by a stochastic process such as a random walk or a simple linear first-order autoregressive process; in the last two papers no economic context was given. Kelley and Friedman (2002) considered learning in an orange juice futures price-forecasting experiment, where prices were driven by a linear stochastic process with two exogenous variables (weather and competing supply). A drawback common to these papers is that the historical or stochastic price series are *exogenous* and there is no feedback from subjects' forecasting behavior.

4.6.1. Learning to Forecast Experiments

In the remaining part of this section we mainly focus on the learning-to-forecast experiments in Hommes et al. (2005). In these experiments, subjects forecast the price of a risky asset, which is determined by market clearing with feedback from individual expectations. Similar experiments have been performed by van de Velden (2001), Gerber et al. (2002), Sutan and Willinger (2005), Adam (2007), Hommes et al. (2007), and Heemeijer (2007); see also the recent survey in Duffy (2008). We are particularly interested in the following questions:

- How do boundedly rational agents form individual expectations, and how do they learn in a heterogeneous world?
- How do individual forecasting rules interact, and what is the aggregate outcome of these interactions?
- Will coordination occur, even when there is limited market information?
- Does learning enforce convergence to rational expectations equilibrium?

In real financial markets, traders are involved in two related activities: prediction and trade. Traders make a prediction concerning the future price of an asset, and given this prediction, they make a trading decision. In the experiments discussed here, the only task of the subjects is to forecast prices; asset trading is computerized and derived from optimal demand (from mean-variance maximization), given the individual forecast. The experiments can therefore be seen as learning to forecast experiments (Marimon and Sunder 1994, p. 134), in contrast to learning to solve intertemporal optimization problems or, more concisely, learning to optimize experiments (Duffy 2006, p. 4), where participants are asked to submit their decisions (e.g., trading or consumption quantities) while their private beliefs about future developments remain implicit. Learning to forecast experiments provide us with "clean" data on expectations, which can be used to test various expectations hypotheses.

In the experiments, each participant is told that he or she is an advisor to a pension fund, with the only task to predict next period's price of a risky asset. Earnings are given by a (truncated) quadratic scoring rule:

$$e_{ht} = \max\left\{ 1300 - \frac{1300}{49}\left(p_t - p^e_{ht}\right)^2, 0 \right\} \qquad \textbf{(4.63)}$$

where 1300 points is equivalent to 0.5 euros, and earnings are 0 in period t when $\left|p_t - p^e_{ht}\right| \geq 7$. Subjects are informed that their pension fund needs to decide how much to invest in a risk-free asset paying a risk-free gross rate of return $R = 1 + r$, where r is the real interest rate, and how much to invest in shares of an infinitely lived risky asset. The risky asset pays uncertain IID dividends y_t with mean $\bar{y}$. The mean dividend $\bar{y}$ and the interest rate r are common knowledge, so that the subjects could compute the (constant) fundamental $p^* = \bar{y}/r = 3/0.05 = 60$. Subjects know that the price of the asset is determined by market clearing. Although they do not know the exact underlying market-clearing equation, they have qualitative information about the market and are informed that the higher their forecast, the larger will be the fraction of money of their pension fund invested in the risky asset and the larger will be the demand for stocks. They do not know the exact investment strategy of their pension fund and the investment strategies of the other pension funds. They also do not know the number of pension funds (which is six) or the identity of the other members of the group.

The experiment lasts for 51 periods. In every period t the participants have to predict the price p_{t+1} of the risky asset in period $t + 1$, given the available information consisting of past prices $p_{t-1}, p_{t-2}, \ldots, p_1$ and the participants' own past individual predictions $p^e_{ht}, p^e_{h,t-1}, \ldots, p^e_{h1}$. Notice that the participants have to make a two-period-ahead forecast for p_{t+1}, since p_{t-1} is the latest available price observation. Subjects are told that their price forecast has to be between 0 and 100 for every period. In periods 1 and 2, no information about past prices is available. At the end of period t, when all predictions for period $t + 1$ have been submitted, the participants are informed about the price in period t and earnings for that period are revealed. On their computer screens, the subjects are informed about their earnings in the previous period, total earnings, a table of the last 20 prices, and their corresponding predictions and time series of the prices and their predictions. Subjects have no information about earnings and predictions of others.

4.6.2. The Price-Generating Mechanism

The asset market is populated by six pension funds and a small fraction of fundamentalist robot traders. Each pension fund h is matched with a participant and makes an investment decision at time t based on this participant's prediction $p^e_{h,t+1}$ of the asset price. The fundamentalist trader always predicts the fundamental price p^f and trades based on this prediction.

The realized asset price in the experiment is determined by market clearing, with the pension fund's asset demand derived from mean-variance maximization, given their advisor's forecast, as in the standard asset-pricing model with heterogeneous beliefs (e.g., Campbell, Lo, and MacKinlay, 1997; Brock and Hommes, 1998; see Section 4.2). The market-clearing price is given by Eq. 4.5:

$$p_t = \frac{1}{1+r}\left[(1 - n_t)\,\bar{p}_t\right] \tag{4.64}$$

where $\bar{p}^e_{t+1} = \frac{1}{6}\sum_{h=1}^{6} p^e_{h,t+1}$ is the *average forecast* for period $t + 1$ of the six participants, n_t is the time-varying weight of the fundamentalist traders, and ε_t is a noise term, representing (small) stochastic demand and supply shocks. Note that the realized asset price p_t at time t is determined by the individual price predictions $p_{h,t+1}$ for time $t+1$. Therefore, when traders have to make a prediction for the price in period $t+1$, they do not know the price in period t yet, and they can only use information on prices up to time $t-1$.

The weight n_t of the fundamental traders in the market is endogenous and depends positively on the absolute distance between the asset price and the fundamental value according to:

$$n_t = 1 - \exp\left(-\frac{1}{200}\left|p_{t-1} - p^f\right|\right) \qquad \textbf{(4.65)}$$

The greater this distance, the more the fundamentalist trader will buy or short the asset. The fundamentalist trader therefore acts as a "stabilizing force" pushing prices in the direction of the fundamental price. Their presence excludes the possibility of everlasting speculative bubbles in asset prices.[27] Also note that $n_t = 0$, if $p_{t-1} = p^f$.

An important feature of the asset-pricing model is its self-confirming nature or positive feedback: If all traders make a high (low) prediction, the realized price will also be high (low). This feature is characteristic for speculative asset markets: If traders expect a high price, the demand for the risky asset will be high, and as a consequence the realized market price will be high, assuming that the supply is fixed.

4.6.3. Benchmark Expectations Rules

Figure 4.13 shows the price dynamics under three benchmark expectation rules: rational expectations, naïve expectations, and a trend extrapolation rule. In the rational expectations' benchmarks, all agents forecast the price to be equal to its fundamental value $p^f = 60$.[28] Realized prices are then given by

$$p_t = p^f + \frac{1}{1+r}\varepsilon_t \qquad \textbf{(4.66)}$$

Therefore, under rational expectations, prices exhibit small random fluctuations around the fundamental price $p^f = 60$. This outcome of the experiment should probably not be expected right from the start, but perhaps subjects can learn to coordinate on the rational, fundamental forecast.

[27] DeGrauwe et al. (1993) discuss a similar stabilizing force in an exchange rate model with fundamentalists and chartists. In the same spirit, Kyle and Xiong (2001) introduce a long-term investor who holds a risky asset in an amount proportional to the spread between the asset price and its fundamental value. Since in the experiments the fundamental value is $p^f = 60$, the weight of the fundamentalist traders is bounded above by $\bar{n} = 1 - \exp\left(-\frac{3}{10}\right) \approx 0.26$. The weight of the other traders is the same for each trader and equal to $(1 - n_t)/6 \leq 0.17$.

[28] Recall that participants know the values of $\bar{y}$ and r and therefore have enough information to compute the fundamental value and predict it for any period.

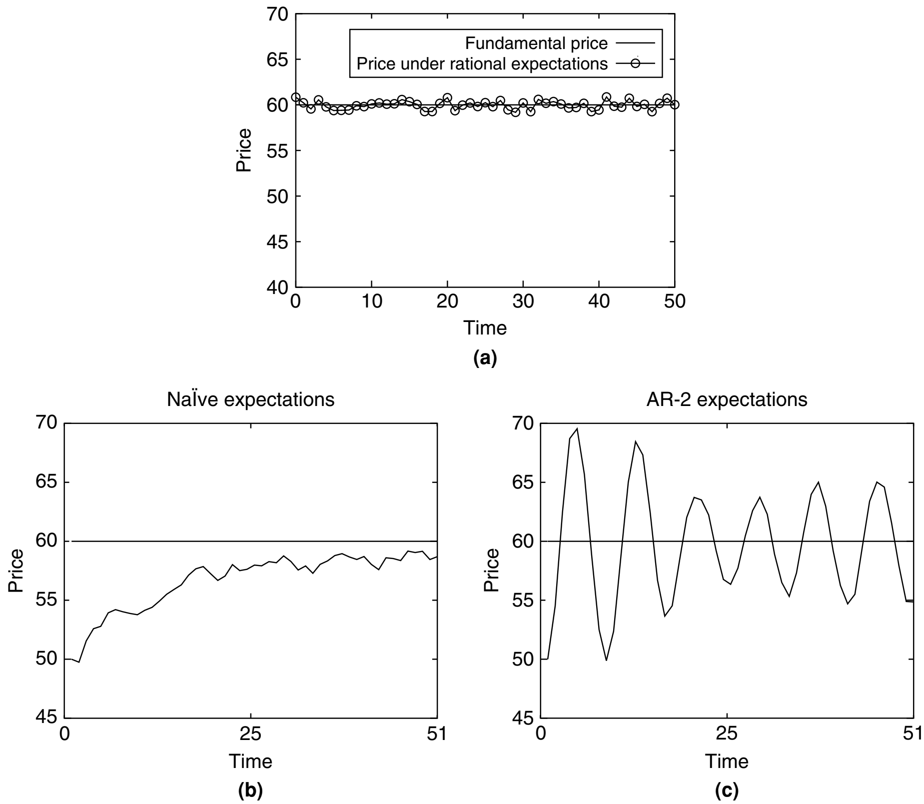

FIGURE 4.13 Realized prices under some benchmark expectation rules. (a) Under rational expectations, prices remain very close to the fundamental price 60; (b) under naïve expectations, prices converge monotonically, but slowly, toward the fundamental price; (c) under a simple linear AR-2 rule, $p^e_{t+1} = (p_{t-1} + 60)/2 + (p_{t-1} - p_{t-2})$, which may be interpreted as an anchor and adjustment rule, prices exhibit persistent oscillations.

Under *naïve expectations* all participants use the last observed price as their forecast, that is, $p^e_{h,t+1} = p_{t-1}$. The asset price then converges (almost) monotonically toward the fundamental price, as illustrated in Figure 4.13. Finally, the figure also illustrates what happens when all subjects use the simple trend extrapolation rule:

$$p^e_{h,t+1} = \frac{(60 + p_{t-1})}{2} + p_{t-1} - p_{t-2} \tag{4.67}$$

If all subjects use the forecasting rule (Eq. 4.67), realized market prices will fluctuate for 50 periods. This simple rule may be viewed as an anchor and adjustment heuristic, following the terminology of Tversky and Kahneman (1974), since it uses an anchor (the average of the fundamental price and the last observed price) and extrapolates the last price change from there. One may wonder how subjects would arrive at this anchor if they do not know the fundamental, but quite surprisingly a number of subjects used a rule very similar to Eq. 4.67.

4.6.4. Aggregate Behavior

Figure 4.14 shows time series of the realized asset prices and individual predictions in the experiments for five different groups. The first three groups illustrate the three typical qualitatively different outcomes in the treatment with robot traders:

- *Monotonic convergence*. The price converges (atmost) monotonically to the fundamental price from below.
- *Persistent oscillations*. The price oscillates with more or less constant amplitude; there is no convergence of the price to its fundamental value.
- *Dampened oscillations*. The price oscillates around the fundamental price with large amplitude initially, but the amplitude decreases over time, indicating (slow) convergence to the fundamental price.

The last two groups in Figure 4.14 illustrate what happens in a different treatment of the experiments without fundamental robot traders. When there are no fundamental robot traders present in the market, persistent price oscillations with large amplitude typically occur. The difference between these last two groups lies in the upper bound for price predictions, set to 100 (as in the case with robot traders) and 1000, respectively. Hommes et al. (2008) ran experiments without robot traders and a high upper bound of 1000 (maintaining the same fundamental price $p^f = 60$) and in six out of their seven markets long-lasting price bubbles (almost) reaching the upper bound were observed, with price levels up to 15 times the fundamental value.

Comparing the experimental results in Figure 4.14 with the simulated benchmarks in Figure 4.13 one observes that realized prices under naïve expectations resemble realized prices in the case with monotonic convergence remarkably well. On the other hand, the case of persistent oscillatory behavior in the experiment is qualitatively similar to the asset price behavior when participants use a simple $AR(2)$ prediction strategy. Clearly, naïve and $AR(2)$ prediction strategies give a qualitatively much better description of aggregate asset price fluctuations in the experiment than does the benchmark case of rational expectations. Recall from Subsection 4.6.3 that an $AR(2)$ rule has a simple behavioral interpretation as an anchor and adjustment trend-following forecasting strategy.

4.6.5. Individual Prediction Strategies

In this subsection we discuss some characteristics and estimation of individual prediction strategies. Some participants try to extrapolate observed trends and by doing so overreact and predict too high or too low. Other participants are more cautious when submitting predictions and use adaptive expectations, that is, an average of their last forecast and the last observed price. An individual degree of overreaction can be

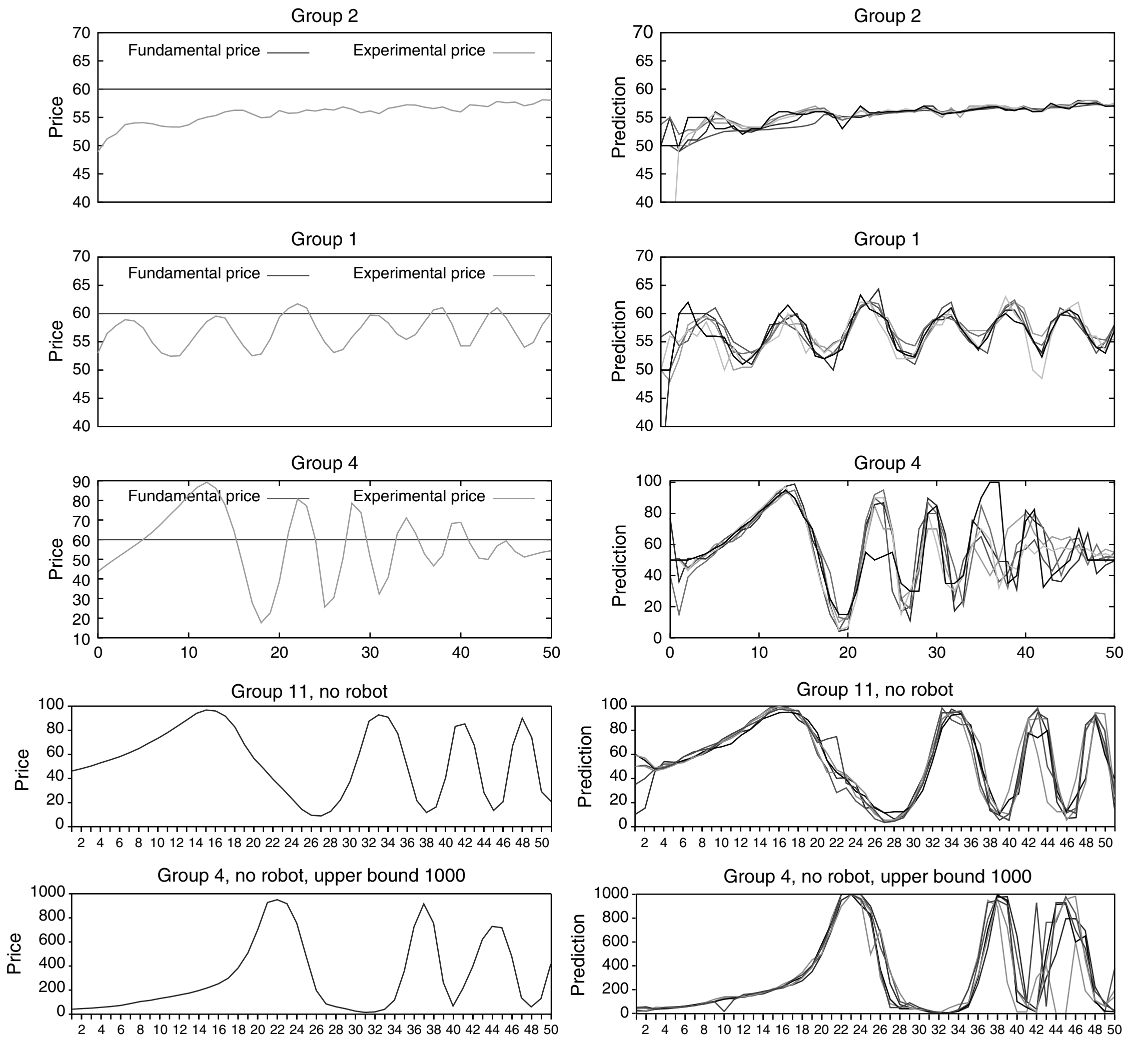

FIGURE 4.14 Realized prices (a) and individual predictions (b) in five typical asset-pricing experiments. The fundamental price $p^f = 60$ is indicated by a horizontal line. The first three examples show three different outcomes in the experiments with robot traders: monotonic convergence, persistent oscillations, and dampened oscillations. The last two examples show experiments without robot traders for upper bounds of 100 and 1000, respectively. The right panels show a striking coordination of individual forecasts.

quantified as the average absolute (one-period) change in predictions of participant h:

$$\triangle_h^e = \frac{1}{41} \sum_{t=11}^{51} \left| p_{ht}^e - p_{h,t-1}^e \right| \tag{4.68}$$

The average absolute change in the price is given by $\triangle = \frac{1}{41} \sum_{t=11}^{51} |p_t - p_{t-1}|$. We will say that individual h *overreacts* if $\triangle_h^e > \triangle$ and we will say that individual h is

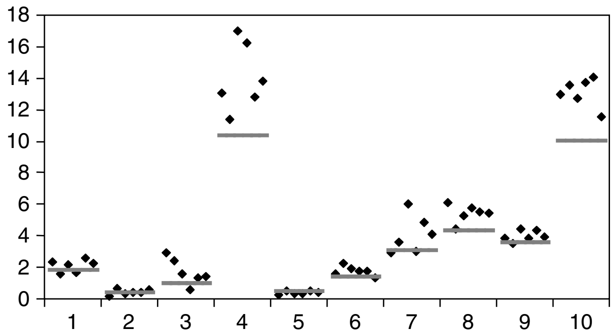

FIGURE 4.15 Individual degrees of overreactions for 10 different groups, all with a robot trader: the first seven with a fundamental $p^f = 60$ and the last three with a fundamental $p^f = 40$. The line segments represent the average absolute price change; the dots represent the average absolute changes in individual forecasts. Dots above the line segments correspond to individual overreaction.

cautious if $\triangle_h^e \leq \triangle$. Figure 4.15 illustrates the individual degree of overreaction for the different groups. In the case of monotonic convergence (groups 2 and 5), there is no overreaction; in the case of permanent oscillations (groups 1, 6, 8, and 9) a majority of subjects shows some overreaction, but it is relatively small. In the case of dampened oscillations (groups 4, 7, and 10), with large temporary bubbles in the initial phases of the experiment, a majority of participants strongly overreacts. Oscillatory behavior and temporary bubbles are thus caused by overreaction of a majority of agents.

Individual prediction strategies have been estimated using a simple linear model:

$$p_{h,t+1}^e = \alpha_h + \sum_{i=1}^{4} \beta_{hi} p_{t-i} + \sum_{j=0}^{3} \gamma_{hj} p_{ht-j}^e + \nu_t \tag{4.69}$$

where ν_t is an IID noise term. This general setup includes several important special cases: (1) naïve expectations ($\beta_{h1} = 1$, all other coefficients equal to 0); (2) adaptive expectations ($\beta_{h1} + \gamma_{h0} = 1$, all other coefficients equal to 0), and (3) $AR(L)$ processes (all coefficients equal to 0, except α_h, $\beta_{h1}, \ldots, \beta_{hL}$). The estimation results for 60 participants (using observations $t = 11$ to $t = 51$) can be summarized as follows:

- For more than 90% of the individuals, the simple linear rule (Eq. 4.69) describes forecasting behavior well.
- In the monotonically converging markets, a majority of subjects uses a naïve, an adaptive, or an AR(1) forecasting rule.
- In the dampened and persistently oscillating markets, a majority of subjects uses simple AR(2) or AR(3) forecasting rules; in particular, a number of subjects use a simple trend-following rule of the form:

$$p_{h,t+1}^e = p_{t-1} + \delta_h (p_{t-1} - p_{t-2}) \qquad \delta_h > 0 \tag{4.70}$$

This forecasting rule corresponds to positive feedback of momentum traders.

Within each group, participants learn to coordinate on a simple forecasting rule, which becomes self-fulfilling. If participants coordinate on an adaptive or AR(1) forecasting rule, the asset price monotonically converges to the fundamental price. In contrast, if the participants coordinate on a trend-following rule, transitory or even permanent price oscillations may arise, with persistent deviations from fundamental price. Anufriev and Hommes (2008) extended the adaptive belief systems in Section 4.2 and developed an evolutionary heuristics-switching model, matching all three different observed patterns in the learning to forecasting experiments remarkably well.

4.6.6. Profitability

In the learning-to-forecast experiments, subjects have been rewarded by their forecasting performance. As discussed in Hommes (2001), the fitness measure of (minus) squared forecasting errors is equivalent to risk-adjusted profits, and therefore it may be a relevant measure in real markets (see Footnote 6). But it is interesting to investigate the corresponding realized profits of the pension funds. In this section therefore we briefly discuss the profitability, that is, the (nonrisk-adjusted) realized profits, of the investment strategies. Realized profits of a mean-variance investment strategy based on a price forecast $p^e_{h,t+1}$ are given by:

$$\pi_{ht} = (p_{t+1} + y_{t+1} - Rp_t)(p^e_{h,t+1} + \bar{y} - Rp_t) \tag{4.71}$$

As a typical example, Figure 4.16 shows the realized profits and the realized accumulated profits in group 4, that is, a group with a robot trader and upperbound 100 exhibiting dampened price oscillations (the third panel in Figure 4.14). Figure 4.16 shows the realized profits corresponding to the six individual forecasts, the realized profits of the fundamental robot trader, and the realized profits of a hypothetical switching strategy, together with the realized price series (scaled by a factor of 3).

Clearly there are large fluctuations in the realized profits and all strategies occasionally suffer from large losses. The fundamental strategy starts with positive profits in periods 1 to 6 as the asset price rises from below the fundamental. When the asset price rises above its fundamental value and the bubble starts, the fundamental strategy makes large losses in periods 7 to 13. At the peak of the first bubble, at period 13, the fundamental strategy has performed second to last on average, with the second to lowest accumulated realized profits. During the crash, however, the fundamentalists make huge profits because they have built a large short position in the risky asset. At the same time, most other (trend-following) strategies suffer large losses because they hold long positions. As the crash continues and the asset price falls below its fundamental value, the fundamental strategy starts making losses again. At the bottom of the market in period 19, fundamentalists make a large loss. Over the full sample of 50 periods, however, on average the fundamental strategy performs very well and accumulates more profits than the other six forecasting strategies. Figure 4.16 also shows that the fundamental strategy is beaten by a switching strategy, always selecting the best (according to last period's

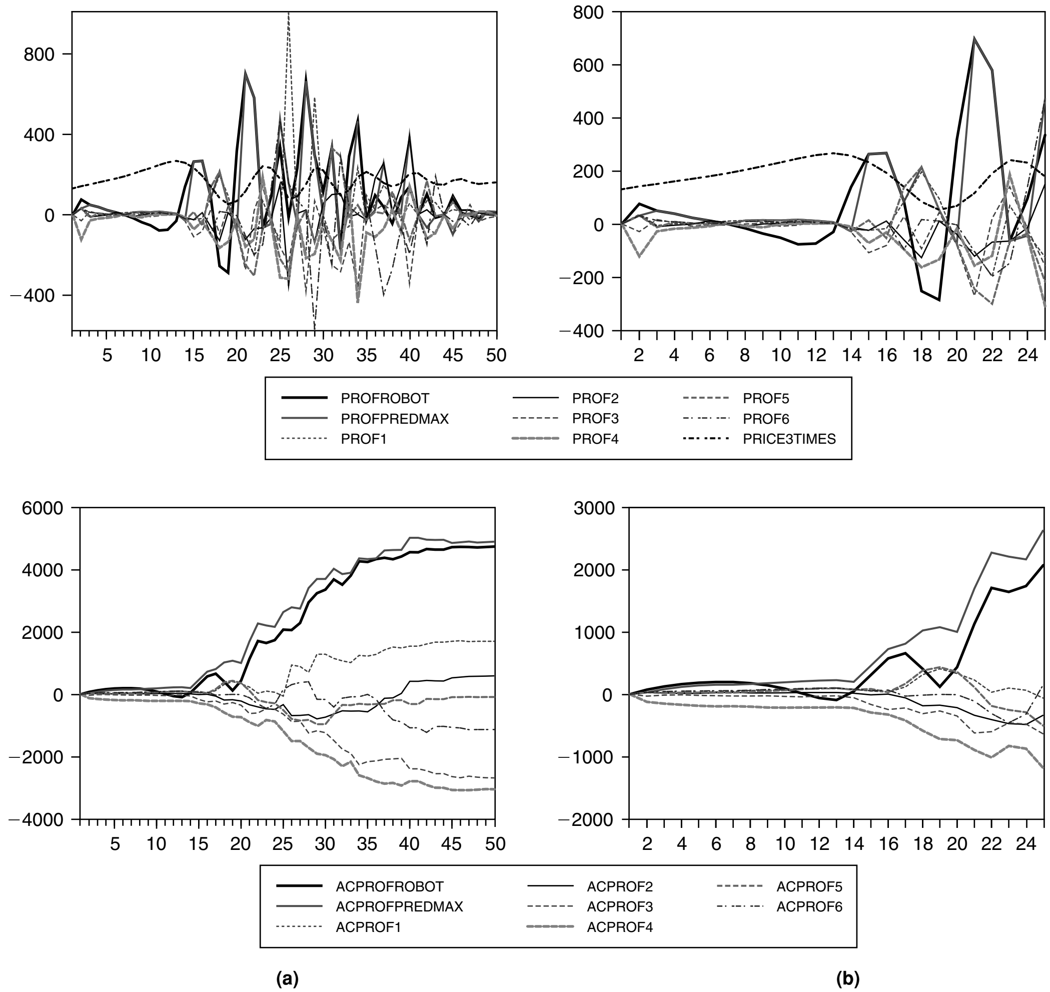

FIGURE 4.16 Realized (nonrisk adjusted) profits (top panel) and accumulated profits (bottom panel) for group 4 (see Figure 4.14) with a robot trader and price upper bound 100. The graphs in (a) show periods 1–50; (b) graphs zoom in to periods 1–25; the realized price series (scaled by a factor of 3) is also shown. The top graphs show the realized profits corresponding to the six individual forecasting strategies, the fundamental robot trader, and a hypothetical switching strategy using the best (according to last period's realized profits) of the other seven strategies. The bottom graphs show the accumulated profits of these eight strategies. Realized profits exhibit large fluctuations over time, and all strategies at times suffer from large losses. On average and in terms of accumulated profits, the fundamental strategy performs very well but is beaten by the switching strategy.

realized profit) out of the seven other strategies. Stated differently, the switching strategy always uses the forecast of the advisor whose pension fund generated the highest realized profit in the previous period. Such a switching strategy beats the fundamental robot strategy.

4.7. CONCLUSION

This chapter has reviewed some behavioral finance models with evolutionary selection of heterogeneous trading strategies and discussed their empirical and experimental validity. When strategy selection is driven by short-run realized profits, trend-following strategies may destabilize asset markets. Asset price fluctuations are characterized by phases in which fundamentalists dominate and prices are close to fundamentals, suddenly interrupted by possibly long-lasting phases of price bubbles when trend-following strategies dominate the market and prices deviate persistently from fundamentals. Even in simple heterogeneous belief models, asset prices are difficult to predict and market timing based on the prediction of the start or the collapse of a bubble is extremely difficult and highly sensitive to noise. Estimation of simple versions of heterogeneous agent models on yearly S&P 500 data suggests that stock prices are characterized by behavioral heterogeneity. Simple evolutionary models therefore could provide an explanation of, for example, the dot-com bubble as being triggered by good news about economic fundamentals and subsequently strongly amplified by trend-following trading strategies. Laboratory experiments using human subjects confirm that coordination on simple trend-following strategies may arise in asset markets and cause persistent deviations from fundamentals.

In a heterogeneous beliefs asset-pricing model, as long as prices fluctuate around their fundamental values, fundamentalist strategies do quite well in terms of accumulated profits. If there are no limits to arbitrage and fundamentalists can survive possibly long-lasting bubbles during which they suffer large losses, their strategy performs very well in the long run and may help stabilize markets. However, fundamentalists can be beaten by a switching strategy based on recent realized profits, thus providing an incentive for investors to keep switching strategies. It should be noted that the models discussed here are very stylized, with a well-defined fundamental price. In real markets, there may be a lot of disagreement about the "correct" fundamental price, and it may then not be so clear what the fundamental strategy would be. Limits to arbitrage may also prevent fundamentalists from holding long-lasting positions opposite to the trend as more and more traders go with the trend based on their recent success.

Most behavioral asset pricing models focus on a single risky asset. Only a few extensions to a multiasset setting have been made until now. Westerhoff (2004) and Chiarella et al. (2007) considered multi-asset markets, where chartists can switch their investments between different markets for risky assets. The interaction between the different markets causes complex asset price dynamics, with different markets exhibiting comovements as well as clustered volatility and fat tails of asset returns. Böhm and Wenzelburger (2005) apply random dynamical systems to investigate the performance of efficient portfolios in a multi-asset market with heterogeneous investors.

Work on complex evolutionary systems in finance is rapidly growing, but little work has been done on policy implications. The most important difference with a representative rational agent framework is probably that, in a heterogeneous, boundedly rational world, asset price fluctuations exhibit excess volatility. If this is indeed the case, it has important implications concerning, for example, the debates on whether a Tobin tax on financial transactions is desirable or whether financial regulation is desirable. Westerhoff and Dieci (2006) use a complex evolutionary system to investigate the effectiveness of a Tobin tax. Investors can invest in two different speculative asset markets. If a Tobin tax is imposed on one market, it is stabilized, while the other market is destabilized; if a tax is imposed on both markets, price fluctuations in both markets decrease.

Brock, Hommes, and Wagener (2008) study the effects of financial innovation on price volatility and welfare. They extend the asset-pricing model with heterogeneous beliefs in Section 4.2 by introducing hedging instruments in the form of Arrow securities, that is, state contingent claims to uncertain future events. They show that more hedging instruments may destabilize markets and decrease welfare when agents are boundedly rational and choose investment strategies based on reinforcement learning. The intuition of this result is simple: Optimistic and/or pessimistic traders take larger positions when they can hedge more risk, and those who happen to be on the right side of the market will be reinforced more. In a world of bounded rationality and learning from past success, more hedging instruments may thus lead to more persistent deviations from market fundamentals. Developing a theory of complex multi-asset market models with heterogeneous interacting trading strategies and empirical and experimental testing will be an important area of research for years to come. From a practitioner viewpoint, this kind of research seems highly relevant to gain more insight into the causes of financial crises, such as the current credit crisis, in order to hopefully avoid such crises in the future.

APPENDIX 4.1: BIFURCATION THEORY

The purpose of this appendix is to show how stability loss of a stable steady state is necessarily connected to one of a small number of standard bifurcations. For this, the first section sketched how a system at stability loss can be reduced to a one- or two-dimensional system. The second section presents the most common bifurcation scenarios: saddle node, period doubling, Hopf, and pitchfork. See Kuznetsov (1995) for a detailed mathematical treatment of bifurcation theory.

A4.1.1. Basic Concepts from Dynamical Systems

Instead of directly dealing with evolution equations of the form:

$$x_t = \varphi(x_{t-1}, \ldots, x_{t-n}) \quad \textbf{(4.72)}$$

dynamical system theory usually considers first-order vector dynamics:

$$y_t = \Phi(y_{t-1}) \quad \textbf{(4.73)}$$

where $y_t = (x_t, \ldots, x_{t-(n-1)})$ and where the system map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is given as:

$$\Phi(y_t) = \begin{pmatrix} \varphi(x_{t-1}, \ldots, x_{t-n}) \\ x_{t-1} \\ \vdots \\ x_{t-(n-1)} \end{pmatrix} \quad \textbf{(4.74)}$$

Given an initial state y_0, the orbit of Φ through y_0 is the sequence $\{y_t\}_{t=0}^{\infty}$, satisfying Eq. 4.73 for every t. For instance, a steady state $\bar{x}$ of the evolution equation, which satisfies $\bar{x} = \varphi(\bar{x}, \ldots, \bar{x})$, corresponds to the constant orbit $y_t = (\bar{x}, \ldots, \bar{x})$ for all t of Φ. Such an orbit is called a *fixed point* of the dynamical system Φ.

In this appendix we discuss only stability changes of fixed points (but stability changes of periodic points can be handled similarly). By definition, a fixed point $\bar{y}$ is *asymptotically stable* if for all y_0 sufficiently close to $\bar{y}$, the orbit of Φ through y_0 tends to $\bar{y}$ as $t \to \infty$. A fixed point $\bar{y}$ is *unstable* if arbitrarily close to it there are initial points y_0, the orbits of which do not tend to $\bar{y}$ as $t \to \infty$.

Let us assume, for simplicity, that $y = 0$ is a fixed point of Φ. The *linearization* of the dynamics of Eq. 4.73 at $y = 0$ is given as:

$$y_t = Ly_{t-1} \quad \textbf{(4.75)}$$

where $L = D\Phi(0)$ is the $n \times n$ Jacobi matrix of Φ at $\bar{y} = 0$. We have this theorem.

Theorem 4.1. *If all eigenvalues λ_j, $j = 1, \ldots, n$, of L satisfy $|\lambda_j| < 1$, then the fixed point $y = 0$ is asymptotically stable. If there is at least one eigenvalue such that $|\lambda_j| > 1$, then $y = 0$ is unstable.*

In general, a fixed point $\bar{y}$ is called *hyperbolic* if no eigenvalue λ of L is on the complex unit circle, that is, $|\lambda| \neq 1$. Note that the stability of a hyperbolic fixed point can be determined just by looking at the eigenvalues.

A *bifurcation* of a system is a qualitative change in the orbit structure as a system parameter is changed. In this appendix, we focus on the simplest kind of bifurcations: the possibilities of a fixed point to lose stability as a one-dimensional parameter is varied.

That is, we consider the case that the system map $\Phi_\mu : \mathbb{R}^n \to \mathbb{R}^n$ depends on a parameter $\mu \in [\mu_1, \mu_2]$. Let us assume that $\bar{y}_1$ is an attracting hyperbolic fixed point of the map Φ_{μ_1}. Note that the fixed point equation:

$$\Phi_\mu(y) - y = 0 \qquad \textbf{(4.76)}$$

can be solved for y as a function of μ whenever $\det(D\Phi_\mu(y) - I) \neq 0$, that is, whenever $\lambda = 1$ is not an eigenvalue of $D\Phi_\mu(y)$. Consequently, we can find a parametrized family of fixed points $\bar{y}_\mu$ such that $\bar{y}_{\mu_1} = \bar{y}_1$, and we can investigate the stability of $\bar{y}_\mu$ as μ varies.

The eigenvalues $\lambda_1, \ldots, \lambda_n$ of $L_\mu = D\Phi_\mu(\bar{y}_\mu)$ are continuous functions of μ; by assumption $|\lambda_j(\mu_1)| < 1$ for all j. Theorem 4.1 implies that stability changes can occur only if for some $\mu_0 \in (\mu_1, \mu_2)$ the fixed point fails to be hyperbolic; that is, if one of the eigenvalues, say $\lambda_1(\mu_0)$, has absolute value 1.

There are three main mechanisms via which a fixed point can fail to be hyperbolic: the saddle-node bifurcation $\lambda_1(\mu_0) = 1$; the period-doubling bifurcation $\lambda_1(\mu_0) = -1$; and the Hopf bifurcation $\lambda_1(\mu_0) = \mathrm{e}^{\mathrm{i}\alpha}$ with $0 < \alpha < \pi$. In the last case, there are two eigenvalues of absolute value equal to 1, since the complex conjugate $\mathrm{e}^{-\mathrm{i}\alpha}$ is necessarily an eigenvalue as well. Of course, it could happen that several eigenvalues have absolute value equal to 1 simultaneously in configurations other than those listed, but it can be shown that these cases are atypical for one-parameter systems.[29]

Here we analyze these three bifurcations one by one. The first step in each analysis is to simplify the system by restricting it to an invariant *center manifold*. Recall that a set W is invariant under Φ if $\Phi(W) = W$.

Theorem 4.2. (center manifold theorem). *Let y_μ be a fixed point of Φ_μ that is nonhyperbolic if $\mu = \mu_0$. For μ sufficiently close to μ_0, there is a family of invariant manifolds W^c_μ, depending differentially on μ, such that $W^c_{\mu_0}$ is tangent to the eigenspace associated to the nonhyperbolic eigenvalues at $\mu = \mu_0$.*

The complete complexity of a bifurcation is retained if the map Φ_μ is resticted to the center manifold W^c_μ.

[29]These configurations are *nonpersistent*: They can be removed by making an arbitrarily small change to the system Φ_μ.

APPENDIX 4.2: BIFURCATION SCENARIOS

A4.2.1. The Saddle-Node Bifurcation

At this bifurcation, two fixed points are created or disappear, depending on the direction of the parameter change.

A saddle-node bifurcation occurs in the case that $\lambda_1(\mu_0) = 1$ and a one-dimensional associated eigenspace. After introducing suitable new variables, the restriction of Φ_μ to the associated one-dimensional center manifold takes the form:

$$\Phi_\mu(x) = x + \mu - x^2 + x^3 g(x, \mu) \tag{4.77}$$

where g is some differentiable function. In fact, it is sufficient to consider the *normal form*:

$$\Phi_\mu^{\mathrm{NF}}(x) = x + \mu - x^2 \tag{4.78}$$

This has to be justified afterward by arguing that the full family Φ_μ has qualitatively the same dynamics. This latter step is not hard but rather technical and is therefore omitted.

Let us analyze the normal form. Note that for $\mu = 0$, the system has a single fixed point $x = 0$ and $L = D\Phi_0^{\mathrm{NF}}(0) = 1$, so we are indeed in the saddle-node case. For general μ, fixed points are the solutions of the equation $x = \Phi_\mu^{\mathrm{NF}}(x)$; for $\mu \geq 0$ they are given as:

$$\bar{x}_1 = \sqrt{\mu} \qquad \bar{x}_2 = -\sqrt{\mu} \tag{4.79}$$

whereas for $\mu < 0$ no fixed points exist.

The linearizations $L_j = D\Phi^{\mathrm{NF}}(\bar{x}_j) = 1 - 2\bar{x}_j$ read as:

$$L_1 = 1 + 2\sqrt{\mu} \qquad L_2 = 1 - 2\sqrt{\mu} \tag{4.80}$$

We conclude that $\bar{x}_1$ is attracting and that $\bar{x}_2$ is repelling.

All this information can be summarized in a bifurcation diagram; see Figure A4.2.1. In the diagram, the location of the fixed points is plotted as a function of the parameter μ; the branch of stable fixed points is indicated by the solid curve, whereas the dashed curve indicates the location of unstable fixed points. At $\mu = 0$, the two branches meet and there is exactly one fixed point; for $\mu < 0$, there are no fixed points.

A4.2.2. The Period-Doubling Bifurcation

In this bifurcation, an attracting fixed point loses stability, and a period-two orbit is generated.

A period-doubling bifurcation occurs if $\lambda_1(\mu_0) = -1$. The eigenspace and the corresponding center manifold are again one-dimensional. The normal form reads in this

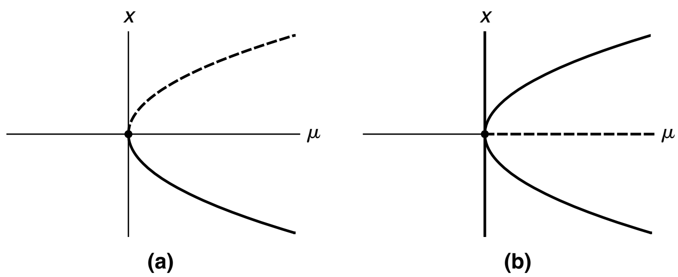

FIGURE A4.2.1 Saddle-node (a) and period-doubling/pitchfork (b) bifurcation diagrams.

case as:

$$\Phi_\mu^{\mathrm{NF}}(x) = -x - \mu x + ax^3 \tag{4.81}$$

with $a = \pm 1$. We only consider the supercritical case $a = 1$, which occurs if the fixed point is globally attracting under Φ_μ for $\mu < 0$. Note that $\bar{x} = 0$ is now a fixed point for all values of μ; moreover,

$$L = D\Phi_\mu^{\mathrm{NF}}(0) = -1 - \mu \tag{4.82}$$

We conclude that $\bar{x}$ is attracting for $-2 < \mu < 0$ and repelling for $\mu > 0$ or $\mu < -2$, implying in particular that the fixed point loses its stability at $\mu = 0$. Moreover, note that:

$$|\Phi^{\mathrm{NF}}(|x|)| = (1 + \mu - |x|^2)|x| = |x| \tag{4.83}$$

if $|x| = \sqrt{\mu}$. This implies that the set $P = \{\sqrt{\mu}, -\sqrt{\mu}\}$ is invariant. As the points in this set are not fixed points, necessarily one is mapped to the other by the map Φ^{NF}.

The set P is a so-called periodic orbit, and its elements are period-2 points. In general, a period-m point is a point that is mapped to itself after m iterations of the system map. It can be verified that nearby orbits are attracted to the period-2 orbit P as $t \to \infty$; the period-2 orbit P is asymptotically stable.

A4.2.3. The Hopf Bifurcation

In a Hopf (or Neimark-Sacker) bifurcation (Figure A4.2.2), a fixed point loses stability and an invariant circle is generated. The bifurcation occurs if $\lambda_1(\mu_0) = \mathrm{e}^{\mathrm{i}\alpha}$ and $\lambda_2(\mu_0) = \mathrm{e}^{-\mathrm{i}\alpha}$, $\alpha \in (0, \pi)\backslash\{\frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}\}$; the associated center manifold is two-dimensional. Again restricting to the supercritical case, the normal form can be expressed as:

$$\Phi^{\mathrm{NF}}(x_1, x_2) = (1 + \mu - x_1^2 - x_2^2)\begin{pmatrix}\cos\vartheta(x) & -\sin\vartheta(x)\\ \sin\vartheta(x) & \cos\vartheta(x)\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \tag{4.84}$$

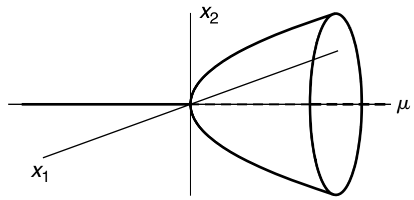

FIGURE A4.2.2 Hopf bifurcation diagram.

where $\vartheta(x) = \alpha + \beta(x_1^2 + x_2^2)$. By introducing polar coordinates $x_1 = r\cos\psi$, $x_2 = r\sin\psi$, the map takes the form:

$$\Phi^{NF}(r,\psi) = \begin{pmatrix} (1+\mu)r - r^3 \\ \psi + \alpha + \beta r^2 \end{pmatrix} \quad \textbf{(4.85)}$$

As in the case of the period-doubling bifurcation, the origin $r = 0$ is stable if $-2 < \mu < 0$ and unstable if $\mu > 0$ (or $\mu < -2$). Moreover, for $\mu > 0$ there is a stable invariant circle:

$$C = \{(r,\psi) : r = \sqrt{\mu}\} \quad \textbf{(4.86)}$$

That is, as the stable fixed point $r = 0$ loses stability, a stable invariant circle branches off. The dynamics on the circle are given as:

$$\psi \mapsto \psi + \alpha + \beta\mu \mod 2\pi \quad \textbf{(4.87)}$$

In the case of the Hopf bifurcation, investigating which properties of the normal form carry over to the full normal form is a nontrivial problem.

A4.2.4. The Pitchfork Bifurcation

We have described the three simplest typical bifurcations of general systems. Sometimes a system has a special symmetry, like a reflection symmetry: $\Phi(-x) = -\Phi(x)$. In the space of such systems, bifurcations that are nontypical for general systems may become typical. An example is the pitchfork bifurcation, which is typical for systems with reflection symmetry.

The normal form for a (supercritical) pitchfork bifurcation reads as:

$$\Phi(x) = (1+\mu)x - x^3 \quad \textbf{(4.88)}$$

Note that $\Phi(-x) = -\Phi(x)$ and that as a consequence the point $x = 0$ is a fixed point for all μ, stable if $-2 < \mu < 0$, unstable otherwise. Moreover, for $\mu > 0$, the points $x = \pm\sqrt{\mu}$ are stable fixed points, branching off the fixed point $x = 0$ as that point loses stability.

Note also that the bifurcation diagram of the pitchfork bifurcation is identical to that of the period-doubling bifurcation but that the interpretation of the branches is different. In a (supercritical) pitchfork bifurcation one steady state changes stability while two new (stable) steady states are created.

References

Adam, K., Experimental evidence on the persistence of output and inflation, *Economic Journal*, 2007, 117, 603–636.

Alchian, A. A., Uncertainty, evolution, and economic theory, *Journal of Political Economy*, 1950, 58, 211–221.

Alfarno, S., T. Lux, and F. Wagner, Estimation of agent-based models: the case of an asymmetric herding model, *Computational Economics*, 2005, 26, 19–49.

Allen, H., and M. P. Taylor, Charts, noise and fundamentals in the London foreign exchange market, *Economic Journal*, 1990, 100, No. 400, Conference Papers, 49–59.

Amir, R., I. V. Evstigneev, T. Hens, and K. R. Schenk-Hoppé, Market selection and survival of investment strategies, *Journal of Mathematical Economics*, 2005, 41, 105–122.

Anderson, P. W., K. J. Arrow, and D. Pines (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, 1988.

Anderson, S., A. de Palma, and J. Thisse, Discrete choice theory of product differentiation, MIT Press, 1993.

Anufriev, M., Wealth-driven competition in a speculative financial market: Examples with maximizing agents, *Quantitative Finance*, 2008, 8, 363–380.

Anufriev, M., T. Assenza, C. H. Hommes, and D. Massaro, Interest rate rules with heterogeneous, expectations, *CeNDEF Working Paper*, August 2008.

Anufriev, M., G. Bottazzi, and F. Pancotto, Equilibria, stability and asymptotic dominace in a speculative market with heterogeneous agents, *Journal of Economic Dynamics and Control*, 2006, 30, 1787–1835.

Anufriev, M., and C. H. Hommes, Evolution of market heuristics, *Knowledge Engineering Review*, 2008, forthcoming.

Arthur, W. B., Complexity in economic and financial markets, *Complexity*, 1995, 1, 20–25.

Arthur, W. B., S. N. Durlauf, and D. A. Lane (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, 1997a.

Arthur, W. B., J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler, Asset pricing under endogenous expectations in an artificial stock market, in W. Arthur, D. Lane, and S. Durlauf (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, pp. 15–44, 1997b.

Baak, S. J., Tests for bounded rationality with a linear dynamic model distorted by heterogeneous expectations, *Journal of Economic Dynamics and Control*, 1999, 23, 1517–1543.

Barberis, N., and R. Thaler, A survey of behavioral finance, in G. M. Constantinidis, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*, Elsevier, pp. 1051–1121, 2003.

Barberis, N., A. Shleifer, and R. Vishny, A model of investor sentiment, *Journal of Financial Economics*, 1998, 49, 307–343.

Benartzi, S., and R. H. Thaler, Heuristics and biases in retirement savings behavior, *Journal of Economic Perspectives*, 2007, 21, 81–104.

Blume, L., and D. Easley, Evolution and market behavior, *Journal of Economic Theory*, 1992, 58, 9–40.

Blume, L., and D. Easley, If you're so smart, why aren't you rich?: Belief selection in complete and incomplete markets, *Econometrica*, 2006, 74, 929–966.

Böhm, V., and J. Wenzelburger, On the performance of efficient portfolios, *Journal of Economic Dynamics and Control*, 2005, 29, 721–740.

Boswijk, H. P., C. H. Hommes, and S. Manzan, Behavioral heterogeneity in stock prices, *Journal of Economic Dynamics and Control*, 2007, 31, 1938–1970.

Branch, W. A., The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations, *Economic Journal*, 2004, 114, 592–621.

Branch, W. A., and G. W. Evans, Intrinsic heterogeneity in expectation formation, *Journal of Economic Theory*, 2006, 127, 264–295.

Branch, W. A., and McGough, A new Keynesian model with heterogeneous expectations, working paper, University of California, Irvine, 2008.

Brock, W. A., Pathways to randomness in the economy: Emergent nonlinearity and chaos in economics and finance, *Estudios Económicos*, 1993, 8, 3–55.

Brock, W. A., Asset price behavior in complex environments, in W. B. Arthur, S. N. Durlauf, and D. A. Lane (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, 385–423, 1997.

Brock, W. A., and C. H. Hommes, A rational route to randomness, *Econometrica*, 1997a, 65, 1059–1095.

Brock, W. A., and C. H. Hommes, Models of complexity in economics and finance, in C. Hey, et al. (eds.), *System Dynamics in Economic and Financial Models*, Wiley, pp. 3–41, 1997b.

Brock, W. A., and C. H. Hommes, Heterogeneous beliefs and routes to chaos in a simple asset pricing model, *Journal of Economic Dynamics and Control*, 1998, 22, 1235–1274.

Brock, W. A., and C. H. Hommes, Rational animal spirits, in P. J. J. Herings, G. van der Laan, and A. J. J. Talman (eds.), *The Theory of Markets*, North-Holland, 109–137, 1999.

Brock, W. A., C. H. Hommes, and F. O. O. Wagener, Evolutionary dynamics in markets with many trader types, *Journal of Mathematical Economics*, 2005, 41, 7–42.

Brock, W. A., C. H. Hommes, and F. O. O. Wagener, More hedging instruments may destabilize markets, CeNDEF working paper, University of Amsterdam, 2008.

Brock, W. A., and B. LeBaron, A structural model for stock return volatility and trading volume, *Review of Economics and Statistics*, 1996, 78, 94–110.

Brooks, C., and A. Katsaris, A three-regime model of speculative behaviour: Modeling the evolution of the S&P 500 Composite Index, *Economic Journal*, 2005, 115, 767–797.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.

Campbell, J. Y., and R. J. Shiller, Valuation ratios and the long-run stock market outlook: An update, in R. H. Thaler (ed.), *Advances in Behavioural Finance*, Vol. 2, pp. 173–201, Princeton University Press, 2005.

Chang, S. K., A simple asset pricing model with social interactions and heterogeneous beliefs, *Journal of Economic Dynamics and Control*, 2007, 31, 1300–1325.

Chavas, J. P., On information and market dynamics: The case of the U.S. beef market, *Journal of Economic Dynamics and Control*, 2000, 24, 833–853.

Chevalier, J., and G. Ellison, Risk-taking by mutual funds as a response to incentives, *Journal of Political Economy*, 1997, 105, 1167–1200.

Chiarella, C., R. Dieci, and X. He, Heterogeneity, market mechanisms, and asset price dynamics, *Handbook of Financial Markets: Dynamics and Evolution,* 2009.

Chiarella, C., and X. He, Asset price and wealth dynamics under heterogeneous expectations, *Quantitative Finance*, 2001, 1, 509–526.

Chiarella, C., and X. He, Heterogeneous beliefs, risk and learning in a simple asset pricing model, *Computational Economics*, 2002, 19, 95–132.

Chiarella, C., and X. He, Heterogeneous beliefs, risk and learning in a simple asset pricing model with a market maker, *Macroeconomic Dynamics*, 2003, 7, 503–536.

Chiarella, C., R. Dieci, and L. Gardini, Speculative behaviour and complex asset price dynamics: A global analysis, *Journal of Economic Behavior and Organization*, 2002, 49, 173–197.

Chiarella, C., R. Dieci, and L. Gardini, Asset price and wealth dynamics in a financial market with heterogeneous agents, *Journal of Economic Dynamics and Control*, 2006, 30, 1755–1786.

Chiarella, C., R. Dieci, and X.-Z. He, Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework, *Journal of Economic Behavior and Organization*, 2007, 62, 408–427.

Cont, R., and J.-P. Bouchaud, Herd behavior and aggregate fluctuations in financial markets, *Macroeconomic, Dynamics* 2000, 4, 170–196.

Cutler, D. M., J. M. Poterba, and L. H. Summers, What moves stock prices?, *Journal of Portfolio Management*, 1989, 15, 4–12.

DeBondt, W. F. M., and R. H. Thaler, Does the stock market overreact? *Journal of Finance*, 1985, 40, 793–805.

DeGrauwe, P., H. Dewachter, and M. Embrechts, *Exchange Rate Theory: Chaotic Models of Foreign Exchange Markets*, Blackwell, 1993.

DeGrauwe, P., and M. Grimaldi, Heterogeneity of agents, transaction costs and the exchange rate, *Journal of Economic Dynamics and Control*, 2005, 29, 691–719.

DeGrauwe, P., and M. Grimaldi, Exchange rate puzzles: A tale of switching attractors, *European Economic Review*, 2006, 50, 1–33.

Del Guercio, D. D., and P. A. Tkac, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis*, 2002, 37, 523–557.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, Noise trader risk in financial markets, *Journal of Political Economy*, 1990a, 98, 703–738.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance*, 1990b, 45, 379–395.

Diks, C. G. H., C. H. Hommes, V. Panchenko, and R. van der Weide, E&F Chaos: A user friendly software package for nonlinear economic dynamics, *Computational Economics*, 2008, 32, 221–244.

Diks, C. G. H., and R. van der Weide, Heterogeneity as a natural source of randomness, CeNDEF working paper 03-05, University of Amsterdam, 2003.

Diks, C. G. H., and R. van der Weide, Herding, a-synchronous updating and heterogeneity in memory in a CBS, *Journal of Economic Dynamics and Control*, 2005, 29, 741–763.

Duffy, J., Agent-based models and human subject experiments, in L. Tesfatsion and K. L. Judd (eds.), *Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics*, Handbooks in Economics Series, Elsevier, 949–1011, 2006.

Duffy, J., Experimental macroeconomics, in L. Blume, S. Durlauf (eds.), *New Palgrave Dictionary of Economics*, Macmillan, 2008.

Dwyer, G. P., A. W. Williams, R. C. Battalio, T. I. Mason, Tests of rational expectations in a stark setting, *Economic Journal*, 1993, 103, 586–601.

Evstigneev, I., T. Hens, K. R. Schenk-Hoppé, Market selection of financial trading strategies: global stability, *Mathematical Finance*, 2002, 12, 329–339.

Evstigneev, I., T. Hens, K. R. Schenk-Hoppé, Globally evolutionary stable portfolio rules, *Journal of Economic Theory*, 2008.

Evstigneev, I., T. Hens, K. R. Schenk-Hoppé, Evolutionary finance, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Fama, E. F., Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, 1970, 25, 383–423.

Fama, E. F., and K. R. French, The equity premium, *Journal of Finance*, 2002, 57, 637–659.

Farmer, J. D., Market force, ecology, and evolution, *Industrial and Corporate Change*, 2002, 11, 895–953.

Farmer, J. D., and S. Joshi, The price dynamics of common trading strategies, *Journal of Economic Behavior and Organization*, 2002, 49, 149–171.

Fisher, K. L., and M. Statman, Blowing bubbles, *Journal of Psychology and Financial Markets*, 2002, 3, 53–65.

Föllmer, H., U. Horst, and A. Kirman, Equilibria in financial markets with heterogeneous agents: A probabilistic perspective, *Journal of Mathematical Economics*, 2005, 41, 123–155.

de Fontnouvelle, P., Information dynamics in financial markets, *Macroeconomic Dynamics*, 2000, 4, 139–169.

Frankel, J. A., and K. A. Froot, Using survey data to test standard propositions regarding exchange rate expectations, *American Economic Review*, 1987a, 77, 133–153.

Frankel, J. A., and K. A. Froot, Short-term and long-term expectations of the yen/dollar exchange rate: Evidence from survey data, *Journal of the Japanese and International Economies*, 1987b, 1, 249–274; also published as NBER working paper 2216, April 1987.

Frankel, J. A., and K. A. Froot, Chartists, fundamentalists and the demand for dollars, in A. S. Courakis, and M. P. Taylor (eds.), *Private Behaviour and Government Policy in Interdependent Economies*, Oxford University Press, pp. 73–126 (also published as NBER working paper no. r1655, October 1991), 1990a.

Frankel, J. A., and K. A. Froot, The rationality of the foreign exchange rate: Chartists, fundamentalists and trading in the foreign exchange market, *American Economic Review*, 1990b, 80, no. 2, AEA Papers and Proceedings, 181–185.

Friedman, M., The case of flexible exchange rates, in *Essays in Positive Economics*, University of Chicago Press, 1953.

Gaunersdorfer, A., Endogenous fluctuations in a simple asset-pricing model with heterogeneous beliefs, *Journal of Economic Dynamics and Control*, 2000, 24, 799–831.

Gaunersdorfer, A., and C. H. Hommes, A nonlinear structural model for volatility clustering, forthcoming in A. Kirman, and G. Teyssière (eds.), *Microeconomic Models for Long Memory in Economics*, Springer-Verlag, pp. 265–288, 2007.

Gaunersdorfer, A., C. H. Hommes, F. O. O. Wagener, Bifurcation routes to volatility clustering under evolutionary learning, *Journal of Economic Behavior and Organization*, 2008, 67, 27–47.

Gerber, A., T. Hens, B. Vogt, Rational investor sentiment, working paper 126, *Institute for Empirical Research in Economics*, University of Zurich, 2002.

Gervais, S., and T. Odean, Learning to be overconfident, *Review of Financial Studies*, 2001, 14, 1–27.

Gilli, M., and P. Winker, A global optimization heuristic for estimating agent-based models, *Computational Statistics and Data Analysis*, 2003, 42, 299–312.

Goldbaum, D., Market efficiency and learning in an endogenously unstable environment, *Journal of Economic Dynamics and Control*, 2005, 29, 953–978.

Gordon, M., The investment financing and valuation of the corporation, Irwin, 1962.

Heemeijer, P., Expectation formation in dynamic market models, Ph-D thesis, Tinbergen Institute Research Series, University of Amsterdam.

Hens, T., and K. R. Schenk-Hoppé, Evolutionary stability of portfolio rules in incomplete markets, *Journal of Mathematical Economics*, 2005, 41, 43–66.

Hey, J. D., Expectations formation: Rational or adaptive or ...?, *Journal of Economic Behavior and Organization*, 1994, 25, 329–349.

Hirschleifer, D., Investor psychology and asset pricing, *Journal of Finance*, 2001, 56, 1533–1597.

Hommes, C. H., Financial markets as nonlinear adaptive evolutionary systems, *Quantitative Finance*, 2001, 1 149–167.

Hommes, C. H., Modeling the stylized facts in finance through simple nonlinear adaptive systems, *Proceedings of the National Academy of Sciences*, 2002, 99, 7221–7228.

Hommes, C. H., Heterogeneous agent models in economics and finance, in L. Tesfatsion, and K. J. Judd (eds.), *Handbook of Computational Economics*, *Vol. 2: Agent-Based Computational Economics*, Elsevier, pp. 1109–1186, 2006.

Hommes, C. H., H. Huang, and D. Wang, A robust rational route to randomness in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 2005, 29, 1043–1072.

Hommes, C. H., J. Sonnemans, J. Tuinstra, and H. van de Velden, Coordination of expectations in asset-pricing experiments, *Review of Financial Studies*, 2005, 18, 955–980.

Hommes, C. H., J. Sonnemans, J. Tuinstra, and H. van de Velden, Expectations and bubbles in asset-pricing experiments, *Journal of Economic Behavior and Organization*, 2008, 67, 116–133.

Hong, H., and J. Stein, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance*, 1999, 55, 265–295.

Hong, H., and J. Stein, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies*, 2003, 16, 487–525.

Iori, G., A microsimulation of traders activity in the stock market: the role of heterogeneity, agents' interactions and trade frictions, *Journal of Economic Behavior and Organization*, 2002, 49, 269–285.

Ippolito, R. A., Consumer reaction to measures of poor quality: Evidence from the mutual fund industry, *Journal of Law and Economics*, 1992, 35, 45–70.

Ito, K., Foreign exchange rate expectations, *American Economic Review*, 1990, 80, 434–449.

Kahneman, D., Maps of bounded rationality: Psychology for behavioral economics, *American Economic Review*, 2003, 93, 1449–1475.

Kahneman, D., and A. Tversky, On the psychology of prediction, *Psychological Review*, 1973, 80, 237–251.

Kantz, H., and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.

Karceski, J., Returns-chasing behavior, mutual funds and beta's death, *Journal of Financial and Quantitative Analysis*, 2002, 37, 559–594.

Kelley, H., and D. Friedman, Learning to forecast price, *Economic Enquiry*, 2002, 40, 556–573.

Kirman, A., Epidemics of opinion and speculative bubbles in Financial Markets, in M. Taylor (ed.), *Money and Financial Markets*, Macmillan, 1991.

Kirman, A., Ants, rationality and recruitment, *Quarterly Journal of Economics*, 1993, 108, 137–156.

Kirman, A., and G. Teyssière, Microeconomic models for long memory in the volatility of financial time series, *Studies in Nonlinear Dynamics and Econometrics*, 2002, 5, 281–302.

Kuznetsov, Y., *Elements of Applied Bifurcation Theory*, Springer Verlag, 1995.

Kyle, A., and W. Xiong, Contagion as a wealth effect, *Journal of Finance*, 2001, 56, 1401–1440.

LeBaron, B., Agent-based computational finance: suggested readings and early research, *Journal of Economic Dynamics and Control*, 2000, 24, 679–702.

LeBaron, B., Agent-based computational finance, in L. Tesfatsion, and K. J. Judd (eds.), *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*, Elsevier, pp. 1187–1232, 2006.

LeBaron, B., W. B. Arthur, and R. Palmer, Time series properties of an artificial stock market, *Journal of Economic Dynamics and Control*, 1999, 23, 1487–1516.

Levy, M., H. Levy, and S. Solomon, A microscopic model of the stock market, *Economics Letters*, 1994, 45, 103–111.

Lucas, R. E., Econometric testing of the natural rate hypothesis, in O. Eckstein (ed.), *The Econometrics of Price Determination*, conference. Board of Governors of the Federal Reserve System and Social Science Research Council, 1971.

Lux, T., Stochastic behavioral asset-pricing models, and the stylized facts, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Lux, T., and M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature*, Vol. 397, February 1999, 498–500.

Lux, T., and M. Marchesi, Volatility clustering in financial markets: a micro-simulation of interacting agents, *International Journal of Theoretical and Applied Finance*, 2000, 3, 675–702.

Manski, C., and D. McFadden (eds.), Structural analysis of discrete data with econometric applications, MIT Press, 1981.

Manzan, S., Essays in nonlinear economic dynamics, *Tinbergen Institute Research Series* 317, University of Amsterdam, 2003.

Marimon, R., S. E. Spear, and S. Sunder, Expectationally driven market volatility: an experimental study, *Journal of Economic Theory*, 1993, 61, 74–103.

Marimon, R., and S. Sunder, Indeterminacy of equilibria in a hyperinflationary world: experimental evidence, *Econometrica*, 1993, 61, 1073–1107.

Marimon, R., and S. Sunder, Expectations and learning under alternative monetary regimes: an experimental approach, *Economic Theory*, 1994, 4, 131–162.

Marimon, R., and S. Sunder, Does a constant money growth rule help stabilize inflation?, *Carnegie-Rochester Conference Series on Public Policy*, 1995, 43, 111–156.

Muth, J. F., Rational expectations and the theory of price movements, *Econometrica*, 1961, 29, 315–335.

Rockinger, M., Determinants of capital flows to mutual funds, working paper, HEC School of Management, 1996.

Scheinkman, J. A., and W. Xiong, Heterogeneous beliefs, speculation and trading in financial markets, in R. A. Carmona, et al. (eds.), Paris–Princeton Lectures in Mathematical Finance 2003, *Lecture Notes in Mathematics*, 2004, 1847, Springer Verlag, pp. 217–250.

Schmalensee, R., An experimental study of expectation formation, *Econometrica*, 1976, 44, 17–41.

Shiller, R. J., Stock prices and social dynamics, *Brookings Papers in Economic Activity*, 1984, 2, 457–510.

Shiller, R. J., Investor behavior in the October 1987 stock market crash: survey evidence, NBER working paper no. 2446, November 1987, in R. J. Shiller, *Market Volatility*, MIT Press, 1989, Chapter 23.

Shiller, R. J., *Market Volatility*, MIT Press, 1989.

Shiller, R. J., Speculative prices and popular models, *Journal of Economic Perspectives*, 1990, 4, 55–65.

Shiller, R. J., Measuring bubble expectations and invester confidence, *Journal of Psychology and Financial Markets*, 2000, 1, 49–60.

Shleifer, A., *Inefficient Markets*, Oxford University Press, 2000.

Shleifer, A., and R. W. Vishny, The limits of arbitrage, *Journal of Finance*, 1997, 52, 35–55.

Sirri, E. R., and P. Tufano, Costly search and mutual fund flows, *Journal of Finance*, 1998, 53, 1589–1622.

Smith, V., G. L. Suchanek, and A. W. Williams, Bubbles, crashes and endogenous expectations in experimental spot asset markets, *Econometrica*, 1998, 56, 1119–1151.

Summers, L. H., Does the stock market rationally reflect fundamental values?, *Journal of Finance*, 1986, 41, 591–602.

Sutan, A., and M. Willinger, Why do we guess better in negative feedback situations? An experiment of beauty contest games with negative feedback and interior equilibria, working paper, University of Montpellier, December 2005.

Taylor, M. P., and H. Allen, The use of technical analysis in the foreign exchange market, *Journal of International Money and Finance*, 1992, 11, 304–314.

Teräsvirta, T., Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association*, 1994, 89, 208–218.

Tesfatsion, L., and K. J. Judd, *Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics*, Elsevier, 2006.

Turnovsky, S. J., Empirical evidence on the formation of price expectations, *Journal of the American Statistical Association*, 1970, 65, 1441–1154.

Tversky, A., and D. Kahneman, Judgment under uncertainty: heuristics and biases, *Science*, 1974, 185, 1124–1131.

van de Velden, H., An experimental approach to expectation formation in dynamic economic systems, Ph.D. thesis, Tinbergen Institute Research Series 268, University of Amsterdam, 2001.

van Norden, S., and H. Schaller, Speculative behavior, regime switching, and stock market crashes, in P. Rothman, (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Kluwer, pp. 321–356, 1999.

Vissing-Jorgensen, A., Perspective on behavioral finance: does 'irrationality' disappear with wealth? Evidence from expectations and actions, in M. Gertler, and K. Rogoff (eds.), *NBER Macroeconomics Annual*, MIT Press, 2003.

Westerhoff, F. H., Multiasset market dynamics, *Macroeconomic Dynamics*, 2004, 8, 596–616.

Westerhoff, F. H., and R. Dieci, The effectiveness of Keynes-Tobin transaction taxes when heterogeneous agents can trade in different markets: a behavioral finance approach, *Journal of Economic Dynamics and Control*, 2006, 30, 293–322.

Westerhoff, F. H., and S. Reitz, Nonlinearities and cyclical behavior: the role of chartists and fundamentalists, *Studies in Nonlinear Dynamics and Econometrics*, 2003, Vol. 7, Issue 4, article 3.

Williams, A. W., The formation of price forecasts in experimental markets, *Journal of Money, Credit and Banking*, 1987, 19, 1–18.

Winker, P., and M. Gilli, Indirect estimation of the parameters of agent based models of financial markets, FAME Research Paper no. 38, University of Geneva, November 2001.

Youssefmir, M., and B. A. Huberman, Clustered volatility in multi agent dynamics, *Journal of Economic Behavior and Organization*, 1997, 32, 101–118.

Zwart, G. J., T. Markwat, L. A. P. Swinkels, and D. van Dijk, Chartists and fundamentalists in emerging markets, working paper, Erasmus School of Economics, 2007.

CHAPTER 5

Heterogeneity, Market Mechanisms, and Asset Price Dynamics

Carl Chiarella
University of Technology, Sydney

Roberto Dieci
University of Bologna

Xue-Zhong He
University of Technology, Sydney

5.1. Introduction 279
5.2. Heterogeneity and Market-Clearing Mechanisms 283
5.2.1. Portfolio Optimization 283
5.2.2. Utility Functions 284
5.2.3. Market-Clearing Mechanisms 285
5.2.4. Noise 287
5.2.5. Expectations Feedback 287
5.3. Price Dynamics Implied by the CARA Utility Function 288
5.3.1. Fundamental Price and the Optimal Demand 288
5.3.2. Formation of Heterogeneous Beliefs 289
5.3.3. Performance Measure and Switching 290
5.3.4. Price Behavior under the Walrasian Auctioneer Mechanism 291
5.3.5. Price Behavior under the Market-Maker Mechanism 298
5.4. Price behavior and Wealth Dynamics Implied by the CRRA Utility 302
5.4.1. Optimal Portfolio and Wealth Dynamics 302

Note: The authors are thankful to participants of Modelli Dinamici in Economia e Finanza (MDEF) 2006 at the University of Urbino for comments on an earlier draft of this work. The authors would also like to thank the anonymous reviewers and the editors for many insightful comments that have helped to greatly improve this chapter. Chiarella and He acknowledge support from the ARC under Discovery Grant Project DP0450526. Dieci acknowledges support from MIUR under the project PRIN-2004137559, "Nonlinear models in economics and finance: interactions, complexity and forecasting."

5.4.2. Price and Wealth Behavior with a Walrasian Auctioneer 303
5.4.3. Price and Wealth Behavior with a Market Maker 306
5.5. Empirical Behavior 314
5.5.1. Stylized Facts in the S&P 500 314
5.5.2. A Market Fraction Model and Its Stylized Behavior 316
5.5.3. Econometric Characterization of the Power-Law Behavior 319
5.6. Heterogeneity in a Dynamic Multiasset Framework 321
5.6.1. Optimization of a Many Risky Asset Portfolio with Heterogeneous Beliefs 322
5.6.2. An Example of Two Risky Assets and Two Beliefs 326
5.7. The Continuous Stochastic Dynamics of Speculative Behavior 330
5.7.1. Stochastic Models with Heterogeneous Beliefs 330
5.7.2. A Continuous Stochastic Model with Fundamentalists and Chartists 331
5.7.3. A Random Dynamical System and Stochastic Bifurcations 332
5.8. Conclusion 338
References 340

Abstract

This chapter surveys the boundedly rational heterogeneous agent (BRHA) models of financial markets, to the development of which the authors and several coauthors have contributed in various papers. We give particular emphasis to the role of the market-clearing mechanism used, the utility function of the investors, the interaction of price and wealth dynamics, portfolio implications, the impact of stochastic elements on market dynamics, and calibration of this class of models. Due to agents' behavioral features and market noise, the BRHA models are both nonlinear and stochastic. We show that the BRHA models produce both a locally stable fundamental equilibrium corresponding to that of the standard paradigm as well as instability with a consequent rich range of possible complex behaviors characterized both indirectly by simulation and directly by stochastic bifurcations. A calibrated model is able to reproduce quite well the stylized facts of financial markets. The BRHA framework is thus able to accommodate market features that seem not easily reconcilable for the standard financial market paradigm, such as fat tails, volatility clustering, large excursions from the fundamental, and bubbles.

Keywords: bounded rationality, interacting heterogeneous agents, behavioral finance, nonlinear economic dynamics, complexity

5.1. INTRODUCTION

One of the main building blocks of the modern theory of finance relies on the paradigm that asserts that asset prices are the outcome of the market interaction of utility-maximizing agents who use rational expectations when forming expectations about future market outcomes. Preferences of agents are assumed to satisfy conditions that enable the mass of investors to be considered as a single representative agent. Since agents rationally impound all relevant information into their trading decisions, the movement of prices is assumed to be perfectly random and hence to exhibit random walk behavior. This view is important in empirical finance because it is the theoretical underpinning of the efficient markets hypothesis and asset-pricing theories generally. It is also the basis of the stochastic price mechanisms assumed in many of the key theoretical models in finance, such as the optimal portfolio rules that have developed out of the work of Markowitz (1952) and Merton (1971); the static and intertemporal capital asset-pricing model of Sharpe (1964), Lintner (1965), Mossin (1966), and Merton (1974); and models for the pricing of contingent claims, beginning with the work of Black and Scholes (1973). The impressive statistical evidence in favor of market efficiency, discussed by, for example, Fama (1976), has been taken as support for the random walk model, and for a long time financial economists were contented with this view as the explanation of the time series behavior of observed asset prices.

A range of empirical studies, however, led to some questioning of the basic tenets of the efficient markets model, or at least of the view that it suggests for asset price dynamics. For a start, there has been a large number of studies reporting various anomalies relating excess returns to a variety of factors such as firm size, leverage, the book value of equity, and earnings/price ratio, to name some of the most widely studied. There is also empirical evidence that stock prices over- or underreact to earnings announcements (see Bernard, 1993), suggesting that information is not being instantaneously impounded into prices, as the efficient-markets paradigm would suggest. Much of the anomalies literature is surveyed by Keim (1988).

A number of authors have carried out so-called volatility tests that basically reveal that market prices have far more volatility than can be ascribed to the underlying fundamentals (supposedly expected future dividends); much of this literature is surveyed by Cochrane (1991), but see also Shiller (1981). It should be pointed out that some authors, such as Marsh and Merton (1986), have criticized volatility tests on methodological grounds. These results have suggested to some authors that asset prices are not only the result of rational investors reacting rationally to the shocks impinging on the market but also contain at least some element of what has come to be called "irrational" behavior, which in this chapter we prefer to call *boundedly rational* behavior. This led to the development of mean-reverting fads models as suggested by Shiller and Perron (1985) and Summers (1986), though these seem unable to account for the departures from the random walk model. According to the efficient markets paradigm, large market movements should be the result of some large news event. But many large market movements, in particular the stock market crash of 1987, some of the "crashettes" during the 1990s, and the large market movement worldwide in February 2007, seem to be not at all related to any specific news event. Certainly the 1987 crash has been well researched, but so far there is no satisfactory explanation of what was the "news event" that triggered such a large market movement.

Significantly, despite the dominance of the random walk/efficient markets paradigm in the academic literature and its espousal by business schools, some major investment institutions continue to devote a considerable amount of valuable resources to technical analysis, as evidenced, for example, in the surveys of Allen and Taylor (1990), Frankel and Froot (1987), and Taylor and Allen (1992). Indeed, there is some evidence that such techniques meet with some measure of success; see, for example, Pruitt and White (1988). From a theoretical perspective there had always been dissatisfaction with the extreme assumptions of rationality and computational power ascribed to the rational economic agent of the standard paradigm, going back at least to the writings of Simon (1997). Furthermore, the concept of the no-trade theorem in efficient markets, popularized by Milgrom and Stokey (1982), seems to stand in stark contrast to the incredible volume of trade observed in financial markets. We should also mention the developments in so-called behavioral finance—see, for example, Stracca (2004)—that question the efficient markets paradigm more from a behavioral or psychological perspective.

As a result of these considerations there developed during the 1990s an alternative paradigm of asset price dynamics that took into consideration the fact that investors in the real world can at best be boundedly rational, have limited computational power, and are heterogeneous with respect to risk preferences and the way they form expectations. There were in fact some important antecedents to this new literature, in particular Zeeman (1974), who postulated two groups of investors, fundamentalists and chartists, and set up the market dynamics within the frameworks of catastrophe theory. Unfortunately, his contribution came at the time that the so-called rational expectations revolution was beginning to take a very strong grip on the mindset of economists. Since Zeeman's model started with the premise that some investors might not be fully rational (in the sense used by rational expectations economists), his contribution was completely overlooked. Furthermore, the fact that the dynamics of Zeeman's model were expressed in the language of catastrophe theory may also have impeded further developments. As we now know, the language of the theory of nonlinear dynamical systems is a far more suitable framework, and the sudden-jump type of behavior in reaction to a smooth change in a parameter observed in catastrophe theory models can be obtained if certain speeds of adjustment go to infinity (so that relaxation cycles result; see Chiarella, 1986). This theory only started its modern phase of development in the late 1970s.

Another important early contribution was that of Beja and Goldman (1980), who also posited a simple differential equation model of fundamentalists and chartists and showed that if the chartists were dominant in some sense, the market would tend to instability and some sort of breakdown. Interestingly enough, this contribution appeared in the pages of the *Journal of Finance*, which some regard as the ultimate bastion of the efficient markets paradigm. Although the Beja and Goldman article found little resonance in the traditional finance literature, not even in the fads literature that developed later, it certainly had an impact on the development of the boundedly rational heterogeneous agent paradigm that we survey here. Frankel and Froot (1990) also developed a fundamentalist and chartist framework to try to explain the long deviation of the U.S. dollar from its supposed fundamental value during the 1980s. These authors also provided some survey evidence of the use of fundamentalism and chartism in foreign exchange markets.

The boundedly rational heterogeneous agent (BRHA) literature can be considered to have properly started with a number of contributions in the early 1990s. Day and Huang (1990) and Chiarella (1992) took the fundamentalist/chartist framework of the earlier cited literature and added to it some type of behavioral nonlinearity (such as an asset-demand function) that resulted in the models exhibiting bounded dynamic behavior when the steady state becomes locally unstable. The contribution by Kirman (1992) was influential in changing the viewpoint about the representative agent. Kirman (1993) further promoted the idea of agents imitating one another, which led to the contributions of Lux (1995, 1998), who took the paradigm much further and in the direction of getting the models to generate the type of return behavior observed in financial markets, in particular fat-tail phenomena. In fact, getting the BRHA models to generate the so-called stylized facts of financial market behavior has become a feature of recent research, as have attempts to estimate such models. The model of Lux and Marchesi (1999) contains fundamentalists and pessimistic as well as optimistic trend followers and allows agents to switch among these groups, depending on the evolution of the market. The model is able to generate periods of extreme volatility and long deviations from the fundamental, without any strong news event triggering such movements. Another influential set of contributions has been that of Brock and Hommes (1997, 1998), who developed and analyzed in detail the concept that investors are choosing from a range of strategies (fundamentalism, pessimistic chartism, optimistic chartism, and so forth) and switch to the more "fit" one (fitness, being, for instance, measured by some weighted average of recent returns) as the market evolves. This idea has been incorporated into many subsequent models.

There have been a number of recent surveys of the BRHA paradigm, most notably those of Hommes (2006) and LeBaron (2006). In this survey we shall focus on the paradigm developed by the authors (and several coauthors) in various papers, giving particular emphasis to the role of the market-clearing mechanism used, the type of utility function of the investors, the interaction of price and wealth dynamics, portfolio implications, the impact of stochastic elements on the markets dynamics, and calibration and estimation of this class of models. The important point to stress is that the model of boundedly rational behavior that we introduce takes the traditional one-period optimization model that is the basis of the CAPM and perturbs it in two essential ways. First, we allow agents to be heterogeneous with respect to their expectations about the distribution of the future returns of risky assets. Agents' expectations will be driven by observations of past returns so that the agents exhibit expectations feedback. This, together with the fact that the one-period optimization is assumed to be continually repeated, gives the models their dynamic behavior. Second, we allow one group of agents, usually known as *fundamentalists*, to have greater knowledge about the economy. These agents have some notion of the fundamental price and perhaps also of the strategies of other agents in the economy.[1]

This chapter unfolds in the following way: In Section 5.2 we give an overview of where we feel the standard paradigm needs to be adjusted and so lay out the basic building blocks of the *boundedly rational heterogeneous agent* (BRHA) framework, including portfolio optimization, agents' utility functions, the market-clearing mechanism, and expectations feedback. When agents have *constant absolute risk aversion* (CARA) utility functions, the dynamics of the wealth process do not feed back into the dynamics governing the price process. Although this separation is convenient from a certain perspective, it has the disadvantage that such models have difficulty in generating the type of growing price process that we observe in reality. When agents have *constant relative risk aversion* (CRRA) utility, the dynamics for price and wealth are intertwined and growing over time and hence follow a nonstationary dynamical process (in the dynamical systems sense of not having a steady state). In Sections 5.3 and 5.4 we

[1] See in particular Wenzelburger (2009).

consider behavior of the price dynamics under two kinds of market-clearing mechanism, the Walrasian auctioneer and the market maker, when agents have CARA and CRRA utility functions, respectively. Section 5.5 reviews the stylized facts of financial markets and we explore whether these can be reproduced by the characteristics of the returns processes of the models of Sections 5.3 and 5.4, when perturbed by a noisy fundamental and market noise, in particular the power-law behavior in returns and volatility clustering. In Section 5.6 we expand the framework of Sections 5.3 and 5.4 to the situation of many risky assets, thus investigating how agent heterogeneity, and in particular their beliefs about covariance between risky assets, affects portfolio diversification. Section 5.7 considers further the issue of the interaction of the bifurcation behavior of a stochastic continuous time model and outlines recent attempts to approach this issue using the tools of stochastic bifurcation theory. Section 5.8 will conclude by pointing out the many unfinished issues in the development of the heterogeneous agent paradigm of financial markets and points to some future research directions.

5.2. HETEROGENEITY AND MARKET-CLEARING MECHANISMS

Although there might be agreement that the standard paradigm does not fully explain what is causing the evolution of speculative asset prices, there may be less agreement on where to start to build an improved paradigm. By and large the view that has been adopted in the BRHA literature is to retain expected utility maximization as the goal of each agent but to allow the agents to have different risk preferences and different expectations, rather than the single homogeneous rational expectation, about future possible returns. The expectations differ since agents are assumed to have different information and beliefs. Thus the asset price-modeling framework that we develop in this section is based on the fact that trading is driven more by differences in expectations than by the random arrival of news events. Within this framework, a simple market of one risky asset and one risk-free asset, with agents having different expectations, is considered with two different types of utility functions and two different market-clearing mechanisms. The framework provides the basic elements and structure of the various models we discuss in the following sections. The basic building blocks of this framework are portfolio optimization, agents' utility functions, the market-clearing mechanism, and expectations feedback. We shall now discuss the role of each of these elements in determining the price map driving the price dynamics.

5.2.1. Portfolio Optimization

We take as our starting point the basic problem of agents allocating their wealth in a market with a risky asset (except Section 5.6, where there are many risky assets) and a risk-free asset. We assume that each agent optimizes one period ahead and applies the resulting optimal investment rule period by period. Let P_t be the risky asset price, y_t the dividend (both at time t), and $R = 1 + r_f$, where r_f is the risk-free interest rate. The asset (proportional) return over the time period $(t, t+1)$ is defined by $r_{t+1} = (P_{t+1} + y_{t+1})/P_t - 1$. For agent i, let $W_{i,t}$ be the wealth, $z_{i,t}$ the holding of the amount of the risky asset, and $\pi_{i,t}$ the proportion of wealth invested in the risky asset at time t, hence $\pi_{i,t} W_{i,t} = z_{i,t} P_t$. Then the wealth dynamics of agent i can be formed either in terms of the number of shares, $z_{i,t}$, as

$$W_{i,t+1} = RW_{i,t} + (P_{t+1} + y_{t+1} - RP_t) z_{i,t} \tag{5.1}$$

or in terms of the wealth proportion, $\pi_{i,t}$, according to

$$W_{i,t+1} = W_{i,t}[R + (r_{t+1} - r_f)\pi_{i,t}] \tag{5.2}$$

Agent i myopically maximizes $\mathbb{E}_{i,t}[U_i(W_{i,t+1})]$ in terms of either the amount of the asset ($z_{i,t}$) or the wealth proportion ($\pi_{i,t}$), where $\mathbb{E}_{i,t}$ denotes the expectation formed by agent i conditional on his information up to time t and U_i is the agent's utility function. An alternative approach to expected utility maximization is to assume a "reasonable" asset-demand function, as is done by Beja and Goldman (1980); Day and Huang (1990); Chiarella (1992); Chiarella, He, and Hommes (2006); and several other authors. It often turns out that such seemingly ad hoc demand functions can be reconciled with some underlying expected utility maximizing story (see, for example, Chiarella, Dieci, and Gardini, 2002).

5.2.2. Utility Functions

The type of utility function employed will determine the nature of the dynamics of the price and wealth processes, for instance, if they are coupled or separated and growing or nongrowing processes. Generally, two classes of CARA and CRRA utility functions have been considered in the literature. We will see that their implications for the price and wealth dynamics are significantly different.

As a CARA utility function, we consider the utility function for agent i given by

$$U_i(W) = -e^{-\alpha_i W} \tag{5.3}$$

where α_i is the agent's absolute risk-aversion coefficient. If the next period wealth is assumed to be conditionally normal, as we shall assume, then maximizing expected utility of wealth, $\max_{z_{i,t}} \mathbb{E}_{i,t}[U_i(W_{i,t+1})]$, is equivalent to mean-variance optimization of the certainty equivalent of wealth,[2] $\max_{z_{i,t}}[\mathbb{E}_{i,t}(W_{i,t+1}) - (\alpha_i/2)\mathbb{V}_{i,t}(W_{i,t+1})]$, where $\mathbb{V}_{i,t}$ denotes the conditional variance. Consequently, the asset-demand (in quantity terms) is independent of wealth. It is this important feature that makes the dynamics of equilibrium prices evolve independently of the dynamics for wealth in the CARA framework. This feature has been used to characterize a nongrowing price process, at least for

[2] Note that if $\tilde{x} \sim N(\mu, \sigma^2)$, then $\mathbb{E}[e^{\tilde{x}}] = e^{\mu + \frac{1}{2}\sigma^2}$.

deterministic models, in the literature. As a result the price dynamics are driven by a nongrowing process, a fact that has allowed many analytical results to be obtained in the literature.

Alternatively, we may assume a CRRA utility function for agent i, namely

$$U_i(W) = \begin{cases} \frac{W^{1-\gamma_i}}{1-\gamma_i} & \gamma_i \neq 1 \\ \ln W & \gamma_i = 1 \end{cases} \tag{5.4}$$

where γ_i is the agent's relative risk-aversion coefficient. If the next period return is assumed to be conditionally normal, as we usually assume for a growing wealth process, then maximizing expected utility of wealth $\max_{\pi_{i,t}} \mathbb{E}_{i,t}[U_i(W_{i,t+1})]$ implies that the optimal wealth proportion $\pi_{i,t}$ is independent of the wealth. Consequently, when the market-clearing price is determined by the excess demand, we will see that this makes the price dynamics much more involved, since the dynamics of the wealth process also need to be taken into consideration. This feature of the CRRA utility function leads to a growing price process, which gives us the realistic feature that both price and wealth are growing. Levy and Levy (1996) have analyzed numerically a heterogeneous agent model with $U(W) = \ln W$ and found that heterogeneous expectations yield more realistic asset price dynamics, compared to the corresponding homogeneous expectations.

For both types of utility function, once individual demands are formed, these come together to form the market aggregate demand, which is most conveniently discussed in quantity terms. Assume that all the agents in the market can be grouped into I types according to their beliefs. Let $n_{i,t}$ denote the fraction of type i agents at time t. Then the average market aggregate demand (in terms of number of shares) z_t can be written

$$z_t = \sum_{i=1}^{I} n_{i,t} z_{i,t} \tag{5.5}$$

The fraction of agents $n_{i,t}$ may either be held fixed in each period, or be allowed to vary (or switch) according to some measure of the fitness of the expectations scheme being used (see Dieci, Foroni, Gardini, and He, 2006). The most common switching strategies are those introduced by Brock and Hommes (1997, 1998) and will be discussed in the next section.

5.2.3. Market-Clearing Mechanisms

These are the mechanisms by which the market price is arrived at. The two most frequently used mechanisms being the Walrasian auctioneer, widely used in economic theory but which, as O'Hara (1995) points out, is used in only one market (the market for silver in London), and the market-maker mechanism, which is close in spirit to the specialist system. We will see that each mechanism leads to a different formulation of the market equilibrium conditions, which in turn have a different impact on the market price dynamics.

In the Walrasian auctioneer scenario, each agent solves her optimization problem treating the market-clearing price in that period as parametric. The auctioneer announces a price and receives from all market participants what their demand/supply would be at that price. The auctioneer then determines the excess demand at that price. The auctioneer keeps announcing prices until a price is arrived at that sets excess demand to zero; this price is then set by the auctioneer as the price for trading in the current period. Via this scenario each agent is effectively giving to the auctioneer her excess demand schedule. The auctioneer then aggregates these to obtain the market-clearing price. Thus essentially the auctioneer mechanism finds the market-clearing price P_t at time period t such that

$$\sum_{i=1}^{I} n_{i,t} z_{i,t} = N_t \tag{5.6}$$

where N_t is the average supply of shares per agent at time t, which can change over time due, for instance, to either new shares being issued or the splitting of shares. It should be stressed that in this scenario since, when agents form their demands, neither the price P_t nor the return r_t is known yet, expectations and hence demands of agents are formed based on information up to time $t-1$. Therefore, the market-clearing price P_t is defined implicitly by Eq. 5.6. It turns out that in the case of CARA utility this procedure leads to a mathematically well-defined price map; however, in the case of CRRA utility, because of the dependence of demand on both price and wealth, this might not always be so.

An alternative market-clearing mechanism is that of the market maker. At time period t, the market maker announces a price P_t based on the excess demand in the previous period $t-1$. Given this price P_t, the agents compute their optimal demands for the asset for time period t. The aggregation of these demands gives the aggregate market demand, which, when matched with the supply of assets, yields the excess demand. The market maker takes an offsetting long or short position in the risky asset so that the excess demand for period t is zero. The market maker then announces a price P_{t+1} for time period $t+1$ that moves the price in the direction of reducing the excess demand. Thus under this scenario the price map has the general form

$$P_{t+1} = P_t + \mu[z_t - N_t] = P_t + \mu\Big(\sum_{i=1}^{I} n_{i,t} z_{i,t} - N_t\Big) \tag{5.7}$$

where $\mu > 0$ measures the market maker's speed of reaction to the excess demand.[3] In contrast to the map obtained under the Walrasian auctioneer scenario, Eq. 5.7 is an explicit map for P_{t+1}. Furthermore, since agents know the price P_t when forming their demands at time t, expectations are based on information up to time t. Thus the price

[3]In many discussions of this mechanism it is often assumed that the supply of the asset $N_t = 0$, so that the excess demand is simply the aggregate demand. As explained in Chiarella, Dieci, and He (2007b), this amounts to having a risk-neutral fundamental price. Cases with nonzero supply of the asset(s) are discussed briefly in Sections 5.4.2 and 5.6.1.

maps that come out of this market-clearing mechanism can be and indeed are quite different from the ones under the Walrasian mechanism.

A potential problem with this mechanism is that the market maker could end up with unsustainably large long or short positions, though for most of the models we consider, the price fluctuations result in the market maker having, over time, a reasonably balanced position overall. Here and in the literature that we review the market maker plays a very simple role. Much more could (and should) be done to model the behavior and incentives of the market maker; see, for example, Ho and Stoll (1981), Peck (1990), and Madhavan (2000). As we shall see in the following analysis, the type of market-clearing mechanism used does in fact affect the dynamic behavior of the models.

5.2.4. Noise

The framework we've developed can also incorporate external noise factors by assuming that as well as the utility maximizing component arising from the optimization problem of all the agents, there is also some exogenous demand component distributed around zero. This may be regarded as being due to some exogenous demand component for each agent or from the market as a whole, perhaps due to news events whose impact is not readily incorporated into the expected utility maximizing calculus, or just so-called noise trader effects. In a great deal of literature this exogenous noisy demand is called irrational, and if one takes the view that any component of demand not arising from expected utility maximization is irrational, the label follows by definition. In any event the aggregate demand would now be written

$$z_t = \sum_{i=1}^{I} n_{i,t} z_{i,t} + \nu \widetilde{\varepsilon}_t \tag{5.8}$$

where $\widetilde{\varepsilon}_t \sim \mathcal{N}(0, 1)$, the standard normal distribution, measures some exogenous noise process affecting the market and ν is the standard deviation. Note that additional sources of noise in Eq. 5.8 could be a noisy fundamental price process and a noisy dividend process. We specify the exact nature of the external noise in the following discussion.

5.2.5. Expectations Feedback

The important feature of the structure of Eqs. 5.6 and 5.7 is that the price-generating mechanism is driven by expectations feedback. Observed market prices are used to form expectations, which in turn feed back to generate prices. It is this feed back mechanism that allows the BRHA models to generate much greater volatility and long fluctuations from the fundamental than seems possible under rational expectations. In fact, the dynamics underlying the rational expectations model are not greatly discussed in the literature, though Wenzelburger (2009) do discuss this case. It is also of interest to note that Böhm and Chiarella (2005) have shown that the dynamics of a market in which agents have homogeneous rational expectations with a positive risk-free rate of interest are divergent.

5.3. PRICE DYNAMICS IMPLIED BY THE CARA UTILITY FUNCTION

As we mentioned in the previous section, an important feature of the CARA utility framework is that the price dynamics constitute a nongrowing process that evolves independently of wealth dynamics. In this section, we assume that agents have the CARA utility function with different risk-aversion coefficients and focus on the different types of market behavior under the two different market-clearing mechanisms. Among the many types of possible heterogeneous agents, we focus on two of the most popular types discussed in the literature: fundamentalists and trend followers. In the following, we first introduce heterogeneous expectations and a switching mechanism that allows agents to switch between different strategies based on some fitness measure. We then examine the price dynamics under the Walrasian auctioneer market-clearing mechanism by focusing on the impact of different risk-aversion coefficients and extrapolation of the trend followers. The study of the price dynamics under the market-maker scenario is then followed by an analysis of the impact of the activities of the market maker and trend followers. A comparison of the price dynamics under the two different market-clearing mechanisms and the same strategy for the trend followers is provided at the end of this section.

5.3.1. Fundamental Price and the Optimal Demand

With the CARA utility function (Eq. 5.3) and the framework in Section 5.2, the optimal asset-demand (number of shares) for agent i turns out to be given by

$$z_{i,t} = \mathbb{E}_{i,t}(R_{t+1})/(\alpha_i \mathbb{V}_{i,t}(R_{t+1})) \tag{5.9}$$

where $R_{t+1} = P_{t+1} + y_{t+1} - RP_t$ denotes the excess (dollar) return on the risky asset.

We are interested in how the market price is related to its fundamental price. There may be many ways to estimate the fundamental price. In this section, we follow Brock and Hommes (1998) and assume that there exists a common "fundamental price" P_t^* that solves

$$P_t^* = \frac{1}{R}\mathbb{E}_t\left(P_{t+1}^* + y_{t+1}\right)$$

if all agents were to hold common beliefs denoted by $\mathbb{E}_t$. Here we assume that the dividend y_t follows an IID process.

It will be convenient to define the deviation from the fundamental, $x_t = P_t - P_t^*$, so that the excess return R_t may be written

$$R_{t+1} = x_{t+1} - Rx_t + \zeta_{t+1} \tag{5.10}$$

where we set $\zeta_{t+1} = P_{t+1}^* + y_{t+1} - \mathbb{E}_t(P_{t+1}^* + y_{t+1})$. We note that this latter quantity forms a martingale difference sequence, since $\mathbb{E}_t(\zeta_{t+1}) = 0$ for all t. Aggregating Eq. 5.9

over all agents, we obtain the aggregate demand

$$z_t = \sum_{i=1}^{I} n_{i,t} \frac{\mathbb{E}_{i,t}(R_{t+1})}{\alpha_i \mathbb{V}_{i,t}(R_{t+1})} \tag{5.11}$$

5.3.2. Formation of Heterogeneous Beliefs

Let us now turn to the question of how agents form their beliefs about conditional means and variances. The beliefs are assumed to have a common component based on knowledge of the fundamental and dividends, and a component particular to each group of agents. Thus the heterogenous beliefs about the conditional mean and variance of the price are assumed to be of the form

$$\mathbb{E}_{i,t}(P_{t+1} + y_{t+1}) = \mathbb{E}_t(P^*_{t+1} + y_{t+1}) + f_{i,t} \tag{5.12}$$

$$\mathbb{V}_{i,t}(P_{t+1} + y_{t+1}) = \mathbb{V}_t(P^*_{t+1} + y_{t+1}) + g_{i,t} = \sigma^2 + g_{i,t} \tag{5.13}$$

where $f_{i,t}$ and $g_{i,t}$ are some deterministic functions of a window of the past L_i price deviations. We assume that each agent has a different window length L_i (an integer), and $\sigma^2 > 0$ is a constant representing the volatility of the fundamental. It follows from Eqs. 5.12 and 5.13 that

$$\mathbb{E}_{i,t}(R_{t+1}) = f_{i,t} - Rx_t, \quad \mathbb{V}_{i,t}(R_{t+1}) = \sigma^2 + g_{i,t} \tag{5.14}$$

For the various market-clearing mechanisms, the conditional mean and variance functions can have different forms. In the Walrasian auctioneer scenario, the market-clearing price P_t at time t is not in the information set of agents at time t. Possible forms for the function $f_{i,t}$ are

$$f_{i,t} = e_i + d_i \bar{x}_{i,t} \tag{5.15}$$

where $\bar{x}_{i,t}$ could have three different forms:

- The moving average process (MAP)

$$\bar{x}_{i,t} = \frac{1}{L_i} \sum_{l=1}^{L_i} x_{t-l} \tag{5.16}$$

- A geometric decay process (GDP)

$$\bar{x}_{i,t} = \left(\sum_{l=1}^{L_i} \delta_i^l \right)^{-1} \sum_{l=1}^{L_i} \delta_i^l x_{t-l} \tag{5.17}$$

where $\delta_i \in [0, 1]$ is the memory decay rate of agent i

- The limiting geometric decay process (LGDP)

$$\bar{x}_{i,t} = \delta_i \bar{x}_{i,t-1} + (1 - \delta_i) x_{t-1} \quad \textbf{(5.18)}$$

Clearly, LGDP is the limiting situation of GDP as $L_i \to \infty$, whereas MAP is obtained from GDP by setting $\delta_i = 1$. In the market-maker scenario, the market price P_t at time t is in the information set of agent i at time t. Hence the summations in Eqs. 5.16 and 5.17 run from $l = 0$ to $l = L_i - 1$, and Eq. 5.18 is replaced by

$$\bar{x}_{i,t} = \delta_i \bar{x}_{i,t-1} + (1 - \delta_i) x_t$$

A possible form for $g_{i,t}$ is

$$g_{i,t} = \sigma^2 v_i(\sigma^2_{i,t}) \quad \textbf{(5.19)}$$

where $\sigma^2_{i,t}$ is the corresponding sample variance to the preceding sample mean process and v_i is a function that bounds the variance belief between some upper and lower values. This variance belief function is inspired by Franke and Sethi (1998) and operates in such a way that agents increase (decrease) their belief about variance according to whether it has recently been high (low). The function v_i plays the role of putting upper and lower bounds on their estimates of future variance. Such an expectations scheme could be based on a belief in some sort of mean reverting process for volatility, which is certainly borne out by empirical studies.

Different investor types are characterized by different values of e_i, d_i, and δ_i in Eqs. 5.15 and 5.17. Thus we can identify three important groups as follows: (1) fundamentalists: $e_i = d_i = 0$; (2) trend followers: $e_i = 0, d_i > 0$; and (3) contrarians: $e_i = 0, d_i < 0$.

5.3.3. Performance Measure and Switching

The quantities that still need to be specified in Eq. 5.11 are the fractions $n_{i,t}$. The scheme that has become widely used is the one proposed by Brock and Hommes (1997), who assume that agents switch strategies based on some measure of the fitness of the various strategies. The most convenient (but by no means the only) fitness measure is the realized profit for agents of type i, namely $R_t z_{i,t-1}$. More generally we could allow the reaction to realize profits to be lagged and consider the geometrically declining weighted average of realized profits calculated according to

$$M_{i,t} = R_t z_{i,t-1} + \eta_i M_{i,t-1} \quad \textbf{(5.20)}$$

where the parameter $\eta_i (> 0)$ represents the memory strength of agent i. Due to space limitations here we only consider the case $\eta_i = 0$, though broadly speaking the effect of $\eta_i > 0$ is to slow the dynamic as it causes agents to react less quickly to the profitability of a particular strategy.

Brock and Hommes (1997) propose that the fractions $n_{i,t}$ be calculated on the basis of the fitness in period $t-1$ according to a discrete choice probability model (see Manski and McFadden, 1981, and Anderson, de Palma, and Thisse, 1993), which under the Walrasian auctioneer scenario assume the form[4]

$$n_{i,t} = \exp[\beta M_{i,t-1}]/Z_t, \quad Z_t = \sum_{i=1}^{I} \exp[\beta M_{i,t-1}] \tag{5.21}$$

where $\beta(>0)$ is the intensity of choice measuring how fast agents switch among different prediction strategies. In particular, $\beta = +\infty$ means that in each period the entire mass of traders switches to the strategy that had the highest fitness in the previous period, whereas $\beta = 0$ means that the mass of traders distributes itself evenly across the set of available strategies. For $0 < \beta < \infty$, a fraction of each group of traders switches to the most recently successful strategy. Note that the fraction $n_{i,t}$ is a random process in general due to the randomness of $\boldsymbol{R}_t$.

A range of BRHA models can be obtained, depending on how we specify expectations, the market-clearing mechanism assumed, and whether agent proportions remain fixed or switch according to some fitness measure. It is also possible to have versions where part of the agent proportions remain fixed and part are switching (see Dieci, Foroni, Gardini, and He, 2006 in this regard). In the following two subsections we will consider both the Walrasian auctioneer and the market-maker price-clearing mechanisms. We shall analyze the type of equilibria that can arise, the stability properties and the ensuing dynamic behavior, and how these can differ for the two different market-clearing mechanisms.

5.3.4. Price Behavior under the Walrasian Auctioneer Mechanism

The Walrasian auctioneer market-clearing mechanism, widely used in economic theory, has been used in the context of BRHA models by, among others, Brock and Hommes (1997, 1998), Gaunersdorfer (2000), and Chiarella and He (2002). In this case, assuming zero supply of shares, we then have from $z_t = 0$, Eqs. 5.11, 5.12, 5.13, and 5.14 that

$$R\left[\sum_{i=1}^{I} \frac{n_{i,t}}{\alpha_i\left(\sigma^2 + g_{i,t}\right)}\right] x_t = \sum_{i=1}^{I} \frac{n_{i,t} f_{i,t}}{\alpha_i\left(\sigma^2 + g_{i,t}\right)} \tag{5.22}$$

which is the general form of the price map implicitly expressing P_t as a function of past prices over a window determined by the largest L_i. It then remains to specify the mean and variance expectation functions $f_{i,t}$ and $g_{i,t}$ to obtain the precise form of the price map.

[4]Note that under the market-maker scenario the information set of agents includes P_t so that in this case we replace $M_{i,t-1}$ in Eq. 5.21 by $M_{i,t}$.

An Asset Price Model of the Fundamentalists and Trend Followers

Chiarella and He (2002) consider three types of agents; fundamentalists, trend chasers, and contrarians. By combining stability and bifurcation analysis and numerical simulation, they examine the market price behavior and various routes to complex dynamics that can occur. Due to limited space we concentrate here on the case of a market consisting of fundamentalists ($i = 1$) and trend followers ($i = 2$). To simplify the analysis, we also do not take the conditional variance component $g_{i,t}$ into account; that is, we set $g_{i,t} = 0$, and refer the reader to Chiarella and He (2002) for a discussion of the general case. In terms of the conditional mean function we've introduced, we assume that

$$f_{1,t} = 0 \quad \text{and} \quad f_{2,t} = d\bar{x}_t \tag{5.23}$$

where $\bar{x}_t = (1/L)\sum_{i=1}^{L} x_{t-i}$ is the price trend based on the moving average over the last L price deviations and $d > 0$ measures the extrapolation rate of the trend followers. Thus, the fundamentalists believe that the price will return to its fundamental value and the trend followers form their expectation of next period's price based on the trend calculated from the past L realized prices. As pointed out by Campbell and Kyle (1993), we would expect the fundamentalists to be more risk averse than the trend followers, so in the subsequent analysis we assume that $a := \alpha_2/\alpha_1 < 1$.

In models with two types of agents, it is convenient to express the fractions $n_{1,t}$ and $n_{2,t}$ in terms of the variable m_t, defined by $m_t = n_{1,t} - n_{2,t}$ so that

$$n_{1,t} = (1 + m_t)/2 \quad \text{and} \quad n_{2,t} = (1 - m_t)/2 \tag{5.24}$$

With the previous specifications, the use of Eqs. 5.23 and 5.24, we find that the price map (Eq. 5.22) can be expressed as

$$x_t = \frac{d}{R}\frac{1 - m_{t-1}}{a(1 + m_{t-1}) + (1 - m_{t-1})}\bar{x}_t \tag{5.25}$$

$$m_t = \tanh\left[\frac{\beta}{2\alpha_1\sigma^2}(Rx_{t-1} - x_t)\left(Rx_{t-1} + \frac{d\bar{x}_{t-1} - Rx_{t-1}}{a}\right) - \frac{\beta C}{2}\right] \tag{5.26}$$

Brock and Hommes (1998) consider Eqs. 5.25 and 5.26 in the case when $L = 1$ and all agent types have the same risk-aversion coefficient so that $a = 1$. Here we consider the case of general L and $a \neq 1$. In the following discussion, we assume that the dividend process y_t follows $y_t = \bar{y} + \epsilon_t$, where $\bar{y}$ is the mean and ϵ_t is an IID noise, which is assumed to be uniformly distributed on an interval $[-\epsilon, \epsilon]$. We first focus on the deterministic dynamics by assuming $\epsilon_t = 0$ for all t. We then compare the dynamics with and without the noise by using numerical simulations.

The Case of $L = 1$

Chiarella and He (2002) consider first the case $L = 1$. Depending on the extrapolation rate of the trend chasers, the system (Eqs. 5.25 and 5.26) may have more than one steady state with different stability properties.

Theorem 5.1. *For the deterministic system Eqs. 5.25 and 5.26, let* $m^{eq} := \tanh(-\frac{\beta C}{2})$, $m^* := 1 - \frac{2aR}{d+R(a-1)}$ *and* x^* *be the positive solution (which exists if and only if* $m^* > m^{eq}$*) of*

$$\tanh\left[\frac{\beta}{2\sigma^2\alpha_1}(R-1)(R+\frac{d-R}{a})(x^*)^2 - \frac{\beta C}{2}\right] = m^* \quad \textbf{(5.27)}$$

- *For* $0 < d < R$, *the fundamental steady state* $E_1 = (0, m^{eq})$ *of the system is globally asymptotically stable*
- *For* $R < d < (1+a)R$
  - *If* $m^* < m^{eq}$, *the fundamental equilibrium* E_1 *is the unique, globally stable steady state of the system*
  - *If* $m^* > m^{eq}$, *the system has three steady states,* $E_1, E_2 = (x^*, m^*)$, *and* $E_3 = (-x^*, m^*)$, *of which* E_1 *is unstable*
- *For* $d > (a+1)R$, *the system has three steady states,* $E_1, E_2 = (x^*, m^*)$, *and* $E_3 = (-x^*, m^*)$, *of which* E_1 *is unstable*

Theorem 5.1 indicates that, when the trend followers extrapolate only weakly ($0 < d < R$), the fundamental steady state $E_1 = (0, m^{eq})$ is globally stable, no matter what risk attitudes agents have. However, when $d > R$, the stability of the fundamental equilibrium E_1 depends on the risk aversion ratio a, which measures the relative risk attitude between the two groups. When the trend followers extrapolate very strongly ($d > (a+1)R$), the fundamental equilibrium E_1 becomes unstable and bifurcates two additional nonzero steady states, E_2 and E_3. In the case of $R < d < (a+1)R$, the fundamental equilibrium E_1 is stable when $m^* < m^{eq}$ and unstable when $m^* > m^{eq}$. One can see that $m^* > m^{eq}$ when a is small, that is when the trend followers become less risk averse than the fundamentalists. This point will become clearer in the following discussion.

To fully understand the impact of the relative risk aversion ratio a on the dynamical behavior of the model when $R < d < (a+1)R$ one needs to study how changes in a affect the bifurcations. Let a^* satisfy

$$\tanh\left(-\frac{\beta C}{2}\right) = 1 - \frac{2a^* R}{d + (a^* - 1)R} \quad \textbf{(5.28)}$$

being the value of a at which $m^{eq} = m^*$. Then we have the following result.

Theorem 5.2. *For the deterministic system Eqs. 5.25 and 5.26,* $m^* < m^{eq}$ *if and only if* $a > a^*$. *In addition,*

- *A pitchfork bifurcation occurs for* $a = a^*$*:*
  - *For* $a > a^*$, E_1 *is the unique stable equilibrium*
  - *For* $0 < a < a^*$, *there are three equilibria,* E_1, E_2 *and* E_3
- *There exists* $a^{**} < a^*$ *such that* E_2 *and* E_3 *are stable for* $a \in (a^{**}, a^*)$ *and unstable for* $a < a^{**}$. *For* $a = a^{**}$, E_2 *and* E_3 *exhibit Hopf bifurcations.*

We note that for $a = 1$, Brock and Hommes (1998) obtained similar dynamics by allowing the switching parameter β to vary. To obtain a better picture of the dynamics, we perform some simulations for the parameter set $R = 1.1, d = 1.2, C = 1.0, \beta = 3.5, \sigma^2 = 1.0, \alpha_1 = 1.0$, with initial value (1.2, 0.7, −0.2) and the different values $\alpha_2 = a = 0.9, 0.95, 1.0, 1.1,$ 1.2. Obviously, $R < d < (a + 1)R$ is satisfied. The pitchfork bifurcation parameter value turns out to be $a^* = 3.01$. Then the fundamental steady state E_1 is globally stable for $a > 3.01$. Figure 5.1 shows plots of the attractors in the (x_t, m_t) plane. There we see that the orbit converges to the positive equilibrium E_2 for $a = 1.2$ and then to an attracting invariant "circle" surrounding E_2 for $a = 1.1$, which indicates that the Hopf bifurcation value $a^{**} \in (1.1, 1.2)$. Then as a decreases further, the "circle" breaks into invariant sets. For $a = 1$, the prices oscillate about the positive equilibrium E_2 such that $m_t < 0$ for all time t and the market is dominated by the chartists. As a increases, the prices are stabilized to E_2. As a decreases, Chiarella and He (2002) present time series that show that the price switches between an unstable phase with an upward trend and a stable phase with prices close to the fundamental

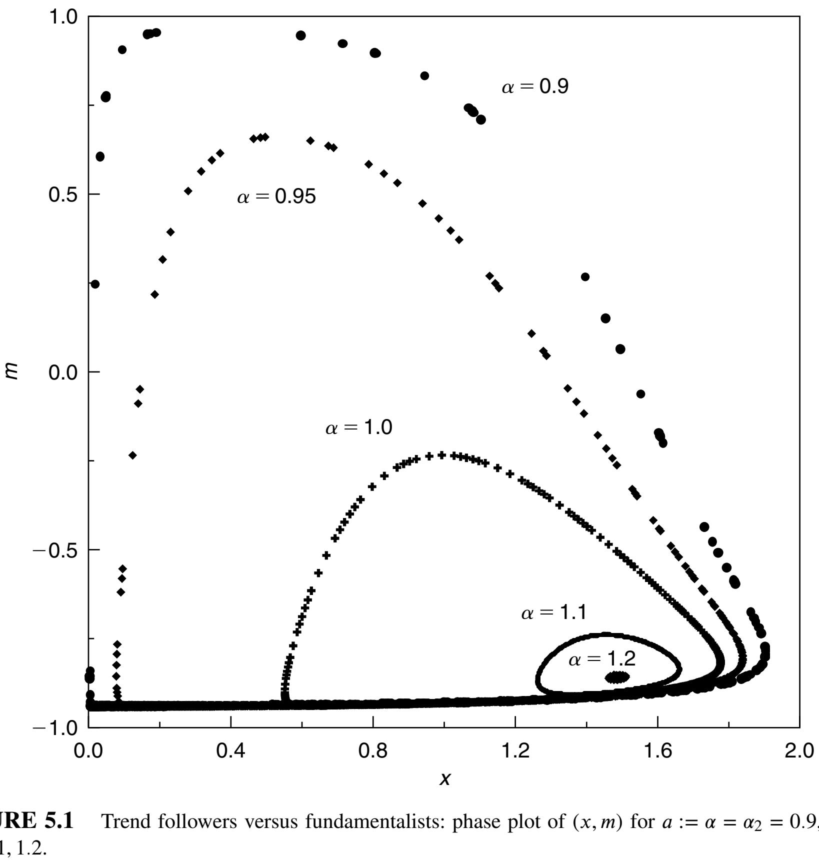

FIGURE 5.1 Trend followers versus fundamentalists: phase plot of (x, m) for $a := \alpha = \alpha_2 = 0.9, 0.95, 1.0, 1.1, 1.2$.

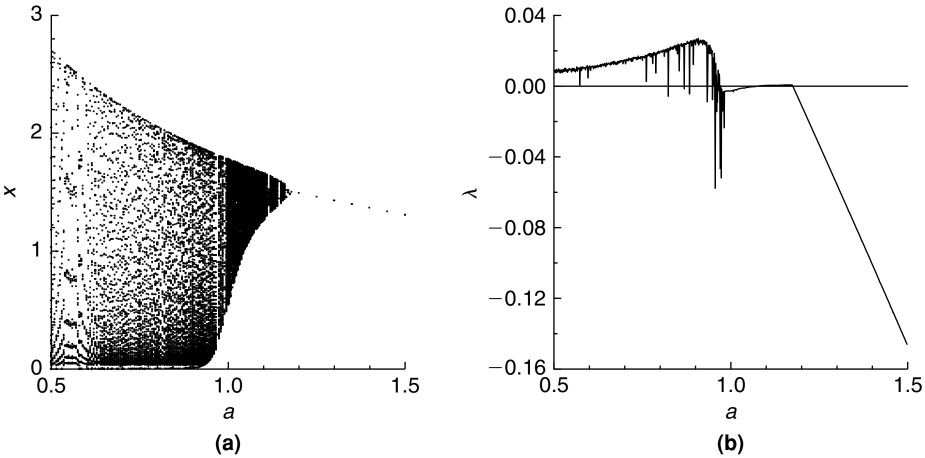

FIGURE 5.2 Trend followers versus fundamentalists: (a) bifurcation diagram and (b) the largest Lyapunov exponent plot.

value. Figure 5.2a shows a bifurcation diagram with respect to the relative risk ratio a, suggesting periodic and quasi-periodic dynamics after the primary Hopf bifurcation as a decreases. Figure 5.2b shows the corresponding largest Lyapunov exponent plot, the positivity of which indicates chaotic behavior. When the fundamentalists are more risk averse than trend followers ($a < 1$), a decrease in a leads to numerical evidence for weakly chaotic asset price fluctuations with an irregular switching between a point close to the fundamental price and upward and downward trends.

The preceding numerical simulations suggest that when the trend followers are more risk averse than the fundamentalists, the market is dominated by the fundamentalists and the prices converge to the fundamental value. When the fundamentalists are more risk averse, the market becomes unstable, even chaotic. Further numerical simulations (not reported here) confirm that when $d < R$, the fundamental equilibrium is globally stable, no matter the degree of risk aversion of both groups. When $d > (a + 1)R$, the fundamental equilibrium is unstable.

The Case of $L \geq 2$

For $L \geq 2$, one can check that the equilibrium of the system is the same as the case when $L = 1$. The system has either one unique equilibrium, E_1, or three equilibria, E_1, E_2 and E_3. In this more general case, we are more interested in the stability of the fundamental equilibrium, about which Chiarella and He (2002) obtain the following result.

Theorem 5.3. *The fundamental steady state E_1 of the deterministic system (Eqs. 5.25 and 5.26) is locally asymptotically stable if and only if*

$$\frac{d}{R}\frac{1 - m^{eq}}{a + 1 + (a - 1)m^{eq}} < 1 \tag{5.29}$$

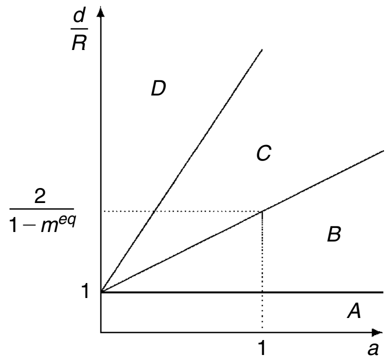

FIGURE 5.3 Stability region of the fundamental steady state E_1: stable in A and B, unstable in C and D.

Note that the condition in Eq. 5.29 is independent of the lag length L. In particular, when $\alpha_1 = \alpha_2$ (so that $a = 1$), the condition can be written as $m^* := 1 - 2R/d < m^{eq}$, which is the same condition derived by Brock and Hommes (1998) . Thus, the preceding analysis provides a generalization of their result to the situation of different risk-aversion coefficients. The region of stability of the fundamental steady state in the $(a, d/R)$ plane is shown in Figure 5.3, where the four regions A, B, C, and D are separated by the lines $\frac{d}{R} = 1, 1 + \frac{1+m^{eq}}{1-m^{eq}} a$ and $1 + a$, respectively. The fundamental steady state is locally stable in regions A and B and unstable in regions C and D. Furthermore, there exist two other steady states, E_2 and E_3, in C and D. Let a^* be the value given by Eq. 5.28; then we can verify for a given value of $d/R(> 1)$ that $a^* = (\frac{d}{R} - 1)\frac{1-m^{eq}}{1+m^{eq}}$. Therefore, E_1 is locally asymptotically stable for $a > a^*$, since it will lie in region B.

We also carry out some simulations using the same parameter set as before and, in addition, set $\beta = 3.8$. For the noisy case,[5] we choose $\epsilon = 0.05$. Now, it turns out that $a^* = 4.06375$. Figures 5.4a1, 5.4b1, and 5.4c1 show the phase plots in the (x_t, m_t) plane for $L = 2$, 5 and 10, respectively, for different values of $\alpha_2 = a$ when $\epsilon_t = 0$. Figures 5.4a2, 5.4b2, and 5.4c2 plot the corresponding time series for x_t without noise, and Figures 5.4a3, 5.4b3 and 5.4c3 with noise. The numerical simulations suggest that:

- Just as in the case when $L = 1$, there exists a second bifurcation value $a^{**} \in (1.0, 1.2)$ for $L = 2$, 5 and 10.
- When lag length L increases, the attractors on the (x_t, m_t) plane become more complicated. Say, for $a = 0.5$, when $L = 2$, the prices switch between an unstable phase with an upward trend and a stable phase with prices close to the fundamental value (see Figure 5.4a2); when $L = 5$ and $L = 10$, the prices fluctuate away from the fundamental value; the periods of upward trend for $L = 10$ are longer than when $L = 5$ (see Figures 5.4b2 and 5.4c3).

[5]As before, we assume that the dividend process y_t follows $y_t = \bar{y} + \varepsilon_t$, where $\bar{y}$ is the mean and ε_t is IID noise distributed uniformly on the interval $[-\varepsilon, \varepsilon]$.

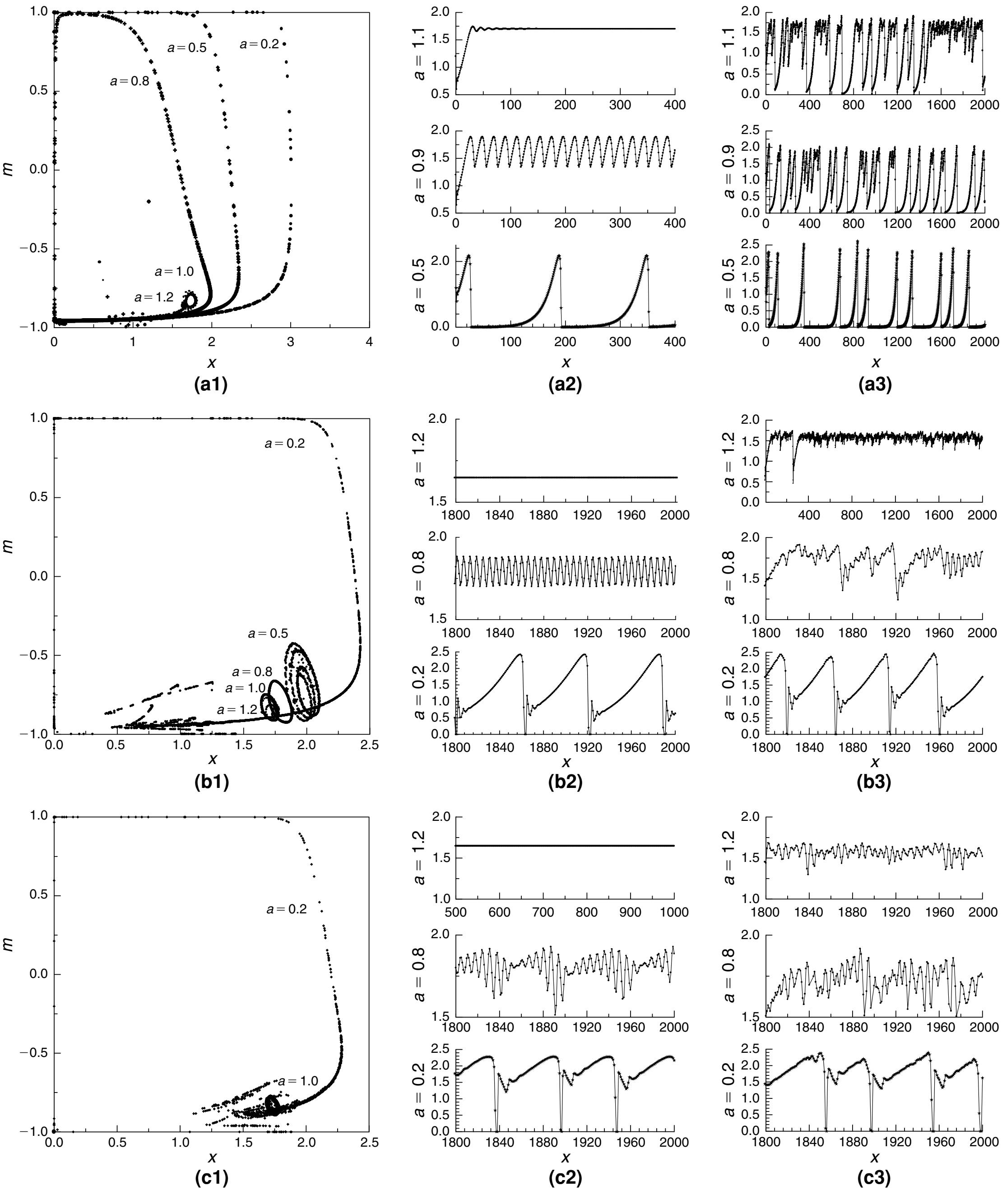

FIGURE 5.4 Trend followers versus fundamentalists: phase plots (x, m) for $L = 2$ (a1), $L = 5$ (b1), and $L = 10$ (c1) without noise; time series of x_t without noise for $L = 2$ (a2), $L = 5$ (b2), and $L = 10$ (c2); and with noise for $L = 2$ (a3), $L = 5$ (b3), and $L = 10$ (c3).

- The external noise has a more significant effect on the dynamical behavior of the model with short lag length ($L = 2$ in Figure 5.4a3) than long lag length ($L = 5$ and 10 in Figures 5.4b3 and 5.4c3). Also, it has more effect for a high-ratio a than for a low-ratio a. In other words, when the system exhibits complicated behavior without noise, adding noise has no significant effect on the dynamics (when the size of the noise is small); however, if the system without noise is stable, adding noise can lead to significant changes in the dynamics of the system.

The impact of the lag length can be very complicated in general, and we refer the reader to Chiarella, He, and Hommes (2006) for a discussion on the dynamics of the moving average.

5.3.5. Price Behavior under the Market-Maker Mechanism

We now examine price behavior when the market price is determined via a market maker mechanism. In this case, as discussed earlier, with a zero supply of shares and a market consisting of fundamentalists ($i = 1$) and trend followers ($i = 2$), the dynamics for the price deviation can be expressed as

$$x_{t+1} = x_t + \mu[(1 + m_t)z_{1,t} + (1 - m_t)z_{2,t}]/2 \qquad \textbf{(5.30)}$$

$$m_t = \tanh(\beta R_t[z_{1,t-1} - z_{2,t-1}]/2 - \beta C/2) \qquad \textbf{(5.31)}$$

The system (Eqs. 5.30 and 5.31) has been extensively analyzed by Chiarella and He (2003) in the case when first-moment beliefs are given by Eq. 5.12 with $f_{1,t} = 0$, $f_{2,t} = d\bar{x}_t$ and second-moment beliefs are constant, so that $g_{i,t} = 0$ in Eq. 5.14. Different from the previous case, the price P_t at time t is part of the agents' information set and hence appears in the geometric moving average, so that

$$\bar{x}_t = \left(\sum_{i=0}^{L-1} \delta^i\right)^{-1} \sum_{i=0}^{L-1} \delta_i^i x_{t-i}, \qquad \delta \in [0, 1]$$

Chiarella and He (2003) show that the equilibrium of Eqs. 5.30 and 5.31 is exactly the same as the corresponding Walrasian auctioneer equilibrium described in the previous subsection. It is in the dynamic behavior that the two market-clearing scenarios differ. Here we just give results for the case of the geometric decay process with finite memory and recall that this includes as a special case the moving average process. We focus in particular on the impact of the speed of adjustment of the market maker (μ), the extrapolation rate of the trend followers (d), the decay rate (δ), and the lag length (L).

The Local Stability of the Fundamental Steady State

It turns out that the dynamics of the market-maker scenario depend on the parameter combination $\bar{\mu} := \mu/4\alpha_2\sigma^2$, which is the speed of adjustment of the market maker weighted (negatively) by the risk aversion of the chartists and the strength of the

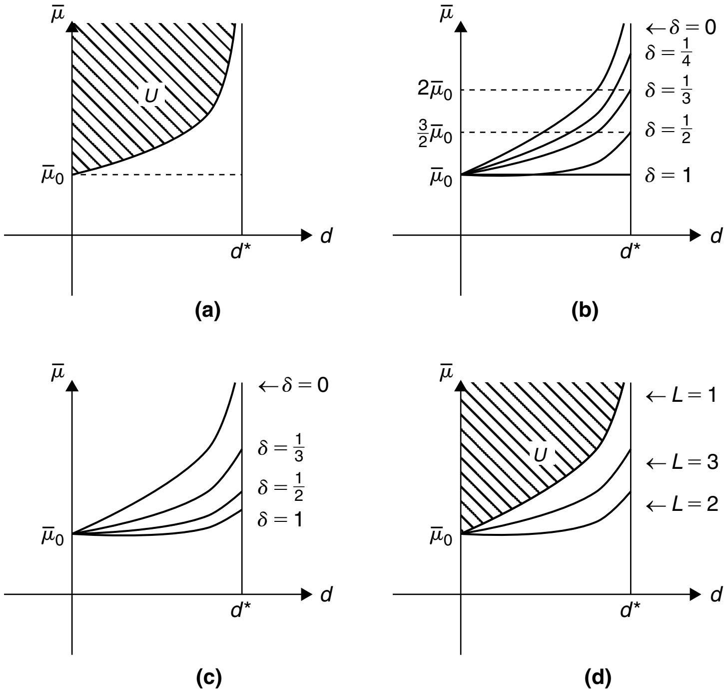

FIGURE 5.5 Unstable region (as indicated by U) and local stability regions (whose upper bounds are given by the various curves) of the fundamental steady state for $L = 1$ (a), $L = 2$ (b), and $L = 3$ (c) with different $\delta \in [0, 1]$. Comparison (d) of the local stability regions for $L = 1, 2, 3$ with fixed $\delta \in (0, 1)$.

variance. Chiarella and He (2003) find that the local stable and unstable regions of the fundamental steady state in the $(d, \bar{\mu})$ plane for various δ and L can be depicted as shown in Figure 5.5 for different lag lengths and different values of the decay rate. With regard to the (local) stability region of the fundamental steady state in the $(d, \bar{\mu})$ parameter space of the extrapolation rates and the adjustment speed of the market maker, we can see the following:

- For fixed δ, an increase in the lag length L does not necessarily enlarge the stability region as illustrated in Figure 5.5d. There does not seem to be any connection between the lag length and the size of the stability region. This observation contradicts a common belief that the stability regions are enlarged when agents include more historical data in forecasting rules.
- For fixed lag length $L = 2, 3$, a decrease in the memory decay rate δ enlarges the local stability region of the fundamental steady state, as illustrated in Figures 5.5b and 5.5c.
- The stability region for $L = 2, 3$ becomes the stability region for $L = 1$ as the decay rate $\delta \to 0^+$, as illustrated in Figures 5.5a–c.

Chiarella and He (2003) use numerical simulation to determine the effect of the lag length L. They show that, when the fundamental steady state becomes unstable, the price series can converge to one of the nonfundamental equilibria for $L = 3$ but to (quasi-)periodic orbits for $L = 5$ and 10 with the phase plots of the prices series being closed orbits encircling the nonfundamental steady states.

Nonlinear Dynamics Under the Moving Average Process

For the moving average process, unlike the asset price model under the Walrasian scenario, the stability conditions are much more related to the speed of the adjustment of the market maker for different lag lengths and, in general, increasing the lag length does not necessary increase the stability of the fundamental steady state. This is also confirmed in Chiarella, He, and Hommes (2006). It may or may not be the case that the stability of the nonfundamental equilibria are improved as the lag length increases. Chiarella and He (2003) show that, when the steady state becomes unstable, the time series for $L = 1$ fluctuates about the fundamental equilibrium initially, is then stabilized to one of the nonfundamental equilibria for $L = 2$ and 3, but increasing L further to 5 and 10 leads to some periodic cycles, which can be regarded as bifurcations from the nonfundamental equilibria.[6]

Nonlinear Dynamics Under the Geometric Decay Process

For the general geometric decay process, a further important effect is that of the speed of adjustment of the market maker μ. For $L = 2$, Figure 5.6 illustrates the time series for

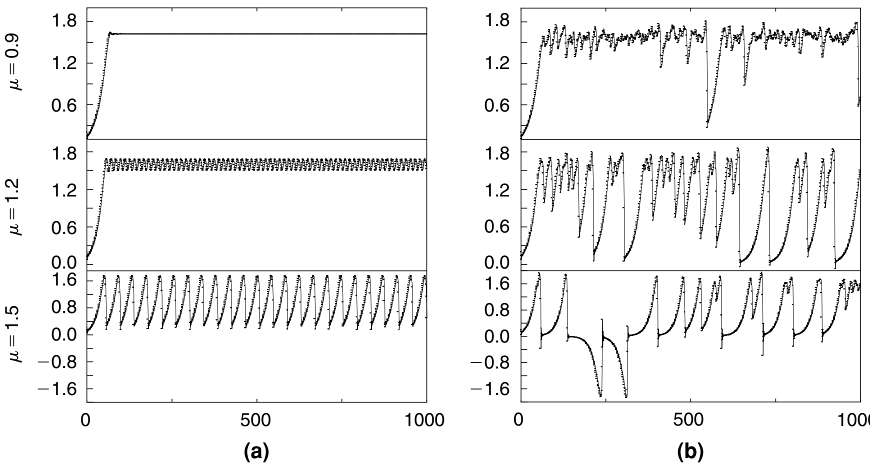

FIGURE 5.6 Time series for $L = 2$ with different values of $\mu = 0.9, 1.2, 1.5$, (a) without noise and (b) with noise for $\delta = 0.5$.

[6]Due to space limitations, we do not discuss the stability of the nonfundamental equilibria.

$\mu = 0.9, 1.2, 1.5$ without and with noise ($\epsilon = 0.05$). Without noise, the figures indicate that, when the trend followers extrapolate strongly, a small adjustment from the market maker leads the price to converge to one of the nonfundamental steady states. However, as the market maker adjusts more strongly in a way such that both forces from extrapolation of the trend chasers and adjustment of the market maker are balanced, the price fluctuates periodically or quasi-periodically around one of the nonfundamental steady states. Large adjustment from the market maker results in the breaking of such balance, leading the price to fluctuate among the three steady states. Adding the noise amplifies these effects. In general, large decay rates δ stabilize the price dynamics, whereas small decay rates destabilize the price dynamics, leading to quasi-periodic cycles.

Comparison of the Two Market-Clearing Mechanisms

By comparing the price dynamics under the Walrasian and market-maker scenarios, we arrive at the following conclusion concerning BRHA models with agents having CARA utility functions:

- When the moving average process is used by the trend followers, with the Walrasian market-clearing scenario, the local stability of the fundamental steady state is completely characterized by the extrapolation rate of the trend followers and the ratio of the risk-aversion coefficients of the fundamentalists and the trend followers but not the lag length. Complicated price dynamics can only be generated through saddle-node-type bifurcations. However, under the market-maker scenario, the stability of the fundamental steady state is maintained only when the speed of adjustment of the market maker is low and balanced with the extrapolation rate of the trend followers. Furthermore, complicated price dynamics can be generated through different types of bifurcations. An increase in lag length under the market-maker scenario might not necessarily enlarge the local stability region. However, different lag lengths can complicate the price dynamics in different ways under both scenarios.
- The geometric decay rate for the trend followers has more complicated effects on the price dynamics under the market-maker scenario. When the trend followers extrapolate weakly, a decrease in the decay rate enlarges the local stability region of the fundamental steady state. However, when the fundamental steady state becomes unstable and the trend followers extrapolate strongly, an increase in the decay rate can stabilize the price to either one of the nonfundamental steady states or to cyclical attractors.
- When the fundamental steady state becomes unstable, the price can be pushed either to one of the two nonfundamental steady states or to cycles enclosing them under the Walrasian scenario. However, under the market-maker scenario, the price can fluctuate among the three steady states. This is in particular the case when noise impinges on the system and volatility clustering can be observed.

In summary, the dynamical behavior of the BRHA asset-pricing model under the market maker scenario is considerably enriched and has some significant differences from its behavior under the Walrasian auctioneer scenario.

5.4. PRICE BEHAVIOR AND WEALTH DYNAMICS IMPLIED BY THE CRRA UTILITY

This section deals with models of asset price dynamics where boundedly rational heterogeneous agents are characterized by CRRA utility of the type represented by Eq. 5.4. The BRHA literature has primarily focused on models with CARA investors, essentially for reasons of analytical tractability. As a matter of fact, it is not possible, in general, to arrive at an exact solution to the agent's expected utility maximization problem in the case of CRRA utility, though of course approximate solutions can be provided (see, for instance, Campbell and Viceira, 2002). However, a number of models assuming CRRA preferences have been proposed in recent years, starting with Levy, Levy, and Solomon (1994, 2000) and Zschischang and Lux (2001), the analysis of which relies essentially on numerical simulation.

One reason for these developments is that the assumption of CRRA utility is considered more realistic than CARA from a number of standpoints.[7] One of the characterizing features of optimal portfolio allocation under CRRA utility is that the proportion of wealth to be invested in the risky asset(s) does not depend on the wealth level. Therefore, the time evolution of the wealth of agent i is conveniently represented by Eq. 5.2, namely

$$W_{i,t+1} = W_{i,t}[R + (r_{t+1} - r_f)\pi_{i,t}]$$

where the optimal wealth proportion invested in the risky asset, $\pi_{i,t}$, turns out to be independent of $W_{i,t}$. As a consequence, unlike the case of CARA utility, the optimal demand in quantity terms ($z_{i,t} = \pi_{i,t} W_{i,t}/P_t$) is proportional, *ceteris paribus*, to $W_{i,t}$. Since prices depend on aggregate excess demand via the particular price-setting mechanism, the dynamics of prices (and returns) will be affected by the dynamics of the wealth of each agent, that is, prices and wealth coevolve over time. This results in general in dynamical systems of higher dimension than in the case of CARA utility. The discussion of this section draws on the two models presented by Chiarella and He (2001) and Chiarella, Dieci, and Gardini (2006). The models are rather stylized but have the advantage of allowing some analytical tractability. They share a number of common features related to the formation of optimal portfolios and the evolution of wealth dynamics.

5.4.1. Optimal Portfolio and Wealth Dynamics

As a solution to agent-type i's maximization problem, $\max_{\pi_{i,t}} \mathbb{E}_{i,t}[U_i(W_{i,t+1})]$, we obtain the approximation (see Chiarella and He, 2001, for details):

[7] As reported, for instance, by Levy, Levy, and Solomon (2000), the results of various experiments support decreasing absolute risk aversion (DARA) rather than CARA preferences, which is consistent with constant relative risk aversion (CRRA). Moreover, as pointed out by Campbell and Viceira (2002), the absence of long-term trends in certain financial variables, such as interest rates and risk premia in face of long-run economic growth, implies that relative risk aversion is almost independent of wealth.

$$\pi_{i,t} = \frac{\mathbb{E}_{i,t}(r_{t+1}) - r_f}{\gamma_i \mathbb{V}_{i,t}(r_{t+1})} \tag{5.32}$$

It is possible to group the population of heterogeneous traders into different agent types so that agents within the same group are homogeneous in terms of their beliefs and risk attitudes: The differences in wealth across agents of the same type have no influence on the dynamics, because the proportion π_t will be the same for all investors within the same group, so that only total wealth of each group (or equivalently, the average wealth of each agent type) matters for the time evolution of the system.[8]

The models formulated in terms of the price, P_t, and the wealth of each group, $W_{i,t}$, may result in dynamical systems where such variables are growing over time. However, the models can be rewritten in terms of return $r_t := (P_t - P_{t-1} + y_t)/P_{t-1}$ (or "price return" $\kappa_t := (P_t - P_{t-1})/P_{t-1}$) and wealth shares $\omega_{i,t} := W_{i,t}/W_t$, $W_t := \sum_i W_{i,t}$, which allows us to deal with "stationary" systems, by which we mean dynamic models that admit steady-state solutions. In particular, the wealth share of each group evolves according to the same basic equation:

$$\omega_{i,t+1} = \frac{\omega_{i,t}\left[R + \pi_{i,t}(r_{t+1} - r_f)\right]}{\sum_j \omega_{j,t}\left[R + \pi_{j,t}(r_{t+1} - r_f)\right]} \tag{5.33}$$

which can be immediately obtained from Eq. 5.2 by noticing that

$$\omega_{i,t+1} = \frac{W_{i,t+1}}{W_{t+1}} = \frac{W_t}{W_{t+1}}\omega_{i,t}\left[R + \pi_{i,t}(r_{t+1} - r_f)\right]$$

where the quantity W_{t+1}/W_t can be expressed as

$$\frac{W_{t+1}}{W_t} = R + (r_{t+1} - r_f)\sum_j \omega_{j,t}\pi_{j,t} \tag{5.34}$$

The two models in Chiarella and He (2001) and Chiarella, Dieci, and Gardini (2006) differ in terms of the market-clearing mechanism and scheme for agents' expectation formation, which affects the way the quantities $\pi_{i,t}$ are determined. The difference in the assumed price-setting rules will determine different laws of motion for the return r_t in the two models, as demonstrated in the following discussion.

5.4.2. Price and Wealth Behavior with a Walrasian Auctioneer

The reader should keep in mind that under the Walrasian mechanism, agents' information set at time t consists of realized returns up to time $t-1$. In period t, agents of type i

[8]Modeling the population of CRRA traders as grouped into different types, which is convenient when the population proportion of each group is fixed over time, may present some difficulties if agents are allowed to switch among groups, depending on some fitness measure. The case of time-varying proportions is discussed in Chiarella and He (2005) and Chiarella and He (2008).

form their beliefs about the first and second moment of next period's return according to

$$\mathbb{E}_{i,t}(r_{t+1}) = f_i(r_{t-1}, r_{t-2}, \ldots, r_{t-L_i}) \tag{5.35}$$

$$\mathbb{V}_{i,t}(r_{t+1}) = g_i(r_{t-1}, r_{t-2}, \ldots, r_{t-L_i}) \tag{5.36}$$

where L_i are integers and f_i, g_i are deterministic functions which can differ across investors. Assuming log-utility, $U_i(W) = \ln W$, an approximate solution for the optimum investment proportion at time t is given by $\pi_{i,t} = \big(\mathbb{E}_{i,t}(r_{t+1}) - r_f\big)/\mathbb{V}_{i,t}(r_{t+1})$, which corresponds to Eq. 5.32 with $\gamma_i = 1$. Such an investment proportion thus turns out to be determined once the price and dividend history $(P_{t-1}, P_{t-2}, \ldots; y_{t-1}, y_{t-2}, \ldots)$ up to time $t-1$ is known. The dividend yield $v_t := y_t/P_{t-1}$ is assumed to follow an IID normal process with mean $\overline{v}$ and variance σ_v^2. Note that agents are not assumed to form beliefs about dividends but to look directly at the total return r_{t+1}.

Market-Clearing

Assuming a constant supply N of shares in the market, under a Walrasian auctioneer scenario the market-clearing condition at time t is

$$\sum_i z_{i,t} = \sum_i \frac{\pi_{i,t} W_{i,t}}{P_t} = N$$

from which $P_t = \frac{1}{N}\sum_i \pi_{i,t} W_{i,t}$, and therefore the return satisfies

$$1 + r_{t+1} = \frac{\sum_i \pi_{i,t+1}\omega_{i,t+1}}{\sum_i \pi_{i,t}\omega_{i,t}} \frac{W_{t+1}}{W_t} + v_{t+1} \tag{5.37}$$

where W_{t+1}/W_t is given by Eq. 5.34, whereas the quantity $\sum_i \pi_{i,t+1}\omega_{i,t+1}$ can be rewritten as[9]

$$\sum_i \pi_{i,t+1}\omega_{i,t+1} = \frac{\sum_i \pi_{i,t+1}\omega_{i,t}\left[R + \pi_{i,t}(r_{t+1} - r_f)\right]}{\sum_i \omega_{i,t}\left[R + \pi_{i,t}(r_{t+1} - r_f)\right]} \tag{5.38}$$

By substituting Eqs. 5.34 and 5.38 into Eq. 5.37 and solving for r_{t+1}, one finally obtains

$$r_{t+1} = r_f + \frac{\sum_i \omega_{i,t}\left[R(\pi_{i,t+1} - \pi_{i,t}) + \pi_{i,t} v_{t+1}\right]}{\sum_i \pi_{i,t}\omega_{i,t}(1 - \pi_{i,t+1})} \tag{5.39}$$

Eqs. 5.33 and 5.39 form a stationary dynamical system in terms of return of the risky asset and wealth shares of each group of traders. The dimension of the system obviously depends on the number of groups and the way beliefs (Eqs. 5.35 and 5.36) are specified for each group.

[9]To obtain this, multiply both sides of Eq. 5.33 by $\pi_{i,t+1}$ and sum across agent types.

Wealth Dynamics

The model (Eqs. 5.33 and 5.39) can be expressed in an equivalent form, which incorporates explicitly information on the relative size of each group in the market, in terms of number of traders. The "market fraction" of a group may represent an important "bifurcation" parameter for understanding the behavior of BRHA models (see, for example, He, 2003 and Dieci, Foroni, Gardini, and He, 2006). Moreover, the explicit consideration of such proportions is required in order to build models (like the ones developed in Section 5.3) where the popularity of each trading strategy varies over time, depending on some fitness measure.

Assume that agents can be grouped into I types according to their beliefs. The parameters that represent the proportions of each group can be incorporated into the model by assuming that group i consists of by l_i agents, and by defining $\overline{W}_{i,t} := W_{i,t}/l_i$ as the average wealth of traders of group $i = 1, 2, \ldots, I$. We then introduce a new variable which represents the "average" wealth share, $\overline{\omega}_{i,t} := \overline{W}_{i,t}/\overline{W}_t$, where $\overline{W}_t := \sum_{i=1}^{I} \overline{W}_{i,t}$, as well as the new parameters $n_i := l_i / \sum_j l_j$, which represent the relative sizes of the groups. Simple algebra reveals that $\bar{\omega}_{i,t}$ turns out to be related to $\omega_{i,t}$, the wealth share of group i, according to

$$\omega_{i,t} = \frac{n_i \bar{\omega}_{i,t}}{\sum_{j=1}^{I} n_j \bar{\omega}_{j,t}}$$

As a consequence, the law of motion (Eq. 5.33) for the wealth shares can be equivalently expressed with respect to the "average" wealth shares, as

$$\bar{\omega}_{i,t+1} = \frac{\bar{\omega}_{i,t}\left[R + \pi_{i,t}(r_{t+1} - r_f)\right]}{\sum_{j=1}^{I} \bar{\omega}_{j,t}\left[R + \pi_{j,t}(r_{t+1} - r_f)\right]} \tag{5.40}$$

while the return equation (5.39) is easily converted into

$$r_{t+1} = r_f + \frac{\sum_{i=1}^{I} n_i \bar{\omega}_{i,t}\left[R(\pi_{i,t+1} - \pi_{i,t}) + \pi_{i,t} \nu_{t+1}\right]}{\sum_{i=1}^{I} \pi_{i,t} n_i \bar{\omega}_{i,t}(1 - \pi_{i,t+1})} \tag{5.41}$$

where the parameters n_i appear explicitly.

Return Behavior

We remind the reader that the dynamical system (Eqs. 5.40 and 5.41) is stochastic since the dividend yield ν_t is assumed to be so. Chiarella and He (2001) analyze the "deterministic skeleton" of this noisy model[10] in the homogeneous case and in a number of heterogeneous cases with two agent types (fundamentalists with different risk attitudes, fundamentalists versus contrarians, fundamentalists versus trend followers, and trend

[10]Obtained by assuming a constant dividend yield equal to the average.

followers with different strategies). These types are modeled via different specifications of $\mathbb{E}_{i,t}(r_{t+1})$ in Eq. 5.35. For all agent types, the conditional variance $\mathbb{V}_{i,t}(r_{t+1})$ is an increasing and bounded concave function of the historical variance, indicating that the agents are cautious about reacting to large historical volatility. In all these cases, multiple steady states emerge, but in general no more than one of them can be locally asymptotically stable. Moreover, bifurcation analysis with respect to extrapolation rates shows that stability switches across steady states following a "quasi-optimal" selection principle, in the sense that the steady state having relatively higher return tends to dominate the market in the long run.[11]

Another feature of the model is that when agents extrapolate weakly, the return converges to one of the fixed equilibra, whereas when agents extrapolate strongly, the fixed equilibria become unstable and can generate periodic cycles, quasi-periodic orbits, and strange attractors for the return series, leading to rich dynamics for the returns and wealth proportions among heterogeneous investors. This suggests that a change of extrapolation rates from time to time might enable the return series to switch among different asymptotic states, thus providing a simple qualitative explanation of changes of the "market environment." Stochastic experiments are then performed to assess the effect of such changes, by introducing independent Poisson jump processes in the extrapolation rates, together with a stochastic normally distributed dividend yield. Simulation results are compared to the statistics of the S&P 500 and demonstrate the capacity of the model to generate the stylized facts observed in financial markets, such as volatility clustering and realistic values of skewness and kurtosis.

In Chiarella and He (2008), the basic model (Eqs. 5.33 and 5.39) has been extended in the direction of allowing the population proportions n_i to evolve endogenously according to a "discrete choice" model similar to the one described by Eq. 5.21. The analysis of the resulting adaptive model focuses on the profitability and the alternating popularity of "momentum" and "contrarian" trading strategies.

5.4.3. Price and Wealth Behavior with a Market Maker

The approach under the market-maker market-clearing mechanism in this subsection is different from the one in the previous subsection. In particular, dividends are expected to grow at a non-negative rate, which imposes a growth rate on the fundamental price. All agents are assumed to take account of this growth rate when forming expectations.

The Fundamental Price

Agents are assumed to share common and correct expectations about the dividend process $\{y_t\}$, so that for any agent i, $\mathbb{E}_{i,t}(y_{t+k}) = \mathbb{E}_t(y_{t+k})$, for $k = 1, 2, \ldots$. In addition the

[11] A discussion of selection between multiple steady states is contained in the closely related work of Anufriev, Bottazzi, and Pancotto (2006), who direcly model the agents' investment choices as smooth functions of the beliefs about expected asset returns and variances, within a framework that is consistent with CRRA utility.

evolution of the dividends is assumed to satisfy

$$\mathbb{E}_t(y_{t+k}) = (1+\varphi)^k y_t \quad k = 1, 2, \ldots \tag{5.42}$$

where φ, $0 \le \varphi < r_f$, represents a non-negative expected dividend growth rate. A benchmark notion of fundamental price, obtained by discounting the expected stream of future dividends, is given by

$$P_t^* = \sum_{k=1}^{\infty} \frac{\mathbb{E}_t(y_{t+k})}{(1+r_f)^k} = \frac{(1+\varphi)y_t}{(r_f - \varphi)} \tag{5.43}$$

Eq. 5.43 implies that $\mathbb{E}_t(P_{t+1}^*) = (1+\varphi)P_t^*$ (and therefore φ also represents the expected growth rate of the fundamental), as well as $\mathbb{E}_t(y_{t+1}) = (1+\varphi)y_t = (r_f - \varphi)P_t^*$.[12] When it comes to simulating the stochastic version of the model, it will be assumed that dividends evolve according to

$$y_{t+1} = (1 + \varphi + \sigma_\epsilon \epsilon_{t+1}) y_t \tag{5.44}$$

so that the fundamental price follows a similar process, $P_{t+1}^* = (1+\varphi+\sigma_\epsilon\epsilon_{t+1})P_t^*$, where $\epsilon_t \sim \mathcal{N}(0,1)$ are IID random shocks and $\sigma_\epsilon > 0$ represents the standard deviation of the dividend (and fundamental) growth rate.[13,14]

Heterogeneous Beliefs: The Fundamentalists and Trend Followers

Two groups of agents, fundamentalists and trend followers, are considered, together with a market maker who mediates transactions. Both the fundamentalists and the market maker are assumed to form correct expectations about the change in the fundamental price, $P_{t+1}^* - P_t^*$, according to

$$\mathbb{E}_t\left(P_{t+1}^* - P_t^*\right) = \varphi P_t^*$$

Fundamentalists, believing that prices will revert to the fundamental, form price expectations according to

$$\begin{aligned}\mathbb{E}_{f,t}(P_{t+1} - P_t) &= \mathbb{E}_t(P_{t+1}^* - P_t^*) + d_f(P_t^* - P_t)\\ &= \varphi P_t^* + d_f(P_t^* - P_t)\end{aligned}$$

[12]For $\varphi > 0$, Eq. 5.43 is known in the finance literature as the Gordon dividend growth model. Such a model seeks to forecast equity prices on the basis of assumptions about the future dividend growth rate. Despite its simplicity, the model plays an important role in applied finance for valuing indices and assessing the behavior of price/earnings and dividend/price ratios (see, for example, Shiller, 2003). Extensions of the basic model have also been used in the behavioral finance literature (for example, Barsky and De Long, 1993).

[13]More generally, Eq. 5.42 is consistent with the assumption of dividends following a geometric random walk of the type $\ln y_{t+1} = \ln y_t + \chi + \xi_{t+1}$, where χ is some constant and $\{\xi_t\}$ is IID noise with $\mathbb{E}(\xi_t) = 0$, provided that one sets: $1 + \varphi := \mathbb{E}\left[\exp(\chi + \xi_{t+1})\right]$.

[14]Note that with the CRRA utility function it seems natural to have the fundamental price follow a geometric random walk, due essentially to the fact that a growing wealth process coevolves with the price process in this framework. This is in contrast to the CARA utility case in Section 5.3, where a stationary process for the fundamental is more natural.

where $0 < d_f < 1$ captures the expected speed of mean reversion. It is assumed that $\mathbb{V}_{f,t}(r_{t+1}) = \mathbb{V}_{f,t}((P_{t+1} - P_t + y_{t+1})/P_t) = \sigma_f^2$, is constant over time. Given their CRRA coefficient γ_1, the optimal investment proportion (Eq. 5.32) for the fundamentalists then becomes

$$\pi_{f,t} = \frac{1}{P_t}\frac{(d_f + \varphi)P_t^* - (d_f + r_f)P_t + (1+\varphi)y_t}{\gamma_1 \sigma_f^2} \tag{5.45}$$

Trend followers do not rely on the knowledge of the fundamental price but try to extrapolate past price movements into the future and form expectations about next period's relative price change $(P_{t+1} - P_t)/P_t$ according to an adaptive scheme. More precisely, by defining $\kappa_{c,t}^e := \mathbb{E}_{c,t}\left[(P_{t+1} - P_t)/P_t\right]$ and $d_c \in (0, 1)$ as the chartist extrapolation parameter, the expectations of trend followers are updated according to the rule

$$\kappa_{c,t}^e = (1 - d_c)\kappa_{c,t-1}^e + d_c \frac{P_t - P_{t-1}}{P_{t-1}} \tag{5.46}$$

From the definition of r_{t+1} and the fact that $\mathbb{E}_t(y_{t+1}) = (1+\varphi)y_t$ it follows that

$$\mathbb{E}_{c,t}(r_{t+1}) = \kappa_{c,t}^e + \frac{(1+\varphi)y_t}{P_t}$$

Despite their attempt to exploit price trends for speculative purposes, it is assumed that trend followers perceive the increasing risk associated with high absolute returns. This is taken into account by assuming that trend followers have state-dependent beliefs about the variance of the return. More precisely, at each time step the estimated variance $\sigma_{c,t}^2 := \mathbb{V}_{c,t}(r_{t+1})$ depends positively on the magnitude of the expected excess return, $\left|\mathbb{E}_{c,t}(r_{t+1}) - r_f\right|$, or put differently, it is a U-shaped function of the quantity $\left(\mathbb{E}_{c,t}(r_{t+1}) - r_f\right)$. As a consequence, the chartist investment fraction in the risky asset turns out to be an increasing and sigmoid-shaped function of the expected excess return, for which Chiarella, Dieci, and Gardini (2006) choose the following specification

$$\pi_{c,t} = \frac{\mathbb{E}_{c,t}(r_{t+1}) - r_f}{\gamma_2 \sigma_{c,t}^2} = \frac{1}{\gamma_2 v_c \theta}\tanh\left\{\theta\left[\kappa_{c,t}^e + \frac{(1+\varphi)y_t}{P_t} - r_f\right]\right\} \tag{5.47}$$

where the parameter γ_2 is the trend followers' CRRA coefficient, v_c represents a minimum level for variance beliefs, and $\theta > 0$ is a constant.

Market Equilibrium

The desired asset holding for agent type i, in number of shares, is given by $z_{i,t} = \pi_{i,t} W_{i,t}/P_t$. Assuming that the risky asset is in zero net supply,[15] the agents' excess

[15] A more general case with positive supply is sketched in Chiarella, Dieci, and Gardini (2006). In such a case, the discount rate in the fundamental price Eq. 5.43 is no longer the risk free rate r_f, but it must include a risk premium that is required by traders to hold a positive amount of the risky asset in equilibrium.

demand turns out to be given by

$$z_t = \frac{1}{P_t}(\pi_{f,t} W_{f,t} + \pi_{c,t} W_{c,t})$$

Under the assumed market-maker scenario, the price adjustment by the market maker takes into account both the expected change φP_t^* of the underlying fundamental, and the sign and magnitude of the excess demand. A sufficiently general specification of the price-setting rule can be found in Chiarella, Dieci, and Gardini (2006). For simplicity, we here assume that the market maker adjusts the market price according to

$$\frac{P_{t+1} - P_t}{P_t} = \varphi \frac{P_t^*}{P_t} + \mu \frac{\pi_{f,t} W_{f,t} + \pi_{c,t} W_{c,t}}{W_t} \tag{5.48}$$

where $\mu > 0$ and $W_t := W_{f,t} + W_{c,t}$.

The resulting discrete-time dynamical system consists of the price-setting rule Eq. 5.48, together with the dynamic Eqs. 5.46 and 5.2 for $\kappa_{c,t}^e$ and $W_{i,t}$, $i \in \{f, c\}$, respectively. It is possible to reduce the deterministic skeleton[16] of the model to a dynamical system in terms of expected and actual price return, $\kappa_{c,t}^e := \mathbb{E}_{c,t}\left[(P_{t+1} - P_t)/P_t\right]$ and $\kappa_{t+1} := (P_{t+1} - P_t)/P_t$, the fundamental to price ratio $\lambda_t := P_t^*/P_t$, and fundamentalist wealth share $\omega_{f,t} := W_{f,t}/W_t$, according to

$$\kappa_{t+1} = \varphi \lambda_t + \mu\left[\omega_{f,t}\pi_{f,t} + (1 - \omega_{f,t})\pi_{c,t}\right] \tag{5.49}$$

$$\lambda_{t+1} = \lambda_t \frac{1 + \varphi}{1 + \kappa_{t+1}} \tag{5.50}$$

$$\kappa_{c,t+1}^e = (1 - d_c)\kappa_{c,t}^e + d_c \kappa_{t+1} \tag{5.51}$$

$$\omega_{f,t+1} = \omega_{f,t} \frac{R + \pi_{f,t}\left[\kappa_{t+1} + (r_f - \varphi)\lambda_t - r_f\right]}{\sum_{i \in \{f,c\}} \omega_{i,t}\left[R + \pi_{i,t}\left(\kappa_{t+1} + (r_f - \varphi)\lambda_t - r_f\right)\right]} \tag{5.52}$$

where

$$\pi_{f,t} = \frac{(d_f + r_f)(\lambda_t - 1)}{\gamma_1 \sigma_f^2}$$

$$\pi_{c,t} = \frac{1}{\gamma_2 v_c \theta} \tanh[\theta(\kappa_{c,t}^e + (r_f - \varphi)\lambda_t - r_f)]$$

Note in particular that the wealth Eq. 5.52 is equivalent to Eq. 5.33, because (in the deterministic skeleton) $y_{t+1} = (r_f - \varphi)P_t^*$ and therefore $r_{t+1} = (P_{t+1} - P_t)/P_t + y_{t+1}/P_t = \kappa_{t+1} + (r_f - \varphi)\lambda_t$. Note also that Eqs. 5.49–5.52 turns out to be a

[16]Obtained by assuming that dividends and fundamental price evolve deterministically according to agents' expectations, that is, $P_{t+1}^* = (1 + \varphi)P_t^*$.

three-dimensional dynamical system since κ_{t+1} (which appears at the right side of Eqs. 5.50–5.52) depends on λ_t, $\kappa^e_{c,t}$, $\omega_{f,t}$, the state variables at time t.

Dynamical Behavior

The model (Eqs. 5.49–5.52) admits two types of steady states, and both can be locally asymptotically stable for particular configurations of the parameters. They are denoted as *fundamental steady states* and *nonfundamental steady states*, respectively. The former are characterized by[17] $\overline{\lambda} = 1$, $\overline{\kappa} = \overline{\kappa}^e_c = \varphi$, any $\overline{\omega}_f \in [0, 1]$. Thus, at a fundamental steady state the price reflects the fundamental and grows at the dividend growth rate, the excess demand is zero, whereas any long-run wealth distribution among agents is possible, depending on the initial conditions.

The nonfundamental steady states are characterized by $\overline{\lambda} = 0$, $\overline{\omega}_f = 0$, $\overline{\kappa} = \overline{\kappa}^e_c = \overline{\psi} > r_f > \varphi$, where $\overline{\psi}$ solves the equation

$$\frac{\overline{\psi}}{\mu} = \frac{1}{\gamma_2 v_c \theta} \tanh\left[\theta(\overline{\psi} - r_f)\right]$$

Therefore, at a nonfundamental steady state, the price grows faster than the fundamental, the market is dominated by trend followers, and there is a permanent positive excess demand. Such attracting nonfundamental equilibria represent a sort of "deterministic bubble" that lasts forever. Of course, such an outcome cannot be sustained in the long run in the real world and is due to the highly simplified setup of the model being considered here (absence of constraints from the side of market maker's inventories, no account of consumption and its effect on wealth dynamics) as well as to its deterministic nature (once noise is added to the system, the behavior of price and wealth becomes more realistic and often phases of booms alternate with phases of crashes in an unpredictable way). Nevertheless, the existence and stability of such nonfundamental equilibria provide the important information that bubbles may arise for particular initial conditions and under particular sets of parameters.

To provide some examples about the rich behavior of the system, depending on the key parameter and the initial conditions, we give some plots illustrating the effect of the risk aversion parameter (Figure 5.7) and the role of initial price and wealth shares in various situations (Figures 5.8 and 5.9). Figure 5.7 is related to a particular case where the market is dominated by trend followers ($\omega_{f,0} = 0$, and thus $\omega_{f,t} = 0$ for any t) and shows the projections in the plane of the variables κ, λ, of different orbits obtained with the same initial condition (close to fundamental) under decreasing values of the chartist risk-aversion coefficient γ_2. The fundamental steady state (F) changes from a stable (a) to an unstable focus (b) through a (supercritical) Neimark-Sacker bifurcation, which creates a stable closed orbit. The amplitude of the oscillations becomes wider for lower risk aversion (c) until the attractor "collapses" onto a stable nonfundamental steady state (NF), with a permanent deviation of the price away from the fundamental (d). Similar

[17] An overbar is used here to denote steady-state quantities.

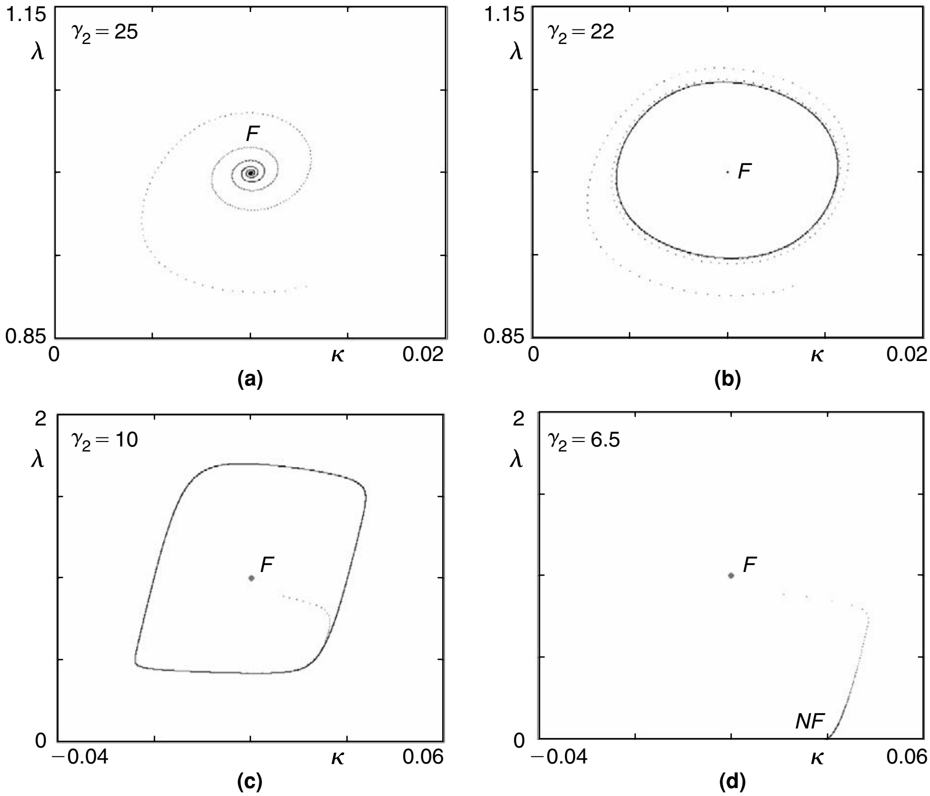

FIGURE 5.7 CRRA trend followers with a market maker. (a) Stability and (b, c, d) instability of fundamental steady state for decreasing values of the risk aversion parameter γ_2, and for $r_f = 0.02$, $\varphi = 0.01$, $\mu = 0.05$, $v_c = 0.002$, $\theta = 100$, $d_c = 0.25$.

phase-space transitions can be obtained by increasing the extrapolation parameter d_c or the speed of price adjustment μ. Although we have set $\omega_{f,0} = 0$ here, the same attractors of Figure 5.7 can be reached also from initial conditions with positive wealth share of the fundamentalists, thus suggesting that fundamentalists do not accumulate in general more wealth than chartists and that chartists may even dominate in the long run.

A related important question concerns the effect of the initial conditions on the asymptotic behavior of the model. For instance, it can be shown numerically that a stable nonfundamental steady state may coexist with a continuum of attracting fundamental steady states, with the phase space shared between different basins of attractions. Figure 5.8 reports an example of such basins of attraction, in the (ω_f, λ) plane. The basin structure depends on the particular parameter set used in the simulation. For instance, higher values of the chartist parameter d_c lead to an increase of the size of the basin of the nonfundamental steady state.

Figure 5.9a reports another situation of coexisting attractors, a fundamental steady state and a periodic orbit. Different from the examples in Figures 5.7 and 5.8, in this

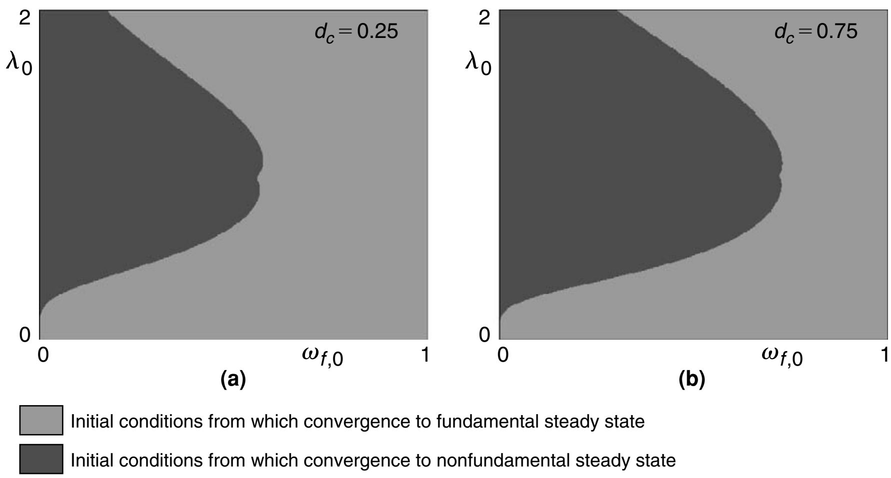

FIGURE 5.8 CRRA heterogeneous agents with a market maker. Basins of attraction of fundamental and nonfundamental steady states for different reaction speeds of chartists—$d_c = 0.25$ (a), $d_c = 0.75$ (b). Parameters are set as follows: $r_f = 0.02$, $\varphi = 0.01$, $\mu = 0.05$, $\sigma_f^2 = v_c = 0.002$, $\theta = 100$, $d_f = 0.3$, and $\gamma_1 = 20$, $\gamma_2 = 6.5$.

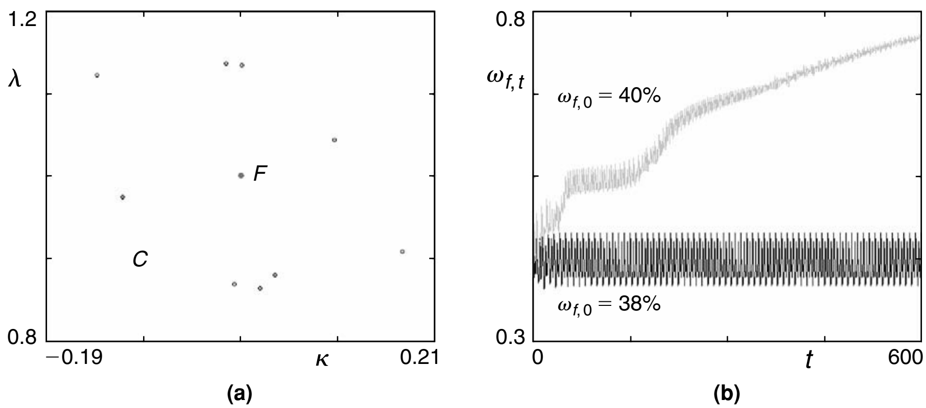

FIGURE 5.9 CRRA heterogeneous agents with a market maker for $r_f = 0.02$, $\varphi = 0.01$, $\mu = 0.1$, $\sigma_f^2 = v_c = 0.004$, $\theta = 50$, $d_f = d_c = 0.8$, $\gamma_1 = 10$, and $\gamma_2 = 2.5$. (a) Coexisting attractors and (b) sensitivity with respect to initial wealth shares.

case each scenario allows both types of agents to survive in the long run, with different behavior of the wealth shares. As shown in Figure 5.9b, for an initial wealth share $\omega_{f,0} = 38\%$, the dynamics converge to a periodic orbit and ω_f eventually fluctuate approximately in the range $[38\%, 47\%]$, whereas for a slightly increased initial wealth share, $\omega_{f,0} = 40\%$, the system converges to a stationary state in which the fundamentalist wealth proportion is much higher. This example also illustrates the extreme sensitivity of the dynamic evolution of the model to the initial values.

The interaction of the nonlinear deterministic dynamics, often characterized by coexisting attractors with simple external noise processes, is able to generate phases of booms, where the price grows faster than the fundamental and the fundamentalist wealth share rapidly declines, followed by crashes, where the price is attracted again toward the fundamental, and the fundamentalist wealth share returns to higher levels. This can be seen from the noisy version of the model, where it is assumed that the dividend and the fundamental price grow at the rate $\varphi + \sigma_\epsilon \epsilon_t$, according to Eq. 5.44. Furthermore, a stochastic term $P_t \sigma_\xi \xi_t$, where $\sigma_\xi > 0$ and $\xi_t \sim \mathcal{N}(0, 1)$ are IID random disturbances representing the influence of noise traders, is added to the price setting equation. Also, such a noisy model can be reduced to a dynamical system formulated in price returns and wealth shares (see Chiarella, Dieci, and Gardini, 2006, for details).

Figure 5.10 presents sample paths of the prices (market price P_t and fundamental price P_t^*) and of the fundamentalist wealth share ($\omega_{f,t}$) as a function of time, under different initial conditions. The parameters are the same as in Figure 5.8a (coexistence of stable fundamental and nonfundamental steady states). Booms and crashes of the price, accompanied by jumps in the wealth shares, are observed when the system starts with a sufficiently small fundamentalist wealth share $\omega_{f,0}$, as is the case of panels (a) and (b). On the other hand, more regular price and wealth paths are obtained starting with a higher fundamentalist wealth proportion, as in the case represented in panels (c)

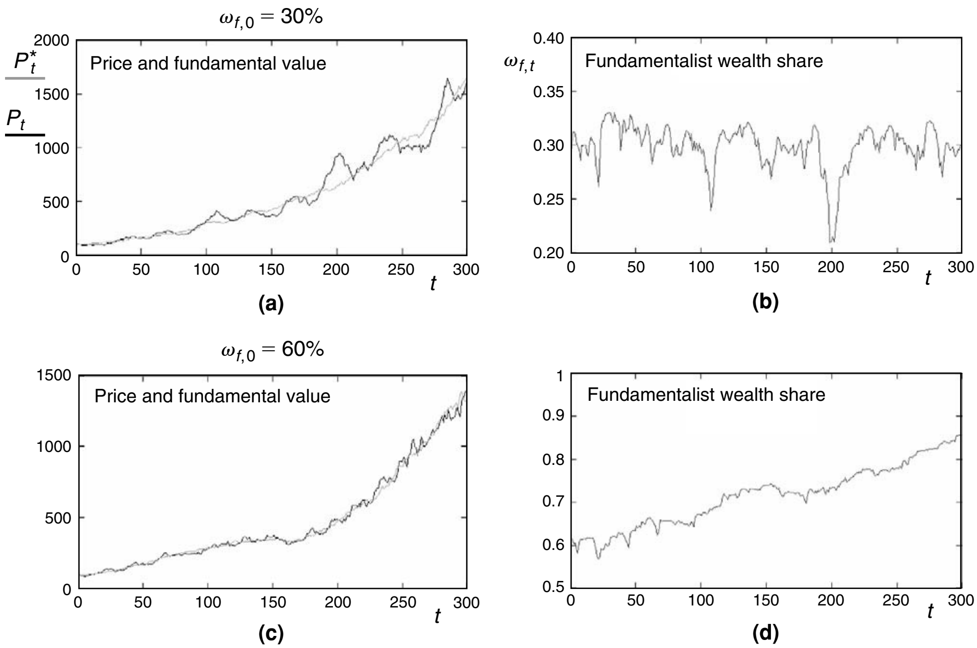

FIGURE 5.10 CRRA heterogeneous agents with a market maker. Price, fundamental price, and wealth shares in a simulation of the noisy model ($\sigma_\epsilon = 0.015$, $\sigma_\xi = 0.03$). Other parameters are as shown in Figure 5.8a. In (a) and (b), initial fundamentalist wealth share is low ($\omega_{f,0} = 30\%$), in (c) and (d), it is higher ($\omega_{f,0} = 60\%$).

and (d). These phenomena may be related to the structure of the basins of attraction of the underlying deterministic model. These simulations give just a glimpse of the effect of the interaction of noise with the nonlinear dynamic elements of BRHA models; this issue is further explored in the next two sections.

Summary

The analysis of this section has shown that under both market-clearing mechanisms, the BRHA framework with agents having CRRA utility exhibits a rich type of dynamics (compared to CARA utility). This framework more readily yields the growing price process that we observe in real markets, it also has a rich bifurcation structure with multiple basins of attraction and switching between fundamental and nonfundamental equilibria. Furthermore, the CRRA framework seems appropriate to discuss the long-run survival of competing trading strategies. However, because the price dynamics and wealth dynamics are intertwined, it is much more difficult to obtain analytical results, so in this case we need to rely much more on computer simulations.

5.5. EMPIRICAL BEHAVIOR

This section reviews some of the work that has been done on the empirics calibration of the BRHA class of models. In the formulation of the models in Sections 5.3 and 5.4 we have indicated that noise could also be incorporated into the modelling framework—for instance, a noisy fundamental, noisy dividends, general market noise due to news arrival and/or noise trading activity. It seems reasonable to expect that the real financial market is the outcome of both nonlinear and stochastic elements. Indeed, simulations reported in Chiarella, Dieci, and Gardini (2002), Chiarella and He (2002), He and Li (2008), among other authors, indicate that the said interaction can indeed lead to price and return distributions that have many of the characteristics of real financial time series. Furthermore, we know from more recent research (to be discussed in Section 5.7) that the stochastic version of the BRHA class of models can undergo stochastic bifurcations. In this section we use a simple market fraction model to show how it is able to reproduce fairly well many of the main stylized facts of real financial time series such as the S&P 500. We refer the reader to He and Li (2007, 2008) for more details on the issues addressed in this section.

5.5.1. Stylized Facts in the S&P 500

As a benchmark, we first briefly review the stylized facts of financial markets based on the S&P 500. Based on daily S&P 500 data from December 5, 1980, to February 23, 2007, Figure 5.11 displays plots including prices, the returns, and the corresponding density distributions, plus autocorrelation coefficients (ACs) of the returns, absolute and squared returns. They share some common stylized facts with high-frequency financial time series, including excess volatility (relative to the dividends

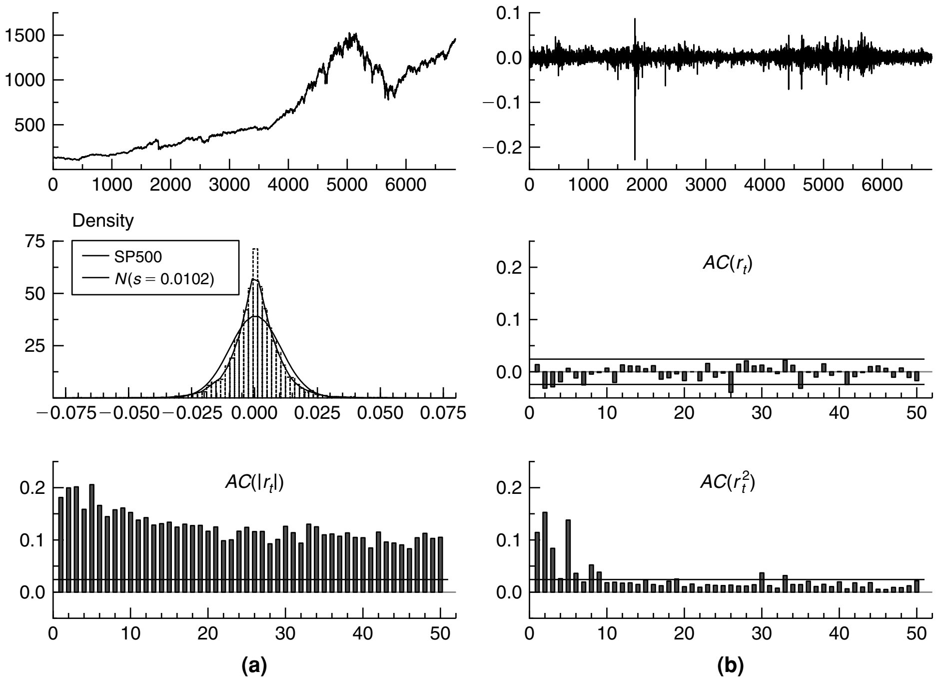

FIGURE 5.11 The daily S&P 500 data, Dec 5, 1980–Feb 23, 2007: prices (a) and returns (b). Return r_t distribution (compared to the corresponding normal distribution), the ACs for the returns $AC(r_t)$, the absolute $AC(|r_t|)$, and squared $AC(r_t^2)$ returns.

and underlying cash flows), volatility clustering (high/low fluctuations are followed by high/low fluctuations), skewness (either negative or positive), and excess kurtosis (compared to normally distributed returns). The returns show (almost) no significant autocorrelations, but the absolute returns and squared returns show slowly decaying autocorrelations. For a comprehensive discussion of stylized facts characterizing financial time series, we refer the reader to Pagan (1996).

Among the stylized facts, volatility clustering and power-law behavior (that is, insignificant ACs of raw returns and hyperbolic decline of ACs of the absolute and squared returns) have been extensively studied since the seminal paper of Ding, Granger, and Engle (1993). Recently, a number of universal power laws[18] have been found to hold in financial markets. This finding has spurred attempts at a theoretical explanation and the search for an understanding of the underlying mechanisms responsible for such power laws.[19] Among which, the herding models (see, for instance, Kirman, 1993; Lux and Marchesi, 1999; and Alfarano, Lux, and Wagner, 2005) and switching models (such as

[18]These include cubic power distribution of large returns, hyperbolic decline of the return autocorrelation function, temporal scaling of trading volume and multiscaling of higher moments of returns.

[19]We refer the reader to Lux (2004) for a recent survey on empirical evidence, models, and mechanisms of various financial power laws.

Brock and Hommes, 1998, and Gaunersdorfer and Hommes, 2000) have shown their potential to explain power-law behavior. Recently, He and Li (2007) consider the market fraction model established in He and Li (2008) and explore the potential of the model to generate the power-law feature observed in empirical data. The next subsection reports the empirical characteristics of the market fraction model obtained in He and Li (2007).

5.5.2. A Market Fraction Model and Its Stylized Behavior

To explain various aspects of financial market behavior and establish the connection between the stochastic model and its underlying deterministic system, He and Li (2008) consider a simple stochastic asset-pricing model, involving two types of agents (fundamentalists and trend followers) under a market-maker market-clearing scenario. The model is a simplified version of the models considered in Section 5.3. It is called a *market fraction model* since the market fraction difference $m = n_1 - n_2 \in [-1, 1]$ is assumed to be a constant parameter, rather than switching dynamically. The market price is determined by $P_{t+1} = P_t + (\mu/2)[(1+m)z_{1,t} + (1-m)z_{2,t}] + \tilde{\delta}_t$, where the demand $z_{i,t}$ is defined by Eq. 5.9 and the IID noise term $\tilde{\delta}_t \sim \mathcal{N}(0, \sigma_\delta^2)$ captures unexpected market news or the excess demand of noise traders.

The fundamentalists trade on the price deviation from the estimated fundamental value P_t^* whose relative return follows a normal distribution $P_{t+1}^*/P_t^* - 1 \sim \mathcal{N}(\bar{P}, \sigma_\epsilon^2)$, where $\bar{P} = \bar{y}/r$ is the expected long-run fundamental value and $\bar{y}$ is the mean of the dividend process y_t. Their beliefs are assumed to follow $\mathbb{E}_{1,t}(P_{t+1}) = P_t + d_f[\mathbb{E}_{1,t}(P_{t+1}^*) - P_t]$ and $\mathbb{V}_{1,t}(P_{t+1}) = \sigma_1^2$, where $d_f \in [0, 1]$ is the fundamentalists' speed of price adjustment toward the fundamental value and σ_1^2 is a constant. Without information on the fundamental value, the trend followers extrapolate the latest observed price change over a long-run sample mean of the price and adjust their variance estimate according to $\mathbb{E}_{2,t}(P_{t+1}) = P_t + d_c(P_t - u_t)$ and $\mathbb{V}_{2,t}(P_{t+1}) = \sigma_1^2 + b_2 v_t$. Here u_t and v_t are sample mean and variance, respectively, which follow the limiting geometric decay processes, $u_t = \delta u_{t-1} + (1-\delta)P_t$ and $v_t = \delta v_{t-1} + \delta(1-\delta)(P_t - u_{t-1})^2$, where $\delta \in [0, 1]$ is the decay rate, $d_c > 0$ measures the extrapolation rate from the trend followers, and $b_2 > 0$ measures the influence of the sample variance.

By assuming the dividend process $y_t \sim \mathcal{N}(\bar{y}, \sigma_y^2)$ with $\sigma_y^2 = r_f^2 \sigma_1^2$, the market fraction model is given by a four-dimensional stochastic system in (P_t, u_t, v_t, P_t^*) involving two noise processes, one for the fundamental value and one for the market noise. He and Li (2008) show that convergence of the market price to the fundamental value, the long- and short-run profitability of the two trading strategies, the survival of trend followers and various under- and overreaction autocorrelation patterns of the stochastic model can be explained by the dynamics, including its stability and bifurcations, of the underlying deterministic system. The model is also able to generate the stylized facts displayed by the return distribution. In the following discussion, we focus on an analysis of the volatility clustering and power-law mechanism of the model.

To understand the mechanism, we first illustrate the different impact of the two noise processes on the underlying deterministic price dynamics. Using the parameter set[20] in Table 5.1, we consider the four cases listed in Table 5.2. Case-00 corresponds to the deterministic case. Case-01 (Case-10) corresponds to the case with noisy fundamental price (noisy excess demand) only and both noise processes appear in Case-11. Figure 5.12 illustrates the price series for the four cases for a typical simulation. The

TABLE 5.1 Parameter Settings and Initial Values

d_f	d_c	α_1	α_2	μ	m	δ	b	σ_ϵ	σ_δ	P_0	P_0^*
0.1	0.3	0.8	0.8	2	0	0.85	1	0.01265	1	100	100

TABLE 5.2 Four Cases of the Noise Effect

Cases	Case-00	Case-01	Case-10	Case-11
$(\sigma_\delta, \sigma_\epsilon)$	(0, 0)	(0, 0.01265)	(1, 0)	(1, 0.01265)

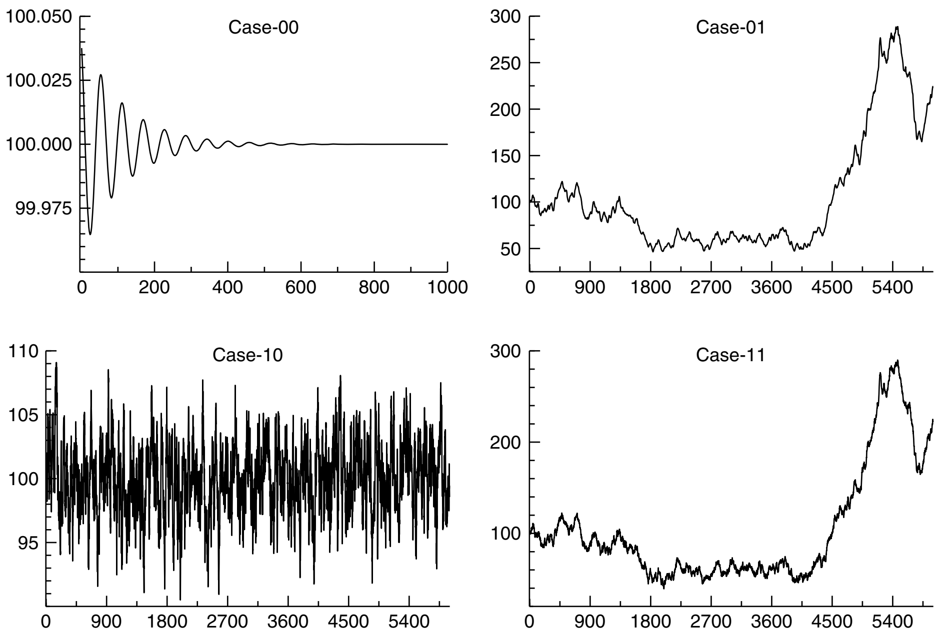

FIGURE 5.12 Time series of prices for the four cases in Table 5.2.

[20]In choosing the parameter values we have been mindful of the fact that most of the stylized facts are observed for high-frequency data (for example, daily) rather than for low-frequency (such as yearly) data. So we have chosen parameter values that are characteristic of daily returns—for instance, the market return volatility is 20% in annualized terms and the risk-free rate is $5\% p.a.$

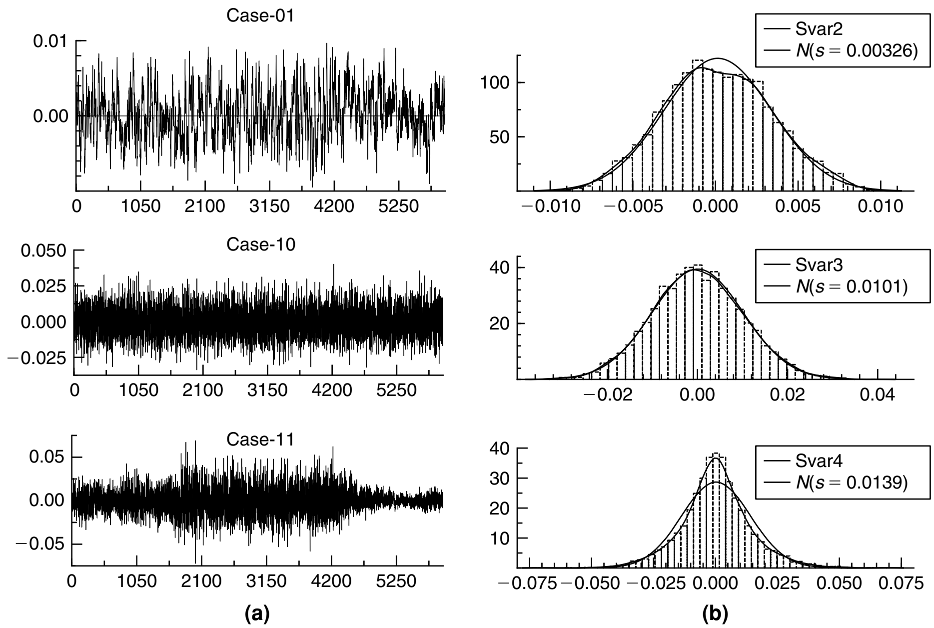

FIGURE 5.13 Return series and their density distributions (b) for Cases 01, 10, and 11 in Table 5.2.

corresponding return series and their density distributions are given in Figure 5.13 for the three cases involving noise. Figure 5.14 shows the ACs of returns, absolute returns and squared returns. For comparison purposes, the same set of noisy demand and fundamental processes is used in Case-11. Both Figures 5.13 and 5.14 show significantly different impacts of the different noise processes on the volatility.

In Case-01, the stochastic fundamental price process is the only external noise source. The market price displays a *strong underreaction* AC pattern of returns, which is characterized by the significantly positive decaying ACs shown in the top left panel in Figure 5.14. This significant AC pattern is also carried on to the AC patterns for the absolute and squared returns. In Case-10, the noisy excess demand is the only external noise source. The market price displays no volatility clustering, which is characterized by insignificant AC patterns for returns, the absolute and squared returns shown in the middle row in Figure 5.14. In Case-11, both the noisy excess demand and noisy fundamental price processes are present. In this case we observe relatively high kurtosis in Figure 5.13 and insignificant ACs for returns, but significant ACs for the absolute and squared returns shown in the bottom panel in Figure 5.14.

These results demonstrate that this simple market fraction model is able to generate realistic price behavior and appropriate power-law behavior for returns when both noise processes are present. The analysis indicates that the noisy demand plays a more important role on the insignificant AC patterns for the returns, whereas the noisy fundamental process plays a more important role on the significant AC patterns for the absolute and squared returns. We refer the reader to He and Li (2007) for an analysis of volatility

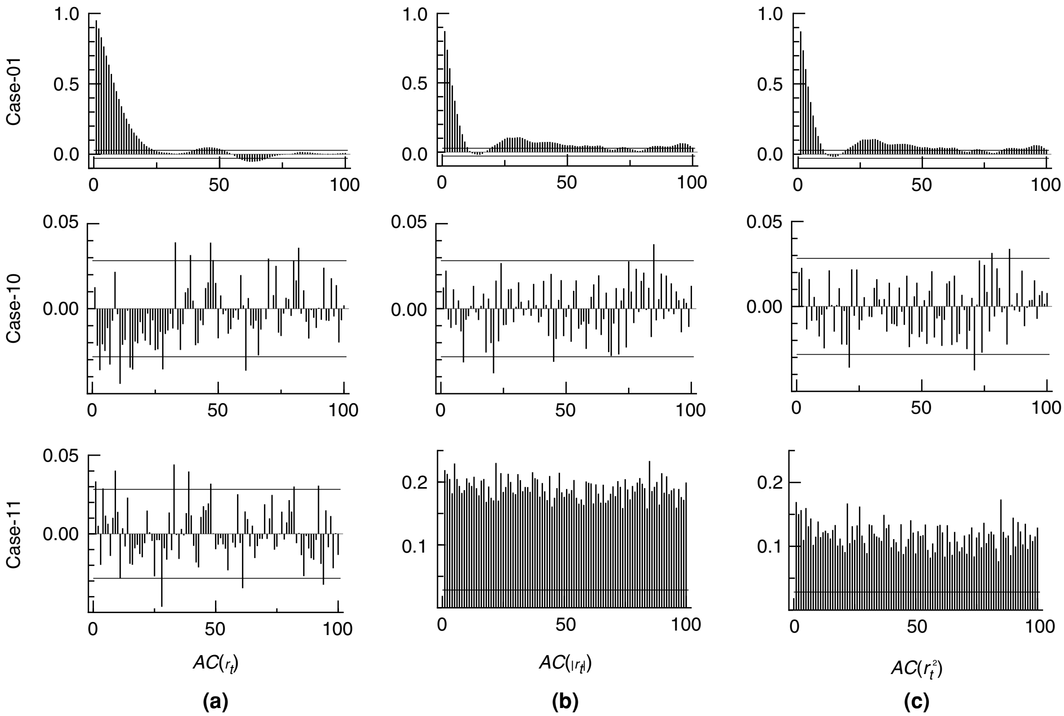

FIGURE 5.14 ACs of returns (a), absolute returns (b), and squared returns (c) for Cases 01, 10, and 11.

clustering and power-law behavior. The discussion in the next section is devoted to an econometric analysis of the power-law behavior.

5.5.3. Econometric Characterization of the Power-Law Behavior

Ding, Granger, and Engle, 1993 investigate autocorrelations of returns (and their transformations) of the daily S&P 500 index over the period 1928 to 1991 and find that the absolute returns and the squared returns tend to have very slow decaying autocorrelations and that the sample autocorrelations for the absolute returns are greater than those for the squared returns at every lag up to at least 100 lags. This AC feature indicates the long-range dependence or the power-law behavior in volatility. By running 1000 independent simulations, we estimate the average autocorrelation coefficients for the market fraction model with the parameter set of Table 5.2 and plot the ACs and their corresponding confidence intervals in Figure 5.15. We see that the patterns of decay of the autocorrelation functions of return, the squared return and the absolute return are quite similar to what we observed for those of the S&P 500. Besides the visual inspection of autocorrelations of r_t, r_t^2 and $|r_t|$, we can also construct models to estimate the decay rate d of the autocorrelations of r_t, r_t^2 and $|r_t|$. Based on the estimates given in He and Li (2007), there is clear evidence of a power law for the squared returns and the absolute returns, which are comparable to those of the actual data.

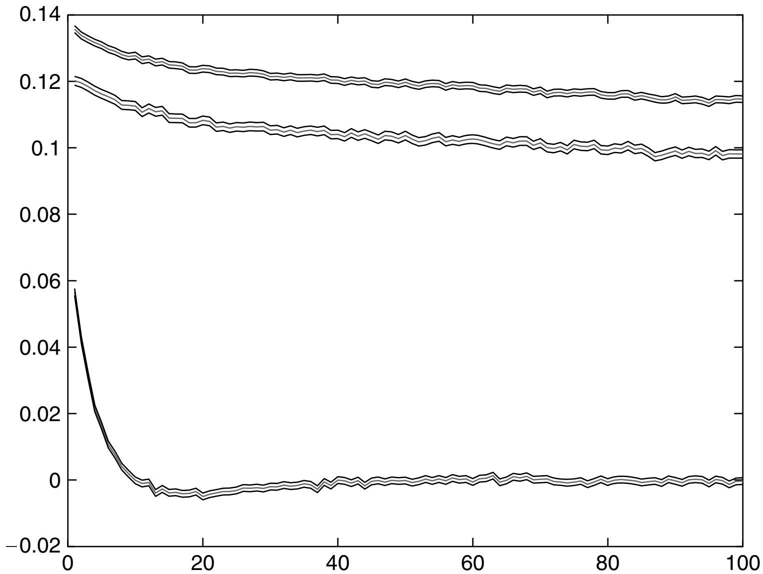

FIGURE 5.15 Autocorrelations of r_t (bottom), r_t^2 (middle), and $|r_t|$ (top) for the market function model.

Another striking feature of financial return series is *volatility clustering*. A number of econometric models of changing conditional variance have been developed to test and measure volatility clustering. Engle (1982) suggested a test where the null hypothesis is that the residuals of a regression model are IID and the alternative hypothesis is that the errors are ARCH(q). For the market fraction model, the null hypothesis is strongly rejected. For the market fraction model, Table 5.3 reports the estimates of the GARCH (1, 1) model (see Bollerslev, 1986). The results from the GARCH model are astonishingly similar to what one usually extracts from real market data: A small influence of the most recent innovation ($\alpha_1 < 0.1$) is accompanied by strong persistence of the variance coefficient ($\beta_1 > 0.9$). It is also interesting to observe that the sum of the coefficients $\alpha_1 + \beta_1 = 0.9928$ is close to one, which indicates that the process is close to an integrated GARCH (IGARCH) process. The GARCH implies that shocks to the conditional variance decay exponentially. However the IGARCH implies that the shocks to the conditional variance persist indefinitely.

In response to the finding that most financial time series are power-law volatility processes, Baillie, Bollerslev, and Mikkelsen (1996) consider the fractional integrated GARCH (FIGARCH) process, where a shock to the conditional variance dies out at a slow hyperbolic rate. Table 5.4 reports the estimates of the FIGARCH models for the market fraction model. For the estimates of the FIGARCH(1, d, 1), we see that the estimate of d is significantly different from zero and one. This is consistent with the well-known findings that the shocks to the conditional variance die out at a slow hyperbolic rate.

TABLE 5.3 The GARCH (1, 1) Parameter Estimates for the Market Fraction Model

$a \times 10^3$	b	$\alpha_0 \times 10^4$	α_1	β
0.0740	0.0725	0.0078	0.0260	0.9738
(0.2300)	(0.0139)	(0.0035)	(0.0032)	(0.0033)
47	77.1	17.7	100	100

Note: The numbers in parentheses are the standard errors, and the numbers in the last row are the percentages that the test statistics are significant at the 5% level over 1000 independent simulations. This also holds for Table 5.4.

TABLE 5.4 The FIGARCH (1, d, 1) Parameter Estimates for the Market Fraction Model

a	b	$\alpha_0 \times 10^4$	d	ϕ_1	β
0.0137	0.0769	0.3620	0.3797	0.3439	0.7933
(0.0010)	(0.0195)	(0.6112)	(0.0386)	(0.0281)	(0.0295)
41.2	72.6	35.6	87.6	83.1	98.5

Overall, we find that the market fraction model does provide a mechanism that can generate the long-range dependence in volatility observed in actual market data. The foregoing analysis shows that agent heterogeneity, risk-adjusted trend chasing through the geometric learning process, and the interplay of noisy fundamental and demand processes and the underlying nonlinear deterministic dynamics can be the source of power-law distributed fluctuations. The noisy demand plays an important role in the generation of insignificant ACs on returns, while the significantly decaying AC patterns of the absolute returns and squared returns seem to be more influenced by the noisy fundamental process. The estimates given are clear evidence that the power-law and (FI)GARCH features can arise even in this simple version of a BRHA model. The results also indicate the need to pursue more deeply the interaction of stochastic and nonlinear elements in BRHA models.

5.6. HETEROGENEITY IN A DYNAMIC MULTIASSET FRAMEWORK

The one risky/one risk-free asset model considered in previous sections is merely a first step to understanding price dynamics under heterogeneous agent interaction. Within a multiple risky asset framework, the way agents form and update their beliefs about the covariance structure also becomes an important factor in determining the dynamics of prices. A number of recent papers deal with the multiple risky asset decision problem in a mean-variance setup, within the BRHA paradigm.[21] We cite in particular Böhm and Chiarella (2005), Böhm and Wenzelburger (2005), and Wenzelburger (2004). Such models establish an overall framework, and focus on various aspects. In particular, the setup of Böhm and Chiarella (2005) is that of a multiasset dynamic CAPM with heterogeneous agents, though the dynamic impact of this heterogeneity has not been analyzed in detail. In Wenzelburger (2004), this setup has been extended to a model in which myopic agents are allowed to switch between different trading strategies similar to Brock and Hommes (1998). See also Raffaelli and Marsili (2006). Böhm and Wenzelburger (2005) apply the general setup to investigate the properties of efficient portfolios under heterogeneous beliefs.

With respect to the existing literature, we here make explicit and more general the updating rules of agents' beliefs, especially with regard to second-moment beliefs. The main point at issue concerns the dynamic effect of the assumed updating mechanism; namely, the question whether the endogenously varying beliefs about correlations—together with the dynamic portfolio allocation among multiple risky assets—may "stabilize" the asset markets, or rather tend to reinforce and to spread the price fluctuations that arise due to the interaction of heterogeneous agents. Put differently, we aim at understanding to what extent such mechanisms may generate interdependence between the price dynamics of different risky assets. A further difference with the existing literature is that a market-maker scenario is used in our setup, instead of a Walrasian auctioneer scenario. The model presented in the current section closely follows Chiarella, Dieci, and He (2007b).

5.6.1. Optimization of a Many Risky Asset Portfolio with Heterogeneous Beliefs

Portfolio Optimization of Many Risky Assets

The asset market is characterized by m risky assets, indexed by $j = 1, 2, \ldots, m$, and a risk-free asset, and by I different trader types, indexed by $i = 1, 2, \ldots, I$. For the risky asset j, the price at time t and the dividend paid in the trading period $(t-1, t)$ are denoted by $P_{j,t}$ and $y_{j,t}$, respectively. Using m-dimensional column vector notation, prices and dividends correspond to P_t and y_t, respectively, whereas the vector $R_t := (P_t + y_t - RP_{t-1})$ collects the excess returns (per share) in the time interval $(t-1, t)$. The vector $z_t^i = (z_{1,t}^i, z_{2,t}^i, \ldots, z_{m,t}^i)^\top$, where $\top$ stands for the transpose, denotes the portfolio held by agent type i in the trading period $(t, t+1)$, so that agent's wealth (W_t^i) evolves according to

$$W_{t+1}^i = RW_t^i + (z_t^i)^\top R_{t+1} = RW_t^i + (z_t^i)^\top (P_{t+1} + y_{t+1} - RP_t)$$

[21]Further, recent studies consider heterogeneous speculators switching among different risky investment opportunities and focus on the resulting comovements of asset prices (see Westerhoff, 2004, and Westerhoff and Dieci, 2006). Such studies do not assume a mean-variance setup explicitly but rather rely on particular behavioral assumptions.

We use $\mathbb{C}ov_t^i$ to denote agent i's conditional covariance operator. Following the same steps as in the single-risky-asset case, and under the usual assumption of conditional multivariate normality of returns (in agents' beliefs), the maximization of the CARA expected utility of wealth at $t+1$ results, at time t, in the optimal portfolio choice

$$z_t^i = (V_t^i)^{-1} m_t^i / \alpha_i \tag{5.53}$$

where α_i is the absolute risk-aversion coefficient of agent type i, $m_t^i := \mathbb{E}_t^i(\boldsymbol{R}_{t+1})$ is the vector of conditionally expected excess returns and $V_t^i := \left[\mathbb{C}ov_t^i(R_{j,t+1}, R_{k,t+1})\right]$, $j, k = 1, 2, \ldots, m$, denotes the conditional variance/covariance matrix (assumed to be positive definite) of the excess returns per share, according to the beliefs of agent type i.

The Fundamentalists and Chartists

We again consider two types of agents, *fundamentalists* ($i = f$), who believe in mean reversion to the fundamental, and *trend followers* or *chartists* ($i = c$), who rely on extrapolation of observed price trends. Both agent types are assumed to share common and correct expectations about dividends, with $\mathbb{E}_t^i[y_{t+1}] = \overline{y}$, $i \in \{f, c\}$, though forming different beliefs about the conditional distribution of asset returns (that is, about $\boldsymbol{R}_{t+1}$). The assumed behavior of the two agent types is basically similar to the previously discussed one-risky-asset cases. In particular, beliefs are specified similarly to the model discussed in Section 5.3.5, which incorporates explicitly an adaptive mechanism used by the chartists to update the variance of prices, based on extrapolation of observed deviations from a sample mean. Such a mechanism is adopted also in the present multiasset model. The obvious difference with the one-risky-asset setup is that chartist extrapolation now also concerns the comovements of prices: The focus is therefore on the additional effects of this updating rule, by which observed comovements contribute to determine current and future portfolio allocations and thus affect the joint dynamics of prices.

To keep matters as simple as possible—and to focus only on the previously described effect—we assume that agents do not care about dividends explicitly when forming second moment beliefs about excess returns $\boldsymbol{R}_{t+1}$. It is simply assumed that the standard deviation of the "cum-dividend" prices $P_{j,t+1} + y_{j,t+1}$ is estimated in proportion to the standard deviation of the price. Namely, for $i \in \{f, c\}$ and j, $k = 1, 2, \ldots m$, we assume[22]

$$\mathbb{V}_t^i(R_{j,t+1}) = q_j^2 \mathbb{V}_t^i(P_{j,t+1})$$
$$\mathbb{C}ov_t^i(R_{j,t+1}, R_{k,t+1}) = q_j q_k \mathbb{C}ov_t^i(P_{j,t+1}, P_{k,t+1})$$

[22]For example, agents would set $q_j = 1 + \delta_j^a / K$, $(j = 1, 2, \ldots, m)$, where δ_j^a is an estimate of the average (per annual) dividend yield of asset j, and K is the trading frequency per year; an alternative choice could be $q_j = R$. Since the parameters used in the numerical simulations at the end of this section have been chosen to represent daily data (for which r_f is of the order of 10^{-4}), in the following discussion we neglect the dividend component and therefore assume $q_j = 1$, for simplicity.

where $q_j > 0$ are constants. For simplicity, we take $q_j = 1$ $(j = 1, 2, \ldots, m)$ in the following discussion.

As usual fundamentalists expect the future prices of the risky assets to evolve according to

$$\mathbb{E}_t^f[P_{t+1}] = P_t + (1 - d_f)(P^* - P_t), \quad (0 \le d_f \le 1) \tag{5.54}$$

where $(1 - d_f)$ represents the expected speed of adjustment of the prices toward their (constant) fundamental values, collected in the vector P^*.[23] Therefore, for the fundamentalists

$$m_t^f := \mathbb{E}_t^f[R_{t+1}] = (d_f - R)P_t + (1 - d_f)P^* + \overline{y}$$

The fundamentalists have the same constant beliefs about the variance/covariance structure of the excess returns, where

$$V_t^f = \overline{V}^f := [\mathbb{C}ov_t^f(P_{j,t+1}, P_{k,t+1})] \quad (j, k = 1, 2 \ldots, m)$$

represents the variance/covariance matrix of the prices.

Chartists expect prices to evolve according to

$$\mathbb{E}_t^c[P_{t+1}] = P_t + d_c(P_t - u_t)$$

where $d_c \ge 0$ is the price extrapolation parameter of the chartists and $u_t = [u_{j,t}]$ $(j = 1, 2, \ldots, m)$ is a vector of sample average prices. As a consequence one obtains

$$m_t^c := \mathbb{E}_t^c[R_{t+1}] = (1 + d_c - R)P_t - d_c u_t + \overline{y}$$

With regard to second-moment beliefs, the covariance matrix of excess returns, V_t^c, is specified as

$$V_t^c = \overline{V}^c + V_t \tag{5.55}$$

which consists of a constant component $\overline{V}^c$ and a time-varying component V_t. The latter is assumed to be updated in each period as a function of deviations of past prices from the sample means u_t. More precisely, the symmetric matrix V_t has the structure $V_t := [v_{jk} s_{jk,t}]$ $(j, k = 1, 2, \ldots, m)$, where $v_{jk} \ge 0$ is a sensitivity coefficient, while $s_{jk,t}$ represents a sample variance/covariance, based on observed data up to time t. We also denote by $\Sigma_t := [s_{jk,t}]$ the matrix of such time-varying sample variances/covariances.

[23]The fundamental price will be defined explicitly later, see Eq. 5.59. We may consider P^* as an exogenous parameter for the moment.

Finally, following He (2003) and Chiarella, Dieci, and Gardini (2005), we specify the learning processes about sample means u_t and sample variances/covariances Σ_t as

$$u_t = \delta u_{t-1} + (1-\delta)P_t \tag{5.56}$$

$$\Sigma_t = \delta\Sigma_{t-1} + \delta(1-\delta)(P_t - u_{t-1})(P_t - u_{t-1})^\top \tag{5.57}$$

where δ, $0 < \delta < 1$, represents a "memory" parameter.

Market-Clearing under a Market Maker

To close the model, the price-setting rule needs to be made explicit. Assuming a market-maker scenario, the price vector adjusts according to

$$P_{t+1} = P_t + \mu(n^f z_t^f + n^c z_t^c - z^s)$$

where $\mu > 0$ is the market maker's price adjustment parameter, n^f and $n^c (= 1 - n^f)$ represent the fractions of fundamentalists and chartists, respectively, and z^s denotes the quantities of shares (per agent) available in the market, so that the vector $(n^f z_t^f + n^c z_t^c - z^s)$ represents the average excess demand of each asset per agent. The asset price dynamics under heterogeneous beliefs are thus expressed by a nonlinear discrete-time dynamical system, in the state variables P_t, u_t, Σ_t.

To further simplify the model, we assume that agents are homogeneous in their risk-aversion coefficients ($\alpha_f = \alpha_c := \alpha$). We also assume that the constant component of the variance/covariance matrix of prices is the same for both agent types, and we denote it by

$$\overline{V}^c = \overline{V}^f = \overline{V} := \left[\rho_{jk}\overline{\sigma}_j\overline{\sigma}_k\right] \quad (j, k = 1, 2, \ldots, m)$$

where $\overline{\sigma}_j$ and ρ_{jk} denote standard deviation of price j and correlation between prices j and k, respectively. So the dynamics of the model evolve according to

$$\begin{cases} P_{t+1} = P_t + \mu[n^f z_t^f + (1-n^f)z_t^c - z^s] \\ u_{t+1} = \delta u_t + (1-\delta)P_{t+1} \\ \Sigma_{t+1} = \delta\Sigma_t + \delta(1-\delta)(P_{t+1} - u_t)(P_{t+1} - u_t)^\top \end{cases} \tag{5.58}$$

where

$$z_t^f = \frac{1}{\alpha}(\overline{V})^{-1}[(d_f - R)P_t + (1-d_f)P^* + \overline{y}]$$

$$z_t^c = \frac{1}{\alpha}(\overline{V} + V_t)^{-1}[(1 + d_c - R)P_t - d_c u_t + \overline{y}]$$

and where the symmetric matrix V_t is a function of Σ_t. For the sake of simplicity, the fundamental prices are set in such a way that P^* turns out to be the steady state of the dynamical system (Eq. 5.58),[24] which results in

$$P^* = \frac{1}{R-1}\left[\overline{y} - \alpha \overline{V} z_s\right] \tag{5.59}$$

where the quantity $\overline{y} - \alpha\overline{V}z_s$ can be interpreted as a vector of "risk-adjusted" dividends.[25]

An immediate remark about the model Eq. (5.58) is that the dynamics of the m asset prices will decouple from each other if $\overline{V}$ and V_t are diagonal matrices, that is, when prices are not correlated in the "fixed" component of agents' beliefs and chartist sensitivity to sample covariances is zero.

5.6.2. An Example of Two Risky Assets and Two Beliefs

To gain some insight into the dynamic behavior of the model and in particular into the role of anticipated correlations, in the following we consider the case of two risky assets, $m = 2$. For simplicity, it is assumed $v_{11} = v_{22} := v$. We also adopt the simplified notation $v_{12} := \lambda$, $s_{12,t} := s_t$, $s_{jj,t} := v_{j,t}$ $(j = 1, 2)$, $\rho_{12} := \rho$, so that the matrices $\overline{V}$, Σ_t and V_t can be rewritten, respectively, as follows:

$$\overline{V} = \begin{bmatrix} \overline{\sigma}_1^2 & \rho\overline{\sigma}_1\overline{\sigma}_2 \\ \rho\overline{\sigma}_1\overline{\sigma}_2 & \overline{\sigma}_2^2 \end{bmatrix} \quad \Sigma_t = \begin{bmatrix} v_{1,t} & s_t \\ s_t & v_{2,t} \end{bmatrix} \quad V_t = \begin{bmatrix} vv_{1,t} & \lambda s_t \\ \lambda s_t & vv_{2,t} \end{bmatrix}$$

In this case Eq. (5.58) reduces to a seven-dimensional dynamical system in which the correlation between prices $P_{1,t+1}$ and $P_{2,t+1}$ estimated by chartists at time t, $\rho_t^{(c)}$ is given by

$$\rho_t^{(c)} = \frac{\rho\overline{\sigma}_1\overline{\sigma}_2 + \lambda s_t}{\sqrt{(\overline{\sigma}_1^2 + vv_{1,t})(\overline{\sigma}_2^2 + vv_{2,t})}} \tag{5.60}$$

Despite the high dimension of the system, analytical results about the local asymptotic stability of the fundamental steady state are possible in the particular case where the "exogenous" part ρ of the correlation of the excess returns, which is a commonly held belief among the agents, is zero. Then the local stability and bifurcation of the dynamic portfolio model with two risky assets is actually determined by two independent dynamic models, each with a single risky asset (and the risk-free asset).[26] In this case it can be shown in general that when the speed of price reaction is not too high, the

[24]This corresponds to the price that would be obtained under homogeneous beliefs about the first and second moments of the cum-dividend price processes. This result can be regarded as a special case of Theorem 3.2 in Böhm and Chiarella (2005).

[25]A similar interpretation is suggested by Brock and Hommes (1998) in the case of one risky asset.

[26]We stress that this property, relative to the case $\rho = 0$, concerns only the behavior of the linerized system around the steady state. In general, the dynamics of the two prices will not decouple, unless V_t is a diagonal matrix (i.e., $\lambda = 0$).

steady state is locally stable for any fundamentalist fraction n^f provided that the chartist extrapolation rate d_c is sufficiently low, whereas for sufficiently high d_c the steady state becomes unstable through a supercritical Neimark-Sacker bifurcation, followed by the emergence of a stable closed invariant curve in the phase-space, resulting in endogenous fluctuations of the asset prices.

The correlation structure among different risky assets plays an important role in the dynamics of the multiperiod setup. Here we present some numerical examples highlighting the impact of agents' beliefs about the covariance structure on the dynamic behavior of the model (Eq. 5.58). Such beliefs play a role in two different ways. First, the coefficient ρ here represents an exogenously set (and common) component of the heterogeneous beliefs about the correlation of the two asset prices. We can interpret this component as a constant part of the beliefs, related to the dividend processes, which characterizes a steady-state situation. Second, the extrapolation of observed price comovements generates endogenously another kind of "perceived" correlation, the impact of which is governed by the parameter λ (sensitivity to sample covariance). This may be interpreted as an out-of-equilibrium component, which may result in a positive or negative correction with respect to the exogenous part ρ.

A natural question concerns the qualitative impact (stabilizing or destabilizing) of each of the two components, in particular whether the endogenously perceived correlation and its updating mechanism tends to dampen or to amplify the price movements that arise from the interaction of heterogeneous traders. Obviously, when $\lambda = 0$ and in addition $\rho = 0$, the demand functions become (for $j = 1, 2$)

$$z^f_{j,t} = \frac{(d_f - R)P_{j,t} + (1 - d_f)P^*_j + \overline{y}_j}{\alpha\overline{\sigma}^2_j}$$

$$z^c_{j,t} = \frac{(1 + d_c - R)P_{j,t} - d_c u_{j,t} + \overline{y}_j}{\alpha(\overline{\sigma}^2_j + \nu v_{j,t})}$$

which are precisely the demand functions that we would obtain in the case of a single risky asset (asset j), that is, for $n = 1$. In other words, in the case $\rho = \lambda = 0$, the prices of the two risky assets evolve independently from each other, whereas for $\rho = 0$, $\lambda = \nu$, agents care about the observed comovements of prices exactly as they do with their sample volatility. In general, $\lambda > 0$ results in interdependent demand functions of the two assets and in prices coevolving over time, even in the absence of exogenous correlation (i.e. when $\rho = 0$). This can be seen from the following numerical simulations, where parameters are chosen in a way that the two assets have equal fundamental prices, which makes possible the use of the same (P, u) plane to represent both asset 1 and asset 2 (via projections onto the planes P_1, u_1 and P_2, u_2, respectively), using different shading. The common set of parameters used in the numerical simulations is the following: $\nu = 0.5$, $\delta = \mu = 0.5$, $n^f = 0.25$, $\alpha = 0.05$, $R = 1.0002$, $\overline{\sigma}^2_1 = 0.9$, $\overline{\sigma}^2_2 = 1.6$, $z^s_1 = z^s_2 = 1$, $\overline{y}_1 = 0.071$, $\overline{y}_2 = 0.106$.

In Figure 5.16 we set $d_f = 0.75$, $d_c = 0.45$. The panels from (a) to (e), where $\rho = 0$, represent a case in which a higher sensitivity to sample covariances is able to stabilize

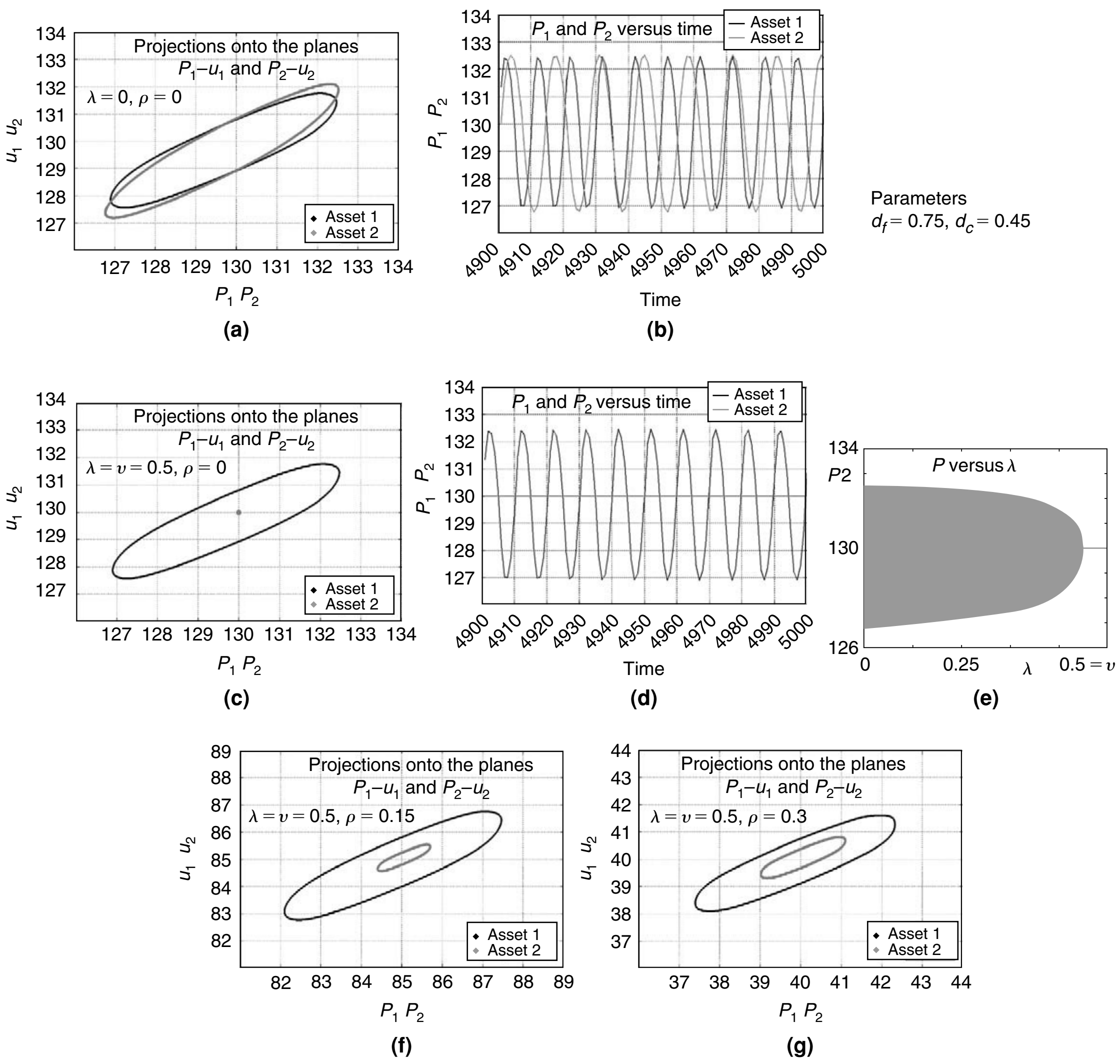

FIGURE 5.16 A two-asset market. Sensitivity to observed comovements can stabilize asset price 2, from (a,b) to (c,d), or in general can reduce the amplitude of fluctuations (e). As an effect of exogenous correlation (c,f,g), the two assets evolve with increasingly similar patterns. Fundamentalist and chartist parameters: $d_f = 0.75$, $d_c = 0.45$.

the dynamics to some extent. This effect is quite often detected when the starting situation with "decoupled" markets ($\rho = 0$, $\lambda = 0$) is one in which the two prices evolve with regular fluctuations of similar amplitude. In particular, the perceived correlation is able to stabilize wide fluctuations of asset 2 into convergence to the fundamental price (compare (a,b) with (c,d)). This is also revealed by the bifurcation diagram (e) of price 2 versus the parameter λ. The higher the sensitivity to observed comovements, the smaller the range of fluctuations of asset 2. Note that the exogenous component ρ causes in general the opposite effect: Larger values of ρ produce wider fluctuations for asset price 2 (by leaving price 1 almost unchanged in this case). This can be seen from the phase plots in Figures 5.16c, 5.16f, and 5.16g, where ρ is increased from 0 to 0.15 and then to 0.3.

We have detected such a destabilizing effect of ρ under several different parameter constellations. In general, the larger is ρ, the more similar is the nature and the amplitude of the fluctuations of the two prices due to a spillover effect from the market, with wider fluctuations to the more tranquil market. In contrast, the qualitative impact of λ is characterized by a strong dependence on the parameters of the model. We reported in Figure 5.16 a situation that is stabilized by the "perceived" correlations, but in other cases the effect is that of producing a transition toward less regular fluctuations and phase plots of increasing complexity. Figure 5.17 (where $d_f = 0.1$, $d_c = 1.2$) shows one such case, which is obtained under a stronger trend extrapolation and fundamentalist reaction than in Figure 5.16. This suggests that in the presence of price fluctuations determined by fundamentalist-chartist interaction, the sensitivity to

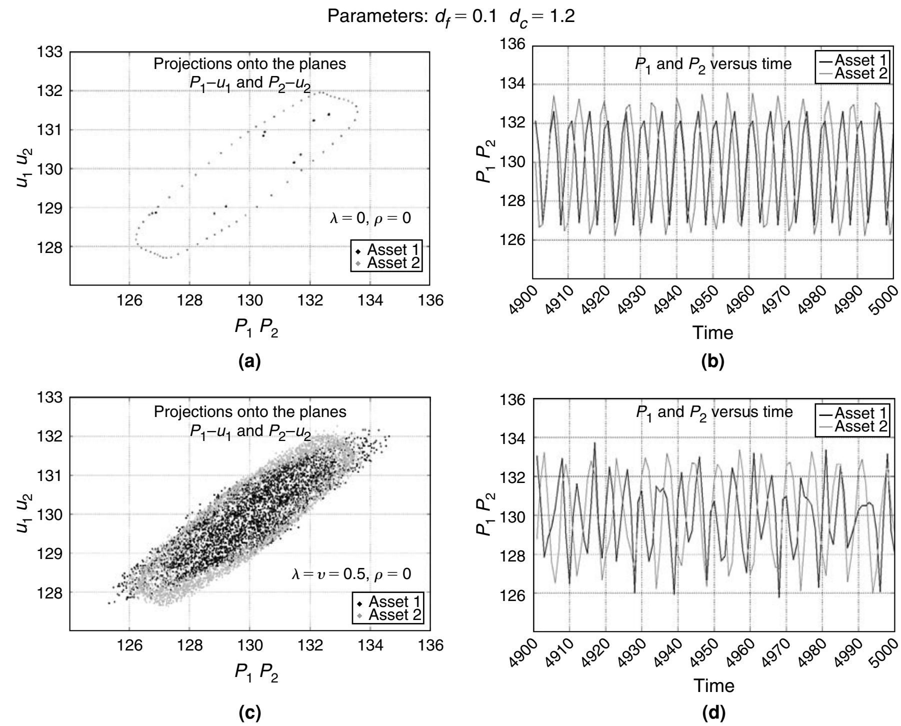

FIGURE 5.17 A two-asset market. Destabilizing effect of the sensitivity to historical covariances: from cyclical behavior on a periodic orbit (a,b) to intricate fluctuations on a chaotic attractor (c,d). Fundamentalist and chartist parameters: $d_f = 0.1$, $d_c = 1.2$.

observed comovements is not always able to stabilize the financial market, but it might even act as a further source of complex behavior.

Summary

Our analysis suggests that combining the BRHA approach with the mean-variance portfolio model with multiple risky assets may provide a suitable framework to address a number of issues concerning comovements in stock prices and in particular the way price fluctuations can propagate themselves across markets as a results of agents' beliefs. It turns out that second-moment beliefs, which are updated dynamically as a function of observed volatility and comovements, affect the agents' portfolio allocation process and may determine different patterns for the joint dynamics of prices. Such mechanisms may play a stabilizing role on the asset prices by reducing the amplitude of fluctuations, but in some cases they may contribute to reinforcing and amplifying price movements. The outcome depends crucially on some behavioral parameters, among which are the fundamentalist reaction and the strength of chartist extrapolation.

5.7. THE CONTINUOUS STOCHASTIC DYNAMICS OF SPECULATIVE BEHAVIOR

As we saw in Section 5.5, an important issue for the BRHA class of models is the interaction of noise with the underlying nonlinear deterministic market dynamics. The approach of Section 5.5, referred to as the *indirect approach*, first considers the corresponding deterministic "skeleton" of the stochastic models in which the noise terms are set to zero and uses stability and bifurcation theory to investigate the dynamics of this nonlinear deterministic system. It then uses simulation methods to examine the interplay of various types of noise and the deterministic dynamics. Hommes (2006) gives many references to this indirect approach. Ideally we would like to deal directly with the dynamics of the stochastic system. For example, in BRHA models with stochastic noise, we would like to know how the statistical properties of the model, which can be characterized by the stationary distribution of the market price process, change as agents' behavior changes and how the market price distribution is influenced by the underlying deterministic dynamics. In particular, we can ask whether there is a connection between different types of attractors and bifurcations of the underlying deterministic skeleton and different types of invariant measures of the stochastic system.

5.7.1. Stochastic Models with Heterogeneous Beliefs

A number of stochastic asset pricing models have been constructed in the heterogeneous agent literature. Brock, Hommes, and Wagener (2005) study the evolution of a discrete financial market model with many types of agents by focusing on the limiting distribution over types of agents that is able to generate important stylized facts. Föllmer, Horst, and Kirman (2005) consider a discrete financial market model with adaptive heterogeneous agents and under certain conditions the price process of which displays fat tails. Rheinlaender and Steinkamp (2004) study a one-dimensional continuously randomized version of the model of Zeeman (1974) and show a stochastic stabilization effect and possible sudden trend reversal.

Other related works include Hens and Schenk-Hoppé (2005), who analyze portfolio selection rules in incomplete markets where the wealth shares of investors are described by a random discrete dynamical system; Lux and Schornstein (2005), who present an adaptive model of a two-country foreign exchange market where agents learn by using genetic algorithms; Böhm and Chiarella (2005), who consider the dynamics of a general explicit random price process of many assets in an economy with overlapping generations of heterogeneous consumers forming optimal portfolios; and Böhm and Wenzelburger (2005), who provide a simulation analysis of the empirical performance of portfolios in a competitive financial market with heterogeneous investors and show that the empirical performance measure may be misleading. Most of the cited papers focus on the existence and uniqueness of limiting distributions of discrete time models rather than the existence and stability of multiple limiting distributions of continuous time models that we consider here.

5.7.2. A Continuous Stochastic Model with Fundamentalists and Chartists

In this section we sketch out a continuous time stochastic version of the most basic of the models discussed earlier.[27] We use a continuous time model since it allows us to use the concepts and stochastic bifurcation techniques from the theory of random dynamical systems (see Arnold, 1998) to conduct a quantitative and qualitative analysis of the stochastic model and examine the existence and stability of invariant measures of the equilibrium market price. Also, we have chosen this framework as it is a stochastic extension of the basic early model of Chiarella (1992). We investigate the equilibrium distribution of the market price through a numerical study of the stochastic bifurcation and show that the market price can display different forms of equilibrium distribution, depending on the speculative behavior of the chartists.

As in previous sections, we consider two types of investors, fundamentalists and chartists. The changes of the risky asset price $P(t)$ are brought about by aggregate excess demand of fundamentalists (D_t^f) and chartists (D_t^c) at a finite speed of price adjustment, so that

$$dp(t) = [D_t^f + D_t^c]dt \tag{5.61}$$

where $p_t = \ln P(t)$ is the logarithm of the risky asset price $P(t)$ at time t. The excess demand of the fundamentalists is assumed to be given by $D_t^f(p(t)) = a[F(t) - p(t)]$, where $F(t)$ denotes the logarithm of the fundamental price that clears fundamental demand at time t so that $D_t^f(F(t)) = 0$ and $a > 0$ is a constant measuring the excess demand of the fundamentalists brought about by the market price deviation from the

[27] This section draws on Chiarella, He, and Zheng (2007) and Chiarella, He, Wang, and Zheng (2007).

fundamental price. The chartists' excess demand is determined by $\psi(t)$, their assessment of the current trend in $p(t)$. Their excess demand is assumed to be given by $D_t^c(p(t)) = h(\psi(t))$, where h is a sigmoid type function. We simply assume that ψ is taken as an exponentially declining weighted average of past price changes, which can be expressed as $d\psi(t) = c[\dot{p}(t) - \psi(t)]dt$, where $c \in (0, \infty)$ is the speed with which chartists adjust their estimate of the trend to past price changes.

We thus obtain the asset price dynamics:

$$\begin{cases} dp(t) \ = a[F - p(t)]dt + h(\psi(t))dt \\ d\psi(t) = [-acp(t) - c\psi(t) + ch(\psi(t)) + acF]dt \end{cases} \tag{5.62}$$

When the fundamental price F is a constant, the system (Eq. 5.62) has a unique steady state $(\bar{p}, \psi) = (F, 0)$, which is locally stable if and only if $c < c^* = a/(b-1)$, where $b = h'(0)$ represents the slope of the demand function at $\psi = 0$ for the chartists. For $c > c^*$ the dynamics are characterized by a stable limit cycle. For the rest of this section, we only consider the case $b > 1$.

Using the notation of stochastic differential equations, the fundamental value $F(t)$ is assumed to follow $dF = \sigma \circ dW$, where W is a two-sided Wiener process[28] and $\sigma > 0$ is the instantaneous standard deviation (volatility) of the fundamental returns. Here the circle $\circ$ indicates that the SDE is interpreted in the Stratonovich sense[29] rather than the Itô sense. The dynamics can be expressed as a nonlinear Stratonovich-SDE system in terms of the variables ψ and ϕ, namely:

$$\begin{cases} d\psi \ = \phi dt \\ d\phi \ = [-a - c + ch'(\psi)]\phi dt - ac\psi dt + ac\sigma \circ dW \end{cases} \tag{5.63}$$

Once the dynamics of $\psi(t)$ have been obtained, the dynamics of the log price $p(t)$ can be obtained by integrating the first equation (5.62).

5.7.3. A Random Dynamical System and Stochastic Bifurcations

Using the methods outlined in Schenk-Hoppé (1996b), it can be shown that Eq. 5.63 defines a global random dynamical system (RDS). An RDS consists of two ingredients: a model describing a dynamical system perturbed by noise and a model of the noise itself. We use stochastic bifurcation theory to help us understand the stochastic nature of the random dynamical system Eq. 5.63, in particular, the study of its limiting distributions. To do so we use numerical tools to analyze the model, largely motivated by the work of Arnold, Sri Namachchivaya, and Schenk-Hoppé (1996) and Schenk-Hoppé (1996a) on the noisy Duffing-van der Pol oscillator.

In the analysis of stochastic differential equations, we may consider two types of bifurcation. This is essentially to do with the fact that SDEs may be considered from

[28] A two-sided Wiener process is one that evolves both backward and forward in time.

[29] The Stratonovich stochastic integral is used in the theory of random dynamical systems as the chain rule of ordinary calculus applies, and this facilitates many of the theoretical developments.

the pathwise point of view or the probability distribution point of view. The first approach corresponds to the so-called dynamical (D)-bifurcation, which examines the simultaneous behavior of paths forward and backward in time and encapsulates all the stochastic dynamics of the SDEs. The second approach corresponds to the so-called phenomenological (P)-bifurcation. The stationary measure is a quantity corresponding to the solution of the corresponding Fokker-Planck equation. As indicated in Schenk-Hoppé (1996a) and the references cited therein, the difference between P-bifurcation and D-bifurcation lies in the fact that the P-bifurcation approach focuses on long-run probability distributions, whereas the D-bifurcation approach is based on the pathwise viewpoint and focuses on invariant measures, the multiplicative ergodic theorem, and Lyapunov exponents.

D-Bifurcation

We take c, the speed of adjustment of the chartist toward the trend, as the bifurcation parameter. A D-bifurcation occurs if a reference invariant measure μ^c depending on the parameter c loses its stability at some point c_D, and another invariant measure $\nu^c \neq \mu^c$ exists for some c in each neighborhood of c_D, with ν^c converging weakly to μ^{c_D} as $c \to c_D$. Thus, the D-bifurcation focuses on the loss of stability of invariant measures and on the occurrence of new invariant measures. The D-bifurcation approach is a natural generalization of deterministic bifurcation theory, if one adopts the viewpoint that an invariant measure is the stochastic analogue of an invariant set—for example, a fixed point—and the multiplicative ergodic theorem is the stochastic equivalent of linear algebra. The D-bifurcation may be examined through the calculation of Lyapunov exponents and random attractors.

Let μ^c be an invariant ergodic probability measure for the random dynamical system generated by Eq. 5.63 depending on the parameter c. We take $h(x) = \alpha \tanh(\beta x)$, where $\alpha, \beta\,(>0)$ so that $b = h'(0) > 1$ is satisfied, which corresponds to the appearance of complex phenomena in the deterministic case. By relying on the multiplicative ergodic theorem, we know that the Lyapunov exponents of the RDS (Eq. 5.63) indicate the stability of the invariant measure. Below a certain critical value c_D the Lyapunov exponents are negative and are associated with a stable invariant measure (stable in the sense that it attracts all trajectories starting sufficiently close to it) that we denote by μ^c. Above the critical value c_D the Lyapunov exponents become positive and a second pair of negative Lyapunov exponents appears. The original invariant measure μ^c, associated with the positive Lyapunov exponents, now becomes unstable (that is, stable in reverse time) and a new stable invariant measure, ν^c, is born and is associated with the negative Lyapunov exponents.

Figure 5.18 illustrates the Lyapunov exponents for c varying between 0 and 6.5. Below the critical value $c_D(\simeq 1)$ we see the two negative Lyapunov exponents $\lambda_1(\mu^c)$, $\lambda_2(\mu^c)$ associated with the stable invariant set μ^c. At around $c \simeq 0.25$ the two values come together and at the same time a rotation number $\kappa(\mu^c)$ appears, indicating (stochastic) cyclical convergence to the steady state. Above c_D the Lyapunov exponent becomes positive (and still has an associated rotation number). Finally, two Lyapunov

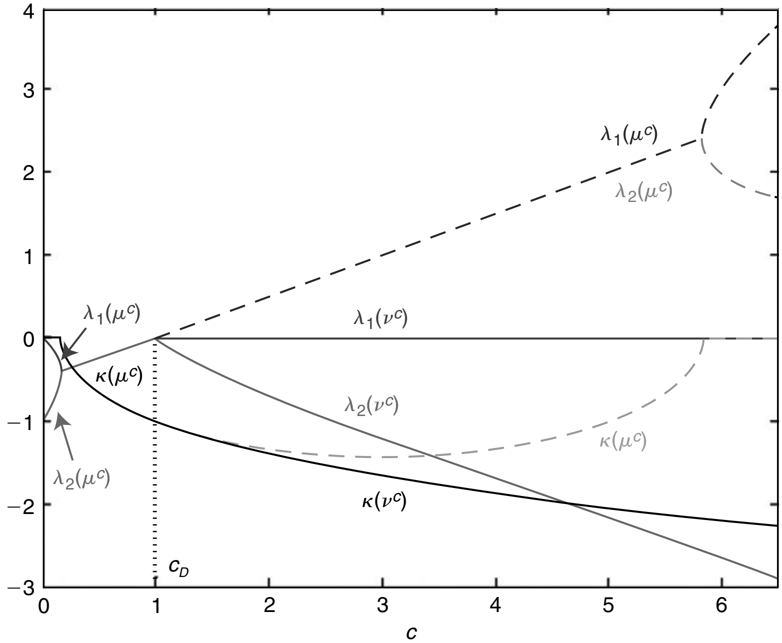

FIGURE 5.18 Lyapunov exponents and rotation number as a function of c for $a = 1$, $\alpha = 2$, $\beta = 1$, and $\sigma = 0.02$.

exponents appear above $c \simeq 5.8$ and the rotation number shrinks to zero. The associated invariant measure μ^c becomes unstable as c passes through c_D.

When $\lambda_{1,2}(\mu^c) > 0$, two other Lyapunov exponents emerge, denoted by $\lambda_{1,2}(\nu^c)$, satisfying $\lambda_2(\nu^c) < \lambda_1(\nu^c) \leq 0$, which indicates that a new stable invariant measure ν^c appears. There is an associated notation number $\kappa(\nu^c)$ indicating fluctuating convergence to the random attractor. This means that there always exists an invariant measure μ^c in the market. However, when the chartists extrapolate the price trend weakly (so that $c < c_D$), this invariant measure is unique and stable; when the chartists extrapolate the price trend strongly (so that $c > c_D$), there exists a new invariant measure ν^c such that the original invariant measure μ^c becomes unstable and the new invariant measure ν^c is stable. We note that the bifurcation value c_D is very close to the bifurcation value c^* for the deterministic case discussed.

The Random Attractors

We may characterize the stochastic Hopf bifurcation of the invariant measure using the concept of random attractors. Changes in the Lyapunov exponents indicate the changes of invariant measures. To actually give more information about invariant measures, we still need to examine global random attractors. These are essentially compact sets that attract the dynamic paths in the long run and are of particular importance, since on them the long-term behavior of the system takes place. The definition of a

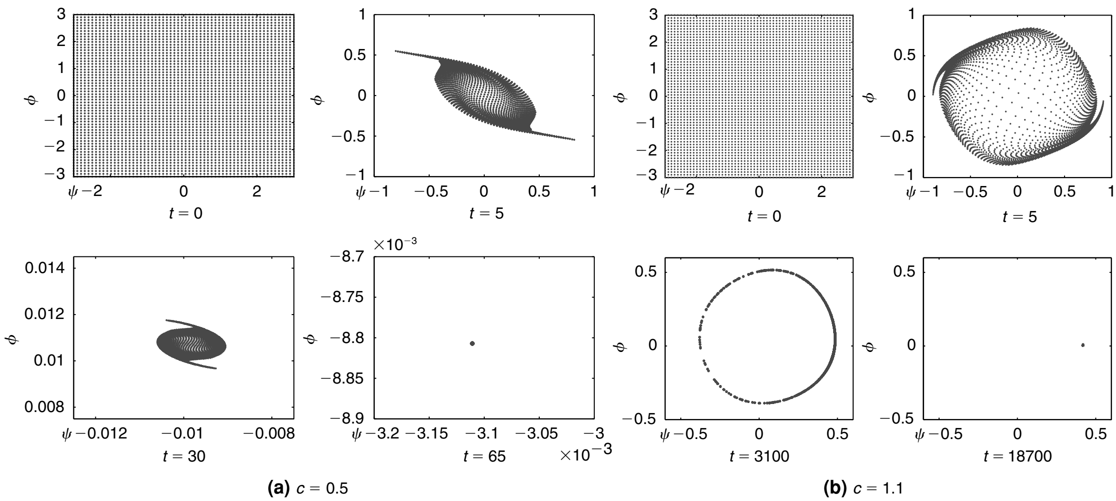

FIGURE 5.19 The random attractors when $a = 1$, $\alpha = 2$, $\beta = 1$, and $\sigma = 0.02$, where the initial value set D comes from a uniform distribution.

global random attractor is based on the so-called pullback process, which consists of simulating from a whole set of initial values moving from time $-t$ to time 0 (and not from 0 to t). This enables us to study the asymptotic behavior as $t \to \infty$ in the fixed fiber[30] at time 0. By increasing t, the mapping is made to start at successively earlier times, corresponding to a pull back in time. The pullback operation is shown in Figure 5.19 with a uniform distribution of initial values at different times t and two values of c, $c = 0.5(<c_D)$ and $1.1(>c_D)$.

When $c = 0.5$, through the pullback process, we can see from Figure 5.19a that all paths from the initial set shrink to a random fixed point $\mathbf{x}^*(\omega)$, which is distinct from zero. Moreover, numerically it is observed that, under time reversion, the solution of the system $\to \infty$ for any $\mathbf{x}_0 \neq \mathbf{x}^*(\omega)$, which implies that there is no other invariant measure. Linking with the calculation of the Lyapunov exponent in Figure 5.18, we know that the system is stable, with the largest Lyapunov exponents being negative, and it has a unique and stable invariant measure, which is a random Dirac measure $\mu_\omega = \delta_{\mathbf{x}^*(\omega)}$ and the global random attractor is $A(\omega) = \{\mathbf{x}^*(\omega)\}$. This is exactly the stochastic analogue of the corresponding deterministic case discussed earlier.

However, when $c = 1.1$, we observe from Figure 5.18 the occurrence of positive Lyapunov exponents. Applying the pullback operation again, we see from Figure 5.19b that a different behavior emerges, compared to the case of $c = 0.5$. A random circle becomes visible (at $t = 100$ in Figure 5.19b) and further convergence takes place on this circle. Finally, all trajectories converge to a random point $\mathbf{x}^{**}(\omega)$. We find that, again, the invariant measure is a random Dirac measure $\nu_\omega = \delta_{\mathbf{x}^{**}(\omega)}$, which is stable with the nonpositive largest Lyapunov exponent. However, through the time-reversed solution,

[30]The fiber is essentially a submanifold of the Euclidean space on which the dynamics evolve.

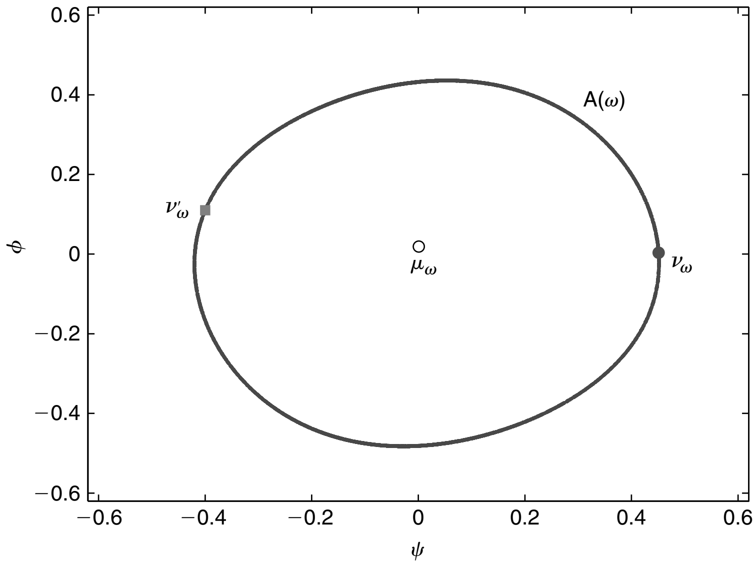

FIGURE 5.20 Global random attractor for $c = 1.1$, $a = 1$, $\alpha = 2$, $\beta = 1$, and $\sigma = 0.02$. μ_ω = unstable invariant measure, ν_ω = stable invariant measure, ν'_ω = additional invariant measure.

we show that the invariant measure $\mu_\omega = \delta_{\mathbf{x}^*(\omega)}$ exists in the interior of the circle, which is illustrated in Figure 5.20. Also, the invariant measure $\mu_\omega = \delta_{\mathbf{x}^*(\omega)}$ is unstable and has two Lyapunov exponents that are positive. In addition, under time reversion, $\mathbf{x}^{**}(\omega)$ is not attracting. As suggested in Schenk-Hoppé (1996a), another invariant measure, say, ν'_ω, on the random circle exists; see Figure 5.20. This analysis implies that, for $c = 1.1$, there exist more than two invariant measures; one is completely stable and one is completely unstable, and the global random attractor $A(\omega)$, which supports all invariant measures, is a random disc, the boundary of which is the random circle shown in Figure 5.19b.

In summary, our analysis on the D-bifurcation gives us insights into the significant impact of the chartists on the market equilibrium distributions. These distributions can be characterized by the invariant measures of the SDEs. We show that there exists a unique stable invariant measure in the market. However, the stable invariant measure changes quantitatively when the chartists change their extrapolation of the trend. The change can be described by the stochastic Hopf bifurcation. We have observed that the Hopf bifurcation remains on the level of the invariant measures as the loss of stability of a measure and occurrence of a new stable measure, and on the level of the global attractor as the change from a random point to a random disc.

P-Bifurcation

The analysis of D-bifurcation gives us a perspective from a dynamical systems viewpoint by focusing on the evolution of the random dynamical system. However, there is also a probability distribution viewpoint, which is best captured by focusing on the

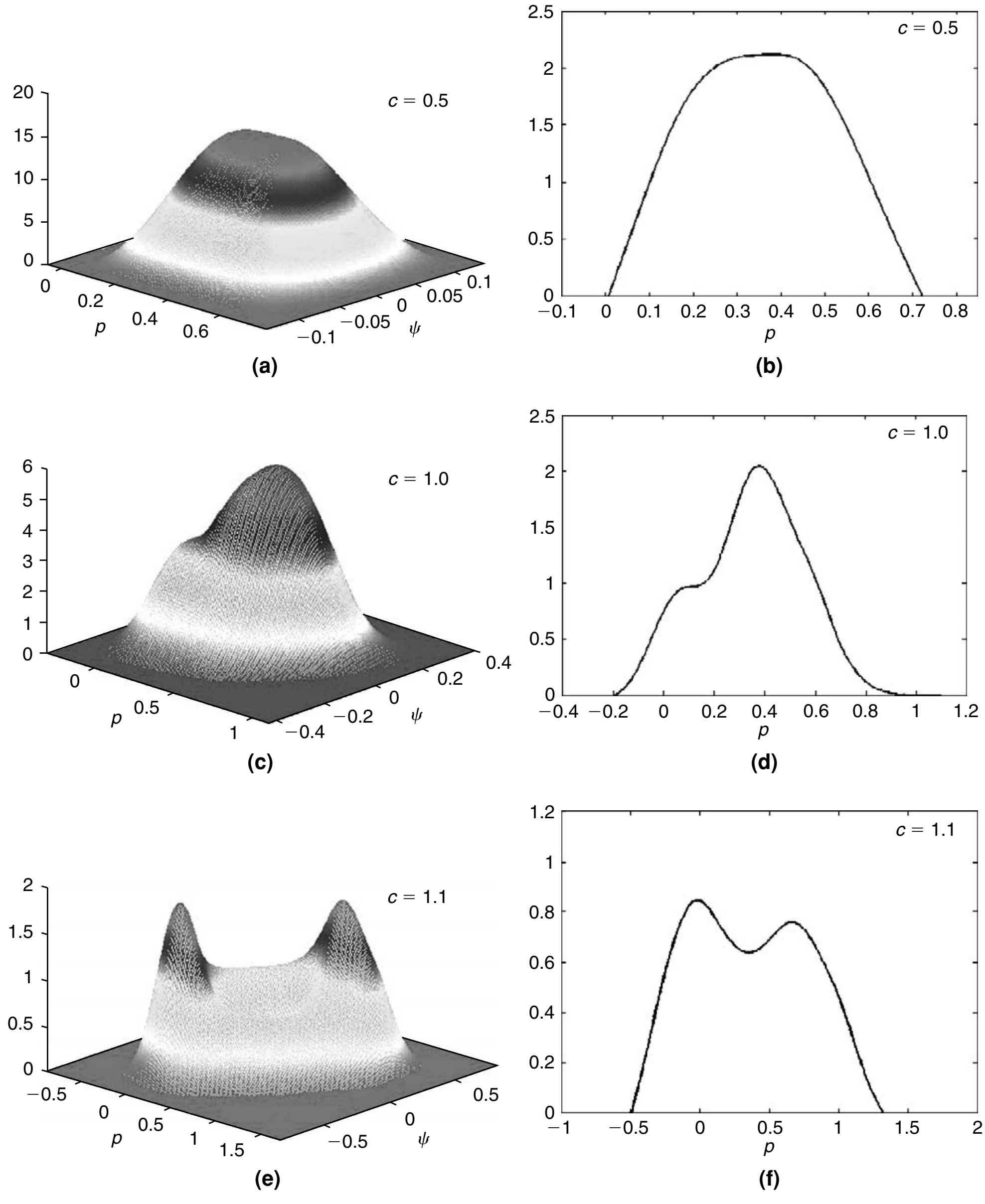

FIGURE 5.21 Joint stationary densities of the log price p and the assessment of the price trend ψ and the corresponding marginal distributions for the log price p.

stationary measure. The P-bifurcation approach to stochastic bifurcation theory examines the qualitative changes of the stationary measures. The stationary measure $\mathbb{P}$ has a probability density that is the stationary solution of the Fokker-Planck partial differential equation associated with the random dynamical system. It can be shown that there is a one-to-one correspondence between the stationary measure $\mathbb{P}$ and the invariant measure μ_ω (see Arnold, 1998).

The P-bifurcation approach studies qualitative changes of densities of stationary measures ρ_c when c varies. Hence, for the P-bifurcation, we are only interested in the changes of the shape of the stationary density. For $a = 1$, $\alpha = 2$, $\beta = 1$, and $\sigma = 0.02$. Figure 5.21 shows the joint stationary densities (the left panel) and the marginal densities (the right panel) for the log price p. We can see from Figure 5.21 that, for $c = 0.5$, the joint density in the (p, ψ) planes has one peak and the marginal densities for p is unimodal, which correspond to the stable Dirac invariant measure $\mu_\omega = \delta_{\mathbf{x}^*}(\omega)$ with the global random attractor $A(\omega) = \{\mathbf{x}^*(\omega)\}$ under the D-bifurcation analysis. However, for $c = 1.1$, the joint density in the (p, ψ) plane has a crater-like shape and the marginal densities for either ψ or p are bimodal. This change is underlined by the stable Dirac invariant measure $\nu_\omega = \delta_{\mathbf{x}^{**}}(\omega)$ with the global random attractor of a random disc under the D-bifurcation analysis. For $c = 1$, the joint and marginal densities can be regarded as the transition from single peak to crater-like (from unimodal to bimodal) densities. Therefore, as the chartists' adjustment parameter c increases, the qualitative changes of the stationary density indicate the occurrence of P-bifurcations.

Summary

Our analysis shows that D- and P-bifurcations characterize the stochastic behavior in different ways. We know that when the chartists extrapolate the trend weakly (so that $c < c_D$), the system only has one invariant measure $\delta_{\mathbf{x}^*(\omega)}$, which is stable. In this case, $\mathbf{x}^*(\omega)$ has a stationary measure that has one peak. However, when the chartists extrapolate the trend strongly (so that $c > c_D$), a new random Dirac measure $\delta_{\mathbf{x}^{**}(\omega)}$ appears and the corresponding stationary measure has a crater-like density. Quantitative changes under the D-bifurcation can help us obtain a better view of the qualitative changes under the P-bifurcation, but the combined analysis of both D- and P-bifurcations certainly gives us a relatively complete picture of the stochastic behavior of the model. Future research in this area will use the stochastic bifurcation concepts discussed here and apply them to more enhanced models of the BRHA type discussed in Section 5.3 and Section 5.4. It will probably also be necessary to introduce market noise as well as fundamental noise, since we have seen in Section 5.5 that this seems necessary to get the BRHA models to generate the stylized facts of financial markets.

5.8. CONCLUSION

This chapter has discussed the introduction of agent heterogeneity with regard to risk preferences and expectations into a repeated one-period optimizing market framework that underpins the traditional financial market paradigm and its well-established pricing theories such as CAPM. We have considered the two most common utility functions (CARA and CRRA) and the two common market-clearing mechanisms (the Walrasian auctioneer and the market maker). Combinations of these two elements provide a rich array of models that are both nonlinear (due to various behavioral features of the agents) and stochastic (typically due to fundamental and market noise). In all cases we have seen that application of the tools of deterministic nonlinear dynamical systems and bifurcation theory enable us to observe that the deterministic skeletons of the models produce both local stability to a fundamental equilibrium corresponding to that of the standard paradigm, as well as local instability of the fundamental, with a consequent rich range of possible complex behaviors involving various types of fluctuations and/or coexisting attractors.

We have also seen how insights into the nonlinear and stochastic versions of the models can be obtained both indirectly by simulation and directly by applying the tools and concepts of stochastic bifurcation theory. Indeed, we have seen that a calibrated version of such a nonlinear stochastic model is able to reproduce quite well the stylized facts of financial markets. The BRHA framework is thus able to accommodate market features that seem not easily reconcilable for the standard financial market paradigm, such as fat-tail behavior, volatility clustering, large excursions from the fundamental, and bubbles. We stress again that here these features are due to agent heterogeneity and the various agent types forming expectations by fairly simple and economically intuitive rules of thumb.

Research in this area is moving in several directions. First, there is a vast amount of ongoing research into exploring further the dynamic behavior of the basic one risky asset/one risk-free asset BRHA model.[31] Second, there is the extension from the single risky asset to multiple risky asset framework. We alluded to some of this work in Section 5.6, but much remains to be done, such as the incorporation of stochastic elements, allowing for switching fractions of agents, and the development of a BRHA version of CAPM so as to make clear the role of agent heterogeneity on relative pricing of assets in financial markets. As well as the work already cited on this topic there is also the work of Chiarella, Dieci, and He (2006, 2007a), which constitutes a preliminary exploration of a BRHA CAPM. We should also cite the work of Li (2007) and Jouini and Napp (2006) on the same theme.

Third, there is the extension of the stochastic bifurcation analysis illustrated on a simple model in Section 5.7 to more fully developed BRHA models containing many of the elements of the models discussed in Sections 5.3 and 5.4. Fourth, much more needs to be done on the calibration and estimation of BRHA models. We discussed one such calibration in Section 5.5, but we should also point out the work of Alfarano, Lux, and Wagner (2005, 2008), who estimate a BRHA model in a number of markets. Lux (2009) surveys recent literature on this topic. Since the BRHA models are essentially partially observed dynamical systems in nature it should be possible to approach the estimation problem using nonlinear filtering methodology. This may prove a fruitful area of research in the near future.

Fifth, the dynamics of the model so far have come from an intertemporal repetition of the one-period optimization model. There is a need to go beyond this simple dynamic

[31] We cite, in particular, Chiarella, Gallegati, Leombruni, and Palestrini (2003); Chiarella, He, and Hommes (2006); Chiarella, He, and Wang (2006); Hommes (2002); Iori (2002); Levy, Levy, and Solomon (1995); and Lux (1995, 1998).

framework and consider the heterogeneous agents optimizing over several periods or over the infinite future—in other words the development of a heterogeneous agent version of the intertemporal optimizing framework. There have already been contributions along these lines in the discrete time framework by Hillebrand and Wenzelburger (2006a, 2006b). There has also been some work within the continuous time framework that is nicely summarized in Ziegler (2003); however, the focus of this latter work is usually on heterogeneity of information rather than heterogeneity of expectations. Furthermore, these models largely lack the expectations feedback aspect that is an essential aspect of the BRHA framework. Work in this area should provide a rich research agenda over the next few years.

Finally, it would seem worthwhile to incorporate BRHA into macrodynamic models. The work to date on BRHA models has shown the importance of agent heterogeneity on financial market behavior. It is also well appreciated that the link between the real and the financial sector is an important one for modern economies. Also, over the last two decades the world has witnessed a number of episodes in which trend-chasing behavior seems to have dominated over rational fundamental behavior, and this behavior in the financial market has spilled over into the real economy. Thus it seems worthwhile to incorporate agent heterogeneity into macrodynamic models. A first very simple attempt in this direction is given by Chiarella, Flaschel, and Hung (2006), but this area still remains largely unexplored.

References

Alfarano, S., T. Lux, and F. Wagner, Estimation of agent-based models: The case of an asymmetric herding model, *Computational Economics*, 2005, 26, 19–49.

Alfarano, S., T. Lux, and F. Wagner, Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach, *Journal of Economic Dynamics and Control*, 2008, 101–136.

Allen, H., and M. Taylor, Charts, noise and fundamentals in the London foreign exchange market, *Economic Journal*, 1990, 100, 49–59.

Anderson, S., A. de Palma, and J. Thisse, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1993.

Anufriev, M., G. Bottazzi, and F. Pancotto, Equilibria, stability and asymptotic dominance in a speculative market with heterogeneous traders, *Journal of Economic Dynamics and Control*, 2006, 30, 1787–1835.

Arnold, L., *Random Dynamical Systems*, Springer-Verlag, 1998.

Arnold, L., N. Sri Namachchivaya, and K. R. Schenk-Hoppé, Toward an understanding of stochastic Hopf bifurcation: A case study, *International Journal of Bifurcation and Chaos*, 1996, 6, 1947–1975.

Baillie, R. T., T. Bollerslev, and H. Mikkelsen, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 1996, 74, 3–30.

Barsky, R. B., and J. B. De Long, Why does the stock market fluctuate?, *Quarterly Journal of Economics*, 1993 108, 291–311.

Beja, A., and M. Goldman, On the dynamic behavior of prices in disequilibrium, *Journal of Finance*, 1980, 34, 235–247.

Bernard, V. L., Stock price reaction to earnings announcements: A summary of recent anomalous evidence and possible explanations, *Advances in Behavioral Finance*, Russell Sage Foundation, 1993, 303–340.

Black, F., and M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy*, 1973, 81, 637–659.

Böhm, V., and C. Chiarella, Mean-variance preferences, expectations formation and the dynamics of random asset prices, *Mathematical Finance*, 2005, 15, 61–97.

Böhm, V., and J. Wenzelburger, On the performance of efficient portfolios, *Journal of Economic Dynamics and Control*, 2005, 29, 721–740.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 1986, 31, 307–327.

Brock, W., and C. Hommes, A rational route to randomness, *Econometrica*, 1997, 65, 1059–1095.

Brock, W., and C. Hommes, Heterogeneous beliefs and routes to chaos in a simple asset-pricing model, *Journal of Economic Dynamics and Control*, 1998, 22, 1235–1274.

Brock, W., C. Hommes, and F. Wagener, Evolutionary dynamics in financial markets with many trader types, *Journal of Mathematical Economics*, 2005, 41, 7–42.

Campbell, J., and A. S. Kyle, Smart money, noise trading and stock price behaviour, *Review of Economic Studies*, 1993, 60, 1–34.

Campbell, J., and L. Viceira, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press, 2002.

Chiarella, C., Perfect foresight models and the dynamic instability problem from a higher viewpoint, *Economic Modelling*, 1986, 3, 283–292.

Chiarella, C., The dynamics of speculative behavior, *Annals of Operations Research*, 1992, 37, 101–123.

Chiarella, C., R. Dieci, and L. Gardini, Speculative behavior and complex asset price dynamics: A global analysis, *Journal of Economic Behavior and Organization*, 2002, 49, 173–197.

Chiarella, C., R. Dieci, and L. Gardini, The dynamic interaction of speculation and diversification, *Applied Mathematical Finance*, 2005, 12, 17–52.

Chiarella, C., R. Dieci, and L. Gardini, Asset price and wealth dynamics in a financial market with heterogeneous agents, *Journal of Economic Dynamics and Control*, 2006, 30, 1775–1786.

Chiarella, C., R. Dieci, and X. He, Aggregation of heterogeneous beliefs and asset-pricing theory: A mean-variance analysis, Working Paper 186, Quantitative Finance Research Centre, The University of Technology, Sydney, 2006.

Chiarella, C., R. Dieci, and X. He, A dynamic heterogeneous beliefs CAPM, Working Paper 181, *Computing in Economics and Finance*, Society for Computational Economics, 2007a.

Chiarella, C., R. Dieci, and X. He, Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework, *Journal of Economic Behavior and Organization*, 2007b, 62, 408–427.

Chiarella, C., P. Flaschel, and H. Hung, A stochastic model of real-financial interaction with boundedly rational heterogeneous agents, in P. Faschel, C. Chiarella, R. Franke, and W. Semmler (eds.), "Quantitative and Empirical Analysis of Nonlinear Dynamics Macromodels," Vol. 277 of *Contribution to Economic Analysis*, Elsevier, 2006, 333–357.

Chiarella, C., M. Gallegati, R. Leombruni, and A. Palestrini, Asset price dynamics among heterogeneous interacting agents, *Computational Economics*, 2003, 22, 213–223.

Chiarella, C., and X. He, Asset price and wealth dynamics under heterogeneous expectations, *Quantitative Finance*, 2001, 1, 509–526.

Chiarella, C., and X. He, Heterogeneous beliefs, risk and learning in a simple asset-pricing model, *Computational Economics*, 2002, 19, 95–132.

Chiarella, C., and X. He, Heterogeneous beliefs, risk and learning in a simple asset-pricing model with a market maker, *Macroeconomic Dynamics*, 2003, 7, 503–536.

Chiarella, C., and X. He, An asset-pricing model with adaptive heterogeneous interacting agents, in T. Lux, S. Reitz, and E. Samanidou (eds.), "Nonlinear Dynamics and Heterogeneous Interacting Agents," Vol. 550 of *Lecture Notes in Economics and Mathematical Systems*, Springer, 2005, 269–285.

Chiarella, C., and X. He, An adaptive model of asset price and wealth dynamics in a market with heterogeneous trading strategies, in D. Seese, C. Weinhardt, and F. Schlottmann (eds.) *Handbook on Information Technology in Finance*, Springer, 2008.

Chiarella, C., X. He, and C. Hommes, A dynamic analysis of technical trading rules in financial markets, *Journal of Economic Dynamics and Control*, 2006, 30, 1729–1753.

Chiarella, C., X. He, and D. Wang, A behavioral asset-pricing model with a time-varying second moment, *Chaos, Solitons and Fractals*, 2006, 29, 535.

Chiarella, C., X. He, D. Wang, and M. Zheng, The stochastic bifurcation behavior of speculative financial markets, *Physica A*, 2008, 387, 3837–3846.

Chiarella, C., X. He, and M. Zheng, The stochastic dynamics of speculative prices, Working Paper 208, Quantitative Finance Research Centre, The University of Technology, Sydney, 2007.

Cochrane, J. H., Volatility tests and efficient markets: A review essay, *Journal of Monetary Economics*, 1991, 27, 463–485.

Day, R., and W. Huang, Bulls, bears and market sheep, *Journal of Economic Behavior and Organization*, 1990, 14, 299–329.

Dieci, R., I. Foroni, L. Gardini, and X. He, Market mood, adaptive beliefs and asset price dynamics, *Chaos, Solitons and Fractals*, 2006, 29, 520–534.

Ding, Z., C. Granger, and R. Engle, A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1993, 1, 83–106.

Engle, R., Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 1982, 50, 987–1008.

Fama, E. F., *Foundations of Finance*, Basic Books, 1976.

Föllmer, H., U. Horst, and A. Kirman, Equilibria in financial markets with heterogeneous agents: A probabilistic perspective, *Journal of Mathematical Economics*, 2005, 41, 123–155.

Franke, R., and R. Sethi, Cautious trend-seeking and complex asset price dynamics, *Research in Economics*, 1998, 52, 61–79.

Frankel, J., and K. A. Froot, Short-term and long-term expectations of the Yen/Dollar exchange rate: Evidence from survey data, *Journal of the Japanese and International Economics*, 1987, 1, 249–274.

Frankel, J., and K. A. Froot, Chartists, fundamentalists and trading in the foreign exchange market, *American Economic Review*, 1990, 80, 181–185.

Gaunersdorfer, A., Endogenous fluctuations in a simple asset pricing model with heterogeneous agents, *Journal of Economic Dynamics and Control*, 2000, 24, 799–831.

Gaunersdorfer, A., and C. Hommes, A nonlinear structural model for volatility clustering, Technical Report 00-02, CeNDF, University of Amsterdam, working paper, 2000.

He, X., Asset pricing, volatility and market behavior: A market fraction approach, Working Paper 95, Quantitative Finance Research Centre, The University of Technology, Sydney, 2003.

He, X., and Y. Li, Heterogeneity, convergency, and autocorrelation, *Quantitative Finance,* 2008, 8, 59–79.

He, X., and Y. Li, Power law behavior, heterogeneity and trend chasing, *Journal of Economic Dynamics and Control*, 2007b, 31, 3396–3426.

Hens, T., and K. Schenk-Hoppé, Evolutionary stability of portfolio rules, *Journal of Mathematical Economics*, 2005, 41, 43–66.

Hillebrand, M., and J. Wenzelburger, The impact of multiperiod planning horizons on portfolios and asset prices in a dynamic CAPM, *Journal of Mathematical Economics*, 2006a, 42, 565–593.

Hillebrand, M., and J. Wenzelburger, On the dynamics of asset prices and portfolios in a multiperiod CAPM, *Chaos, Solitons and Fractals*, 2006b, 29(3), 578–594.

Ho, T., and H. Stoll, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics*, 1981, 9, 47–73.

Hommes, C., Modeling the stylized facts in finance through simple nonlinear adaptive systems, *Proceedings of National Academy of Science of the United States of America*, 2002, 99, 7221–7228.

Hommes, C., Heterogeneous Agent Models in Economics and Finance, in L. Tesfatsion, and K. Judd (eds.), "Agent-based Computational Economics," Vol. 2 of *Handbook of Computational Economics*, North-Holland, 2006, 1109–1186.

Iori, G., A Microsimulation of traders activity in the stock market: The role of heterogeneity, agents' interactions and trade frictions, *Journal of Economic Behavior and Organization*, 2002, 49, 269–285.

Jouini, E., and C. Napp, Heterogeneous beliefs and asset pricing in discrete time: An analysis of pessimism and doubt, *Journal of Economic Dynamics and Control*, 2006, 30, 1233–1260.

Keim, D. B., Stock market regularities: A synthesis of the evidence and explanations, in Elroy Dimson (ed.), *Stock Market Anomalies*, Cambridge University Press, 1988.

Kirman, A., Whom or what does the representative agent represent?, *Journal of Economic Perspectives*, 1992, 6, 117–136.

Kirman, A., Ants, rationality and recruitment, *Quarterly Journal of Economics*, 1993, 108, 137–156.

LeBaron, B., 24 agent-based computational finance, in L. Tesfatsion and K. Judd (eds.), "Agent-based Computational Economics," Vol. 2 of *Handbook of Computational Economics*, North-Holland, 2006.

Levy, M., and H. Levy, The danger of assuming homogeneous expectations, *The Financial Analyst Journal*, May 1996, 65–70.

Levy, M., H. Levy, and S. Solomon, A microscopic model of the stock market: Cycles, booms, and crashes, *Economic Letters*, 1994, 45, 103–111.

Levy, M., H. Levy, and S. Solomon, Microscopic simulation of the stock market: The effect of microscopic diversity, *Journal de Physique I (France)*, 1995, 5, 1087–1107.

Levy, M., H. Levy, and S. Solomon, *Microscopic Simulation of Financial Markets*. Academic Press, 2000.

Li, T., Heterogeneous beliefs, asset prices, and volatility in a pure exchange economy, *Journal of Economic Dynamics and Control*, 2007, 31, 1697–1727.

Lintner, J., The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 1965, 47, 13–37.

Lux, T., Herd behavior, bubbles and crashes, *Economic Journal*, 1995, 105, 881–896.

Lux, T., The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions, *Journal of Economic Behavior and Organization*, 1998, 33, 143–165.

Lux, T., Financial Power Laws: Empirical evidence, models and mechanisms, in C. Cioffi (ed.), *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium in the Social Universe*, Cambridge University Press, submitted 2004.

Lux, T., Stochastic behavioral asset pricing and stylized facts, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Lux, T., and M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, *Nature*, 1999, 397, 498–500.

Lux, T., and S. Schornstein, Genetic learning as an explanation of stylized facts of foreign exchange markets, *Journal of Mathematical Economics*, 2005, 41, 169–196.

Madhavan, A., Market microstructure: A survey, *Journal of Financial Markets*, 2000, 3, 205–258.

Manski, C., and D. McFadden, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, 1981.

Markowitz, H. M., Portfolio Selection, *Journal of Finance*, 1952, 7, 77–91.

Marsh, T., and R. C. Merton, Dividend variability and variance bounds tests for the rationality of stock market prices, *American Economic Review*, 1986, 76, 483–498.

Merton, R. C., Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory*, 1971, 3, 373–413.

Merton, R. C., An intertemporal capital asset pricing model, *Econometrica*, 1974, 41, 867–887.

Milgrom, P., and N. Stokey, Information, trade and common knowledge, *Journal of Economic Theory*, 1982, 12, 112–128.

Mossin, J., Equilibrium in a capital asset market, *Econometrica*, 1966, 34, 768–783.

O'Hara, M., *Market Microstructure Theory*, Blackwell Publishers, 1995.

Pagan, A., The econometrics of financial markets, *Journal of Empirical Finance*, 1996, 3, 15–102.

Peck, J., Liquidity without money: A general equilibrium model of market microstructure, *Journal of Financial Intermediation*, 1990, 1, 80–103.

Pruitt, S. W., and R. E. White, The crisma trading system: Who says technical analysis can't beat the market, *Journal of Portfolio Management*, 1988, 14, 55–58.

Raffaelli, G., and M. Marsili, Dynamic instability in a phenomenological model of correlated assets, *Journal of Statistical Mechanics*, August 2006, L08001.

Rheinlaender, T., and M. Steinkamp, A stochastic version of Zeeman's market model, *Studies in Nonlinear Dynamics and Econometrics*, 2004, 8, Article 4.

Schenk-Hoppé, K. R., Bifurcation scenarios of the noisy Duffing-van der Pol oscillator, *Nonlinear Dynamics*, 1996a, 11, 255–274.

Schenk-Hoppé, K. R., Deterministic and stochastic Duffing-van der Pol oscillators are non-explosive, *Zeitschrift für Angewandte Physik*, 1996b, 47, 740–759.

Sharpe, W. F., Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance*, 1964, 19, 425–442.

Shiller, R. J., Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 1981, 71, 421–435.

Shiller, R. J., From efficient markets theory to behavioral finance, *Journal of Economic Perspectives*, 2003, 17, 83–104.

Shiller, R. J., and P. Perron, Testing the random walk hypothesis: Power versus frequency of observation, *Economics Letters*, 1985, 18, 381–386.

Simon, H., *Models of Bounded Rationality*. Vol. 3, MIT Press, 1997.

Stracca, L., Does liquidity matter? Properties of a divisia monetary aggregate in the euro area, *Oxford Bulletin of Economics and Statistics*, 2004, 66, 309–331.

Summers, L. H., Does the stock market rationally reflect fundamental values? *Journal of Finance*, 1986, 41, 591–601.

Taylor, M., and H. Allen, The use of technical analysis in the foreign exchange market, *Journal of International Money and Finance*, 1992, 11, 304–314.

Wenzelburger, J., Learning to predict rationally when beliefs are heterogeneous, *Journal of Economic Dynamics and Control*, 2004, 28, 2075–2104.

Wenzelburger, J., Perfect forecasting, behavioral heterogeneties, and asset prices, *Handbook of Finanical Markets: Dynamics and Evolution*, 2009.

Westerhoff, F., Multiasset market dynamics, *Macroeconomic Dynamics*, 2004, 8, 596–616.

Westerhoff, F., and R. Dieci, The effectiveness of Keynes-Tobin transaction taxes when heterogeneous agents can trade in different markets: A behavioral finance approach, *Journal of Economic Dynamics and Control*, 2006, 30, 293–322.

Zeeman, E., On the unstable behavior of stock exchanges, *Journal of Mathematical Economics*, 1974, 1, 39–49.

Ziegler, A., *Incomplete Information and Heterogeneous Beliefs in Continuous-Time Finance*. Springer-Verlag, Berlin, 2003.

Zschischang, E., and T. Lux, Some new results on the Levy, Levy and Solomon microscopic stock market model, *Physica A*, 2001, 291, 563–573.

CHAPTER 6

Perfect Forecasting, Behavioral Heterogeneities, and Asset Prices

Jan Wenzelburger
Keele University

6.1. Introduction 346
6.2. The CAPM as a Two-Period Equilibrium Model 349
6.2.1. Portfolio Selection with One Risk-Free Asset 349
6.2.2. Two-Fund Separation 351
6.2.3. Existence and Uniqueness of Equilibrium 355
6.2.4. Equilibria with Heterogeneous Beliefs 358
6.3. Heterogeneous Beliefs and Social Interaction 359
6.3.1. Temporary Equilibria 359
6.3.2. Perfect Forecasting Rules 362
6.3.3. Systematic and Nonsystematic Risk 366
6.3.4. Selecting Mediators 369
6.3.5. Dynamic Stability with Rational Expectations 371
6.4. Multiperiod Planning Horizons 374
6.4.1. Overlapping Cohorts of Investors 375
6.4.2. Temporary Equilibria 378
6.4.3. Perfect Forecasting Rules 379
6.4.4. Portfolio Holdings 385
6.5. Nonergodic Asset Prices 387
6.5.1. Characterization of Long-Run Equilibria 389
6.5.2. Convergence to Long-Run Equilibria 390
6.5.3. Performance of Efficient Portfolios 391
6.5.4. A Boom-and-Bust Scenario 394
6.6. Conclusion 397
References 398

Acknowledgment: I would like to thank Marten Hillebrand, Ulrich Horst, Gauthier Lanot, Klaus Reiner Schenk-Hoppé, Tim Worrall, and Volker Böhm for very valuable comments and discussions. All errors are mine.

Abstract

This survey reviews a dynamic multiasset framework in which heterogeneous agents with multiperiod planning horizons interact. The framework distinguishes between *temporary equilibrium maps* describing the basic market mechanism of an asset market, *forecasting rules* that model the way in which expectations are formed, and a model for exogenous *random perturbations*. Perfect forecasting rules that provide correct forecasts for first and second moments of future prices are introduced. Based on these perfect forecasting rules, fundamental concepts of the traditional CAPM are extended to a setting in which beliefs are heterogeneous. We review a multifund separation theorem and introduce the notion of a generational portfolio that is held by investors with homogeneous beliefs and identical planning horizons. It is shown that social interaction among consumers may endogenously create risk, leading to nonergodic behavior of asset prices. The stochastic dynamics of asset prices, beliefs, portfolio holdings, and market shares are illustrated with numerical simulations.

Keywords: CAPM, agent heterogeneity, market selection, nonergodicity, price dynamics, rational expectations, social interaction, systemic risk

6.1. INTRODUCTION

In recent years research in financial economics has increasingly focused on the interaction between heterogeneous agents as a key force that drives the evolution of financial markets. It is now a widely accepted view that agents in financial markets trade not only because they differ in their attitudes toward risk or their wealth positions but also because they disagree on how markets will evolve in the future. This view stands in contrast to the rational expectations paradigm, which assumes that agents make no systematic errors in predicting the future.

The rational expectations paradigm is one of the tenets in at least three strands of literature on financial markets. First, much of the finance literature has used the rational expectations paradigm to substitute market equilibrium conditions with no-arbitrage conditions. In this context, asset prices are assumed to follow an exogenously given stochastic process to which agents' expectations and their trading activities adjusts without any feedback on the process (see Merton, 1973). An excellent account on this approach is Duffie (1996). Second, initiated by Lucas (1978), dynamic general equilibrium models have used the rational expectations paradigm to reduce the heterogeneities of agents to a single infinitely lived representative agent. These models have been extended in many ways and are applied to a wide range of questions. Early contributions in this direction are Mehra and Prescott (1980, 1985) and Cox, Ingersoll, and Ross (1985). The capital asset-pricing model (CAPM), developed by Sharpe (1964), Lintner (1965), and Mossin (1966), has been extended by Magill and Quinzii (1996, 2000) to a stochastic dynamic equilibrium setting. Kübler and Schmedders (2003) explored the role of assets with collateral. The extent to which agents who do not learn to make accurate predictions are driven out of the market has been analyzed by Blume and Easley (1992, 2002), Sandroni (2000), and others. Angeletos and Calvet (2005, 2006) developed a macroeconomic extension with heterogeneous agents.

Due to the mathematical complexity of these models, however, the diversity of agents regarding preferences and beliefs is relatively limited. Their basic planning character seems to leave relatively little scope for market interaction between heterogeneous agents. In fact, according to Cox, Ingersoll, and Ross (1985), an interpretation as an intertemporal market mechanism requires a set of relatively strong economic assumptions. Moreover, as noted by Judd, Kübler, and Schmedders (2003), such dynamic general equilibrium models seem to have difficulties in matching trading activities that are usually observed in markets.

Third, quite a number of authors have successfully integrated asset markets into models with overlapping generations. Early contributions are Huberman (1984) and Huffman (1985, 1986). More recently, Eckwert (1992), Orosel (1996, 1997, 1998), and Kübler and Polemarchakis (2004) have investigated a number of aspects of financial markets in an OLG setting. Agents in these models are described as expected-utility maximizers. Trade in OLG models occurs naturally, but the assumption of rational expectations provides no operational description of the way expectations are formed on the basis of observations.

From a theoretical perspective, the rational expectations paradigm is unsatisfactory. It ascribes unreasonably strong assumptions on agents' capabilities to acquire information and to foresee the future. Agents in financial markets have different expectations concerning the future evolution of asset prices, simply because they may have distinct information on the assets they are trading. If these agents participate on different sides of the market, the evolution of asset prices cannot be self-confirming for all of them. Situations in which agents correctly anticipate the distribution of future asset prices should for this reason be the special, rather than the general, case of a descriptive model. The assumption of rational expectations, by contrast, offers no insight into the feedback of erroneous expectations concerning the price process. Indeed, the equilibrium concept adopted in both classes of intertemporal models is one in which the notion of expectations is reduced to the assumption of mutual consistency between expectations and realizations. As Lucas puts it: "… *the system is closed with the assumption of rational expectations: the market clearing price function p implied by consumer behavior is assumed to be the same as the price function p on which consumer decisions are based*" (1978, p. 1431). This precludes a description of the way expectations are formed.

The literature on financial markets has responded to the dissatisfaction with the rational expectations paradigm with different lines of research, often carried out simultaneously. Models with heterogeneous boundedly rational agents generally have three structural elements in common. First, a *temporary equilibrium map* determines asset prices in each period, given the characteristics of investors such as preferences and subjective beliefs regarding the future evolution of markets. Second, *forecasting rules* stipulate the way in which agents form and update these beliefs. Third, a model for exogenous perturbations captures all influences that are not modeled explicitly, such as dividend payments, random endowments, or noise-traders' activities.

The first two components constitute a deterministic dynamical system in which agents' beliefs feed back into the actual evolution of asset prices and portfolio holdings. Together with the third component, one obtains a deterministic dynamical system in a random environment. The temporary equilibrium map is most often understood in the sense of Grandmont (1982) and determines market-clearing prices from an aggregate excess demand function. This demand function may be derived from expected-utility maximizing investors as in, for example, Böhm, Deutscher, and Wenzelburger (2000) and Horst (2005), or from mean-variance preferences as in Brock and Hommes (1997a, 1998), Chiarella and He (2002), Wenzelburger (2004), or Böhm and Chiarella (2005). The temporary equilibrium map may be replaced by an *economic law* in the sense of Böhm and Wenzelburger (1999) to include other price mechanisms such as the "market-maker" scenarios considered in Chiarella and He (2003), Chiarella, Dieci, and He (2007), or Horst and Rothe (2008). Some authors take individual demand functions as primitive objects of the model; see, for example, Evstigneev, Hens, and Schenk-Hoppé (2002, 2006, 2008).

Most behavioral models use forecasting rules either explicitly or implicitly to model the way in which agents form and update their beliefs. This behavioral approach is reconciled with the rational expectations paradigm in Wenzelburger (2004) by introducing *perfect forecasting rules*, which generate rational expectations for a group of investors in the sense that the first two moments of the price process are correctly anticipated, whereas the beliefs of other market participants may be erroneous. It turns out that perfect forecasting rules for first and second moments are the key to extending fundamental concepts of the traditional CAPM to a setting with diverse beliefs.

The concept that investors choose between portfolio strategies according to some performance indicator has become popular since Brock and Hommes (1997a, 1998). This choice behavior is mostly modeled by a standard discrete-choice approach, as described in Anderson, de Palma, and Thisse (1992). It adds an evolutionary feature to behavioral models, thereby generating a random environment for asset prices. This random environment is created by the *endogenous* process that governs the choice behavior of agents and exists in addition to the randomness of an *exogenous* stochastic process that models the effects of noise trading, dividend payments, or random endowments. As agents' choices depend on performance indicators, asset prices feed back into the random environment. It is this feedback effect that distinguishes this approach from the models of Evstigneev, Hens, and Schenk-Hoppé (2002, 2006, 2008), Blume and Easley (1992, 2002), and Sandroni (2000), in which investors never change their portfolio strategies.

If the rational expectations paradigm is not immediately discarded, it is sensible to distinguish conceptually between two levels of rationality—first, rationality in the sense of Sargent (1993), and second, rationality in the sense of Simon (1982). In Wenzelburger (2004), a stylized fund manager is characterized by a forecasting technology and carries out portfolio decisions. She is rational in the sense of Sargent and might want to improve her ability to make accurate predictions of the future. Consumers are rational in the sense of Simon. They base their investment choice on a performance indicator that measures the success a fund manager's investment strategy has had in the past. For this setup, Horst and Wenzelburger (2008) show that an asset price process is *ergodic* as long as the dependence of agents' investment decisions on performance indicators is sufficiently weak. This ergodicity disappears if interactive complementarities between agents are too powerful. Although asset prices still converge, their asymptotic behavior now becomes path-dependent. This result is consistent with many findings in the social interaction literature—such as those of Blume (1993), Brock and Durlauf (2001), or Horst and Scheinkman (2006, 2008), which establish uniqueness of equilibria for weak interactions and show that strong interactive complementarities may generate nonergodic dynamics. The chapter develops a common framework for the deterministic approach initiated by Brock and Hommes (1997a, 1998), with its rich dynamics, and the probabilistic approach of Föllmer and Schweizer (1993), Föllmer, Horst and Kirman (2005), and Horst (2005), who obtain rigorous mathematical results but use assumptions that preclude many interesting phenomena such as nonergodic dynamics.

The chapter unfolds as follows: Section 6.2 reviews some classical results from the static CAPM to provide the foundation for a stochastic overlapping generations model, which is discussed in Section 6.3. An extension to a model with overlapping cohorts of investors with *multiperiod* lives is discussed in Section 6.4. Section 6.5 discusses the nonergodic behavior of asset prices; Section 6.6 concludes.

6.2. THE CAPM AS A TWO-PERIOD EQUILIBRIUM MODEL

The capital asset-pricing model as developed by Sharpe (1964), Lintner (1965), and Mossin (1966) is at the center of modern finance. It is an equilibrium theory built on the foundations of the portfolio theory initiated by Markowitz (1952) and Tobin (1958). The standard CAPM is a static two-period model that describes how investors trade infinitely many assets to transfer wealth into the future. Comprehensive treatises of the CAPM with different emphases are presented (e.g., in Cochrane, 2005; Cuthbertson, 1996; or LeRoy and Werner, 2001). In light of recent developments in dynamic models with heterogeneous interacting agents, this section reviews some of its basic properties. Following Nielsen (1987, 1988, 1990a), Dana (1999), Hens, Laitenberger, and Löffler (2002), Böhm and Chiarella (2005), and references therein, it is appropriate to adopt a formulation in terms of prices rather than returns. Mean-variance behavior of investors can be made consistent with expected utility maximization when von Neumann–Morgenstern utility functions are combined with the appropriate probability distributions. As described by Chamberlain (1983) and Owen and Rabinovitch (1983), these include normal and elliptical distributions. To obtain an analytically tractable model with multiple risky assets, this survey treats mean-variance behavior as an alternative rather than a special case of expected-utility maximization.

6.2.1. Portfolio Selection with One Risk-Free Asset

Consider a two-period model in which an investor needs to transfer her initial wealth from the first into the second period. The investment opportunities are K risky assets and one risk-free bond. All prices and payoffs are denominated in a nonstorable consumption good that serves as the numéraire. The K risky assets are characterized by stochastic gross returns $\tilde{q} = (\tilde{q}^{(1)}, \ldots, \tilde{q}^{(K)})$ per unit, which take values in $\mathbb{R}_+^K$. The risk-free bond pays a constant return $R_f = 1 + r_f > 0$ per unit. A portfolio is represented by a vector $(x, y) \in \mathbb{R}^K \times \mathbb{R}$. The vector $x = (x^{(1)}, \ldots, x^{(K)})$ represents the portfolio of risky assets, with $x^{(k)}$ denoting the number of shares of the k^{th} risky asset. The scalar y describes the number of risk-free bonds in the portfolio (x, y). The total amount of risky assets is $x_m \in \mathbb{R}_+^K$ and is referred to as the *market portfolio* of the economy.

Assume, for simplicity, that the investor's wealth consists of $e > 0$ units of the consumption good, that he does not consume in the first period, and that there are no short-sale constraints. If $p = (p^{(1)}, \ldots, p^{(K)}) \in \mathbb{R}^K$ denotes the price vector of risky assets, the investor's budget constraint is

$$e = y + \langle p, x \rangle = y + \sum_{k=1}^{K} p^{(k)} x^{(k)}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product on $\mathbb{R}^K$. Substituting for y, the investor's second-period wealth associated with the portfolio $x \in \mathbb{R}^K$ of risky assets becomes

$$w(e, p, \tilde{q}, x) = R_f e + \langle \tilde{q} - R_f p, x \rangle$$

The uncertainty of second-period wealth rests with the random gross return $\tilde{q}$ of risky assets, when the investor treats the asset price p of the first period as a parameter of his decision problem. This uncertainty is described by a probability space $(\mathbb{R}^K, \mathcal{B}, \nu)$, where $\nu \in \text{Prob}(\mathbb{R}^K)$ is a probability distribution for $\tilde{q}$ and $\text{Prob}(\mathbb{R}^K)$ denotes the set of all Borelian probability measures.

Rather than using the full probability distribution ν, the CAPM assumes that an investor bases his evaluation of the uncertain return solely on the mean and the variance of second-period wealth. Denote *expected gross returns* by

$$\overline{q} = \mathbb{E}[\tilde{q}] := \int_{\mathbb{R}^K} q \, \nu(\mathrm{d}q) \in \mathbb{R}^K$$

and the *(variance)-covariance matrix* of future returns by

$$V = \mathbb{V}[\tilde{q}] := \int_{\mathbb{R}^K} [q - \overline{q}][q - \overline{q}]^\top \nu(\mathrm{d}q) \in \mathcal{M}_K$$

where $\mathcal{M}_K$ denotes the set of all symmetric and positive definite $K \times K$ matrices. The kl^{th} entry of the V is the covariance $V_{kl} = \mathbb{C}\text{ov}[\tilde{q}^{(k)}, \tilde{q}^{(l)}]$ between the gross returns of the k^{th} and the l^{th} risky asset. The positive definiteness of V ensures that there are no redundant assets in the market. Based on ν, the expected wealth and its standard deviation associated with a portfolio of risky assets $x \in \mathbb{R}^K$ becomes

$$\mu_w(e, p, x) := \mathbb{E}\big[w(e, p, \cdot, x)\big] = R_f e + \langle \overline{q} - R_f p, x\rangle$$

and

$$\sigma_w(x) := \mathbb{V}\mathrm{ar}\big[w(e, p, \cdot, x)\big]^{\frac{1}{2}} = \langle x, V x\rangle^{\frac{1}{2}}$$

respectively.[1] The investment behavior of an investor is now based on the following assumption.

Assumption 6.1. *An investor is characterized by a utility function U, which is a function of the mean and the standard deviation of future wealth and a probability distribution ν for future gross returns. These satisfy the following:*

1. *The utility function $U : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is continuously differentiable, strictly increasing in μ, strictly decreasing in σ, and strictly concave.*[2]
2. *The probability distribution $\nu \in \mathrm{Prob}(\mathbb{R}^K)$ is parameterized by a pair $(\overline{q}, V)$, where $\overline{q} \in \mathbb{R}^K$ is the mean return and $V \in \mathcal{M}_K$ describes the covariance structure.*

With this notation, the decision problem of an investor takes the form

$$\max_{x \in \mathbb{R}^K} U\big(\mu_w(e, p, x), \sigma_w(x)\big) \tag{6.1}$$

Assumption 6.1 implies that the objective function in Eq. 6.1 is strictly concave in x. To ensure boundedness of the asset demand derived from Eq. 6.1, we need the concept of a *limiting slope* of an indifference curve as adopted in Nielsen (1987). Recall that the slope of any indifference curve in the $\mu - \sigma$ plane is given by the marginal rate of substitution between risk and return $-\frac{\frac{\partial U}{\partial \sigma}(\mu,\sigma)}{\frac{\partial U}{\partial \mu}(\mu,\sigma)}$, which is a measure of the investor's risk aversion. Consider the indifference curve through the point $(R_f e, 0)$ and denote by

$$\rho_U := \sup\left\{ -\frac{\frac{\partial U}{\partial \sigma}(\mu, \sigma)}{\frac{\partial U}{\partial \mu}(\mu, \sigma)} \; : \; (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+ \text{ s.t. } U(\mu, \sigma) = U(R_f e, 0) \right\}$$

the limiting slope of this indifference curve. From convex analysis it is well known that ρ_U is either positive and finite or plus infinity. Since U is concave, all indifference curves have the same limiting slope ρ_U (see Rockafellar, 1970).

6.2.2. Two-Fund Separation

The key property of the asset demand function derived from the optimization problem (Eq. 6.1) is known as the *two-fund separation theorem* or the *mutual fund theorem*. This theorem and its implications will be reviewed next. Let $\pi := \overline{q} - R_f p$ be the vector

[1] To relate the notation to the one adopted in the incomplete markets literature (e.g., LeRoy and Werner, 2001), note that each function $q \mapsto w(e, p, q, x)$ is an element of the Hilbert space $L^2(\nu)$.

[2] To include interesting examples, it is sometimes convenient to bound the domain of U from below.

of expected excess returns and consider the following mean-variance optimization problem:

$$\max_{x \in \mathbb{R}^K} \mu_w(e, p, x) \quad \text{s.t.} \quad \sigma_w(x) \le \sigma \tag{6.2}$$

where $\sigma \ge 0$. Markowitz (1952) found that for $\pi \ne 0$, the unique portfolio maximizing Eq. 6.2 is given by

$$x_{\text{eff}}(\sigma) := \frac{\sigma}{\langle \pi, V^{-1}\pi \rangle^{\frac{1}{2}}} V^{-1}\pi$$

The portfolio $x_{\text{eff}}(\sigma)$ is *(mean-variance) efficient* because it yields the highest expected wealth given the upper bound σ on the standard deviation of wealth $\sigma_w(x)$ of any portfolio x. The standard deviation of the wealth associated with $x_{\text{eff}}(\sigma)$ is σ. Thus, the risk expected wealth characteristics of any portfolio $x_{\text{eff}}(\sigma)$, $\sigma \ge 0$ satisfy the linear relationship

$$\mu_w\big(e, p, x_{\text{eff}}(\sigma)\big) = R_f e + \rho\,\sigma \tag{6.3}$$

where $\rho := \langle \pi, V^{-1}\pi \rangle^{\frac{1}{2}}$ is called the *market price of risk*. Observe that Eq. 6.3 is nothing but the *efficient frontier* expressed in terms of wealth rather than returns. Indeed, if $r = \frac{w(e,p,q,x_{\text{eff}}(\sigma))}{e} - 1$ denotes the return of $x_{\text{eff}}(\sigma)$, then Eq. 6.3 implies the well-known efficient frontier

$$\mu_r = r_f + \rho\,\sigma_r$$

stating that the relationship between the expected rate of return μ_r and the standard deviation σ_r of any efficient portfolio is linear.

The efficient frontier (Eq. 6.3), also referred to as the *capital market line*, is now used to formulate the *(two-fund) separation theorem.*

Theorem 6.1. *Under the hypotheses of Assumption 6.1, for any $e > 0$ and any $0 \ne \pi \in \mathbb{R}^K$ with $\langle \pi, V^{-1}\pi \rangle^{\frac{1}{2}} < \rho_U$, the optimization problem (Eq. 6.1) has a unique maximizer $x_\star \in \mathbb{R}^K$. The corresponding asset demand function Φ takes the form*

$$x_\star = \Phi(e, \pi, V) := \frac{\varphi\big(e, \langle \pi, V^{-1}\pi \rangle^{\frac{1}{2}}\big)}{\langle \pi, V^{-1}\pi \rangle^{\frac{1}{2}}} V^{-1}\pi \tag{6.4}$$

where for each $\rho \in [0, \rho_U)$,

$$\varphi(e, \rho) := \operatorname*{argmax}_{\sigma \ge 0} U\big(R_f e + \rho\,\sigma, \sigma\big) \tag{6.5}$$

is bounded from above.

The separation theorem states that given π and V, the optimal portfolio $x_\star$ is collinear to $x_{\text{eff}}(1)$ and hence efficient. Thus investors with the same beliefs $(\overline{q}, V)$ will invest in

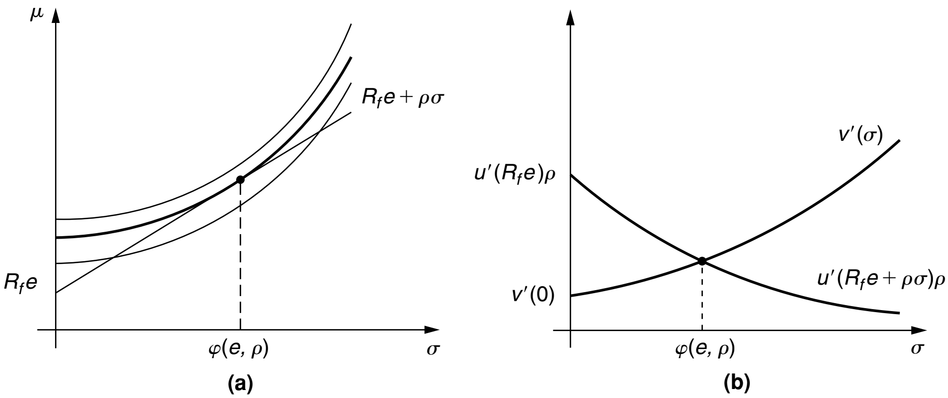

FIGURE 6.1 Portfolio selection: (a) separation principle; (b) optimality with separable utility.

the same two funds: the risk-free asset and a "mutual fund" with the same mix of risky assets $x_{\text{eff}}(1)$. Since $\sigma_w(x_\star) = \varphi(e, \rho)$, the function φ describes the investor's willingness to take risk. Unlike the portfolio mix $x_{\text{eff}}(1)$, this willingness and hence the amount of the endowment invested into risky assets depends on the investor's preferences.

The separation theorem is illustrated in Figure 6.1a showing that in an optimum, the marginal rate of substitution between risk and return has to be equal to the market price of risk ρ. An intuitive proof of the theorem may be found in any finance textbook (e.g., see Copeland, Weston, and Shastri, 2005). The theorem was first proved by Tobin (1958) and then by Lintner (1965) and Merton (1972). The formulation of Theorem 6.1 is, in essence, that of in Böhm and Chiarella (2005, Lemma 2.3). It has been applied by Rochet (1992) and others to describe the behavior of commercial banks.[3] The important assumption here is that the investor may short-sell risky assets as well as borrow and lend at the risk-free rate r_f. Short-sale and borrowing constraints are, for example, investigated in Lintner (1965), Black (1972), and Rochet (1992).

It has long been recognized in the literature that the investor's decision problem of the CAPM can be reduced to the problem of finding an optimal allocation of two artificial commodities: mean return μ and risk σ (e.g., see Dana, 1999; LeRoy and Werner, 2001; or Hens, Laitenberger, and Löffler, 2002). Using the separation theorem, one may now apply standard microeconomic consumer theory to the decision problem (Eq. 6.5). For the special case of separable utility functions of the form $U(\mu, \sigma) = u(\mu) - v(\sigma)$, the following result summarizes several findings of the literature.

[3]See also Freixas and Rochet (1997, Chap. 8) and references therein. The fact that the objective function in Eq. 6.5 has no upper bound for $\rho \geq \rho_U$ so that the asset demand Φ is then undefined has been recognized by Nielsen (1987). See also further references therein. An elementary proof of Theorem 6.1 may be obtained from a comparison of the first-order conditions of the two maximization problems (Eq. 6.1 and Eq. 6.5; see Wenzelburger, 2008).

Proposition 6.1. *Under the hypotheses of Theorem 6.1, assume that U is twice continuously differentiable and of the form*

$$U(\mu, \sigma) = u(\mu) - v(\sigma), \quad \mu \in \mathbb{R},\ \sigma \in \mathbb{R}_+$$

Let $R_f e > 0$ be arbitrary but fixed. Then $\varphi(e, \cdot)$ is differentiable on $(0, \rho_U)$ except for $\rho = \frac{v'(0)}{u'(R_f e)}$. Moreover:

1. *$\varphi(e, \rho) = 0$ for all $0 \leq \rho \leq \frac{v'(0)}{u'(R_f e)}$.*
2. *If $\lim_{\mu \to \infty} u'(R_f e + \mu)\mu = \infty$, then for each risk $\sigma \in \mathbb{R}_+$, there exists a market price of risk $\rho_\sigma \in [0, \rho_U)$ such that $\varphi(e, \rho_\sigma) = \sigma$.*
3. *$\varphi(e, \rho)$ is strictly increasing for all $\frac{v'(0)}{u'(R_f e)} < \rho < \rho_U$, if and only if the elasticity of $u'(R_f e + \mu)$ satisfies $\frac{u''(R_f e+\mu)\mu}{u'(R_f e+\mu)} > -1$ for all $\mu > 0$.*
4. *$\varphi(e, \rho)$ is decreasing in e for each $\frac{v'(0)}{u'(R_f e)} < \rho < \rho_U$, if u is strictly concave.*

The proof is derived from the first-order conditions and illustrated in Figure 6.1b. Assertion 2 gives a condition under which φ is surjective on $\mathbb{R}_+$, implying that an investor is prepared to take any risk, provided that the market price of risk is high enough. The condition has, in essence, been given in Böhm (2002, Thm. 2.2). It corresponds to Dana (1999, Prop. 3.1), who showed that the surjectivity of φ is the crucial property for the existence of asset-market equilibria. The elasticity condition of Assertion 3 guarantees that φ is strictly increasing in ρ and guarantees uniqueness of asset-market equilibria. This condition is due to Dana (1999, Prop. 3.4). For general utility functions $U(\mu, \sigma)$, Hens, Laitenberger, and Löffler (2002, Lemma 1) showed that φ is increasing in ρ, if the slope of the indifference curves is non-increasing in means. Assertion 4 states that risk is an inferior good whenever *u* is strictly concave and a result attributed to Lajeri and Nielsen (2000).

Before addressing the issue of existence of equilibria, let us illustrate Proposition 6.1 with three examples. Consider first the ubiquitous and famous case

$$U(\mu, \sigma) = \mu - \tfrac{1}{2a}\sigma^2 \tag{6.6}$$

where $a > 0$ denotes risk tolerance. Then $\varphi(e, \rho) = a\rho$ is a linear and hence surjective function with limiting slope $\rho_U = \infty$. Moreover, φ is independent of *e*. This is the case for all quasi-linear functions of the form $U(\mu, \sigma) = \mu - v(\sigma)$ with $v'(0) = 0$, so that risk is a *normal* good.

Second, risk can be decreasing in the market price of risk. This is seen from considering

$$U(\mu, \sigma) = \frac{\mu}{1 + \mu} - \sigma, \quad \mu > -1,\ \sigma \in \mathbb{R}_+ \tag{6.7}$$

The limiting slope is $\rho_U = \infty$ and the willingness to take risk is given by

$$\varphi(e, \rho) = \max\Big\{ \tfrac{1}{\rho}\big[\sqrt{\rho} - (R_f e + 1)\big], 0 \Big\}, \quad \rho \in \mathbb{R}_+$$

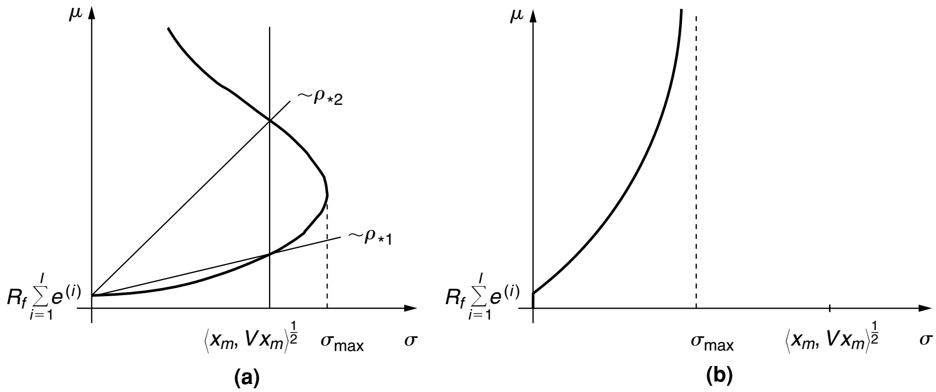

FIGURE 6.2 Multiplicity and nonexistence of equilibria: (a) two equilibria, $\rho_{\star 1}$ and $\rho_{\star 2}$, $U^{(i)}(\mu,\sigma) = \frac{\mu}{1+\mu} - \sigma$; (b) no equilibrium, $U^{(i)}(\mu,\sigma) = \ln\mu - \sigma$.

For each $e > 0$, the map $\varphi(e,\rho)$ is unimodal with respect to ρ and attains its maximum at $\rho_{\max} = 4(R_f e + 1)^2$. The maximum risk the investor is willing to take is $\sigma_{\max} = \frac{1}{4}(R_f e + 1)^{-1}$, so that $\varphi(e,\cdot)$ is not surjective on $\mathbb{R}_+$. Observe that both conditions stated under 2 and 3 of Proposition 6.1 are violated in this case. In the $\mu - \sigma$ plane, the "offer curve" for risk is backward bending, as in Figure 6.2a.

Third, the limiting slope ρ_U may be finite. Consider

$$U(\mu,\sigma) = \sqrt{(\mu^2 + 2\mu)} - \sigma, \quad (\mu,\sigma) \in \mathbb{R}^2_+ \tag{6.8}$$

Then $\rho_U = 1$ and

$$\varphi(e,\rho) = \max\left\{ \frac{1}{\rho}\left[\frac{1}{\sqrt{1-\rho^2}} - (R_f e + 1)\right], 0 \right\}, \quad \rho \in [0,1)$$

An investor characterized by Eq. 6.8 is prepared to take any risk $\sigma \in \mathbb{R}_+$ as $\varphi(e,\cdot)$ is surjective on $\mathbb{R}_+$. However, in this case φ, and hence the corresponding asset demand (Eq. 6.4) is undefined for all $\rho \geq 1$.

6.2.3. Existence and Uniqueness of Equilibrium

The literature has addressed the existence and uniqueness of asset-market equilibria in the traditional CAPM and its extensions with great generality (e.g., see Nielsen, 1988, 1990a,b; Allingham, 1991; Dana, 1993, 1999; or Hens, Laitenberger, and Löffler, 2002). For the case under consideration, we follow a basic line of reasoning in Dana (1999, Sec. 3), which has been adapted in Böhm (2002).

Consider an asset market with $i = 1, \ldots, I$ investors who are all characterized by Assumption 6.1. Suppose that utility functions and endowments are heterogeneous but that their expectations regarding future gross returns are identical and given by $(\overline{q}, V)$. Let $e^{(i)}$ denote investor i's endowment and $\varphi^{(i)}$ be her willingness to take risk derived from some $U^{(i)}$ satisfying Assumption 6.1. For fixed endowments $e^{(i)}$, define *aggregate willingness to take risk* by

$$\phi(\rho) := \sum_{i=1}^{I} \varphi^{(i)}(e^{(i)}, \rho), \qquad \rho \in [0, \overline{\rho}) \tag{6.9}$$

where $\overline{\rho} := \min\{\rho_U^{(i)} : i = 1, \ldots, I\}$ is the minimum of all limiting slopes $\rho_U^{(i)}$ of $U^{(i)}$. Using Theorem 6.1, the asset-market equilibrium condition takes the form

$$\frac{\phi(\langle \pi, V^{-1}\pi\rangle^{\frac{1}{2}})}{\langle \pi, V^{-1}\pi\rangle^{\frac{1}{2}}} V^{-1}\pi = x_m \tag{6.10}$$

where $\pi = \overline{q} - R_f p \neq 0$ is the vector of excess returns. Computing standard deviations of the wealth associated with x_m and with the portfolio on the LHS of Eq. 6.10, $\pi_\star$ is a solution to Eq. 6.10, if $\rho_\star := \langle \pi_\star, V^{-1}\pi_\star\rangle^{\frac{1}{2}} \in (0, \overline{\rho})$ solves

$$\phi(\rho) = \langle x_m, V x_m\rangle^{\frac{1}{2}} \tag{6.11}$$

Vice versa, if some $\rho_\star \in (0, \overline{\rho})$ solves (6.11), then $\pi_\star := \frac{\rho_\star}{\langle x_m, V x_m\rangle^{\frac{1}{2}}} V x_m$ is a solution to Eq. 6.10. Hence, in equilibrium, aggregate willingness to take risk must be equal to the aggregate risk of the market $\langle x_m, V x_m\rangle^{\frac{1}{2}}$. Define the upperbound bound of risk the investors are willing to accept by $\sigma_{\max} := \sup\{\phi(\rho) : \rho \in [0, \overline{\rho})\}$. The existence of $\rho_\star$ now follows from the intermediate-value theorem, observing that $\phi(0) = 0$. Summarizing, the following result is a refinement of Böhm and Chiarella (2005, Lemma 2.5) that includes the case in which the limiting slope $\overline{\rho}$ is finite.

Theorem 6.2. *Let $(\overline{q}, V)$ and $e^{(1)}, \ldots, e^{(I)} > 0$ be given. Assume that aggregate willingness to take risk $\phi : [0, \overline{\rho}) \to \mathbb{R}_+$ is a continuous map with respect to ρ. Then the following holds:*

1. *For each $0 \neq x_m \in \mathbb{R}_+^K$ with $\langle x_m, V x_m\rangle^{\frac{1}{2}} < \sigma_{\max}$, there exists an asset-market equilibrium with market-clearing prices:*

$$p_\star = \frac{1}{R_f}\left[\overline{q} - \frac{\rho_\star}{\langle x_m, V x_m\rangle^{\frac{1}{2}}} V x_m\right] \tag{6.12}$$

   where $\rho_\star \in (0, \overline{\rho})$ is a solution to Eq. 6.11.

2. *If, in addition, ϕ is strictly increasing with respect to all ρ for which $\phi(\rho) > 0$, then the asset-market equilibrium (Eq. 6.12) is uniquely determined.*

The pricing rule (Eq. 6.12) reveals three features of the CAPM. First, the market portfolio x_m is efficient and in equilibrium any investor will hold a proportion of x_m, that is., a portfolio with the same mix of risky assets as x_m. Second, the equilibrium price of risk $\rho_\star$ responds to changes in second-moment beliefs V but not to changes in first-moment beliefs $\overline{q}$. As a consequence, any change in first-moment beliefs $\overline{q}$

changes the corresponding market-clearing asset prices in a linear fashion, irrespective of any nonlinearities in investors' utility functions. Third, the equilibrium price of risk is bounded by the lowest limiting slope $\overline{\rho}$. This confirms that Nielsen's earlier requirement that all limiting slopes be infinite is not necessary.

Theorem 6.2 reduces the existence and uniqueness of an equilibrium in K markets to an invertibility condition of a one-dimensional demand function Eq. 6.9. It is illustrated in Figure 6.2 with two panels displaying "aggregate offer curves" for risk. Panel (a) depicts a backward-bending offer curve implying the existence of multiple equilibria. Panel (b) depicts a situation in which aggregate willingness to take risk is increasing but not surjective. Then an asset-market equilibrium does not exist if the aggregate risk of the market $\langle x_m, V x_m \rangle^{\frac{1}{2}}$ is above the upper bound of risk $\sigma_{\max}$ that investors are prepared to accept.

Eq. 6.12 is a vector version of what is known as the *certainty equivalent pricing formula* of the CAPM (see Luenberger, 1998). To see this, denote by $r_m = \frac{\langle q, x_m \rangle}{\langle p_\star, x_m \rangle} - 1$ the return of the market portfolio x_m and by $\sigma_m = \frac{\langle x_m, V x_m \rangle^{\frac{1}{2}}}{\langle p_\star, x_m \rangle}$ its standard deviation. Using the pricing formula (Eq. 6.12), it is readily seen that the expected rate of return μ_m of x_m satisfies $\mu_m = r_f + \rho_\star \sigma_m$, so that the risk-return characteristics of x_m lie on the capital market line. Thus the equilibrium price of the k^{th} asset ($k = 1, \ldots, K$) takes the form

$$p_\star^{(k)} = \frac{1}{R_f} \left[\overline{q}^{(k)} - \left(\frac{\mu_m - r_f}{\sigma_m^2} \right) \mathbb{C}\mathrm{ov}[q^{(k)}, r_m] \right]$$

The term in the bracket is called the *certainty equivalent* of the k^{th} asset because this value may be treated as the certain amount of the asset's proceeds before discounting it to obtain $p_\star^{(k)}$.

The next corollary is immediate from Proposition 6.1 and Theorem 6.2 and a refinement of Böhm (2002, Thm. 3.2). Closely related results are Dana (1999, Prop. 3.4) and Hens, Laitenberger, and Löffler (2002, Thm. 1). The key observation is that aggregate willingness to take risk ϕ is invertible with respect to all $\rho \in (0, \overline{\rho})$, if all individual demand functions $\varphi^{(i)}$ are nondecreasing in ρ with at least one demand function being increasing for positive ρ and surjective on $\mathbb{R}_+$.

Corollary 6.1. *Under the hypotheses of Theorem 6.2, suppose that the willingness to take risk of all investors is nondecreasing in ρ and that the preferences of at least one investor satisfy the conditions of Proposition 6.1 stated in 2 and 3. Then, for any market portfolio $0 \neq x_m \in \mathbb{R}_+^K$, there exists a unique asset-market equilibrium.*

Although Corollary 6.1 is quite elementary, it reveals that two tasks have to be tackled when addressing existence and uniqueness issues in more general setups with heterogeneous beliefs, asset endowments, or both. First, invertibility and surjectivity of the aggregate asset demand function, given beliefs about future gross returns. Second, the specification of preferences that guarantees these two properties. Although the first task can be addressed using the Global Inverse Function Theorem (e.g., see Deimling, 1980, Thm. 15.4, p. 153, or Gale and Nikaido, 1965), the second one is significantly harder, mainly because the dimensionality of the first problem cannot be reduced to one as in Theorem 6.2, if investors have either initial endowments of assets or heterogeneous beliefs.

6.2.4. Equilibria with Heterogeneous Beliefs

Consider now an economy in which the investors have linear mean-variance preferences and heterogeneous beliefs regarding future gross returns of assets. If $(\overline{q}^{(i)}, V^{(i)}) \in \mathbb{R}^K \times \mathcal{M}_K$ denotes investor i's subjective belief, her asset demand function takes the form

$$\Phi^{(i)}(\overline{q}^{(i)}, V^{(i)}, p) = a^{(i)} V^{(i)-1}[\,\overline{q}^{(i)} - R_f p], \quad p \in \mathbb{R}^K \tag{6.13}$$

The market-clearing condition reads

$$\sum_{i=1}^{I} \Phi^{(i)}(\overline{q}^{(i)}, V^{(i)}, p) = x_m \tag{6.14}$$

Inserting Eq. 6.13, we obtain an explicit market-clearing price vector $p_\star$, which solves Eq. 6.14 as follows. For arbitrary second-moment beliefs $V^{(i)} \in \mathcal{M}_K$, $i = 1, \ldots, I$, all $K \times K$ matrices

$$A := \tfrac{1}{R_f}\left(\sum_{i=1}^{I} a^{(i)} V^{(i)-1}\right)^{-1} \quad \text{and} \quad A^{(i)} := a^{(i)} A V^{(i)-1}, \quad i = 1, \ldots, I \tag{6.15}$$

are well defined. Solving Eq. 6.14 for p then yields an explicit *equilibrium map*

$$p_\star = G\big((\overline{q}^{(i)}, V^{(i)})_{i=1}^{I}\big) := \sum_{i=1}^{I} A^{(i)}\,\overline{q}^{(i)} - A x_m \tag{6.16}$$

which determines the market-clearing price vector. The functional form[4] of G reflects the fact that asset prices are essentially determined by subjective beliefs of investors $(\overline{q}^{(i)}, V^{(i)})$. Note for the sake of completeness that for homogeneous beliefs, that is, $\overline{q}^{(i)} = \overline{q}$ and $V^{(i)} = V$ for all $i = 1, \ldots, I$, we recover the pricing rule (Eq. 6.12), which now takes the form

$$p_\star = \tfrac{1}{R_f}\left[\overline{q} - \tfrac{1}{a^{(1)}+\cdots+a^{(I)}} V x_m\right]$$

Despite the existence and uniqueness results for nonlinear mean-variance preferences in the literature, explicit equilibrium maps G, unfortunately, remain the exception.

[4]It is relatively straightforward to establish a map for the case in which each investor trades with a different subset of risky assets.

6.3. HETEROGENEOUS BELIEFS AND SOCIAL INTERACTION

A major caveat of the traditional CAPM with homogeneous beliefs is that it cannot explain trading between investors, because according to the theory investors will hold a certain proportion of the market portfolio that is constant over time. This unrealistic property seems to reappear in many representative agent models (see Judd, Kübler, and Schmedders, 2003). To remedy this caveat, this section adopts a modeling approach of Böhm, Deutscher, and Wenzelburger (2000). The key idea is to extend a static two-period model to a dynamic model by linking together an infinite series of two-period economies. In terms of the CAPM introduced in Section 6.2, this is achieved by first reinterpreting the static equilibrium map (Eq. 6.16) as a *temporary equilibrium map* and then linking prices and allocations of two consecutive periods by specifying the *forecasting rules* investors use to form and update their beliefs.

The model presented here was originally built on a structure with overlapping generations of consumers and developed in Wenzelburger (2004). We can equally well assume that consumers have infinite lives and maximize wealth myopically as in Horst and Wenzelburger (2008) because the results regarding the dynamics of asset prices and allocations are exactly the same. The model includes Böhm and Chiarella (2005) as a special case with homogeneous beliefs, noting that their temporary equilibrium map is of the form of Eq. 6.12 and hence linear in first-moment beliefs. The model may be seen as an extension of Brock and Hommes (1997a, 1998) to *multiple types* of agents and *multiple risky* assets. Multiple risky assets are essential for the investigation of the trade-off between risk and return, because the concept of a Sharpe ratio is only meaningful when more than one risky asset can be traded. Recently, more and more research has been undertaken into the dynamics of markets with multiple assets (e.g., see Chiarella, Dieci, and He, 2007).

This section adopts the assumption that consumers have no direct access to the asset markets and instead select a stylized professional financial investor who solves their individual investment problems. This is done to separate the effects of social interaction among consumers, such as herding and imitation, from the effects of the boundedly rational behavior of a professional investor who, for example, may want to learn systematically about the economic environment in which she lives. For the sake of simplicity, we consider only one type of consumer and two financial investors. The general case with multiple types of consumers and investors is found in Wenzelburger (2004).

6.3.1. Temporary Equilibria

Suppose that there are overlapping generations of consumers who live for two periods. Each consumer receives an initial endowment $e > 0$ of a nonstorable consumption good in the first period of life and does not consume. Her risk-taking behavior is characterized by linear mean-variance preferences with risk tolerance $1/\alpha$ as in Eq. 6.16 and subjective beliefs regarding her future gross return on investment. Assuming that the consumption good cannot be stored directly, each consumer needs to transfer wealth from the first to the second period of life, in which she consumes the proceeds of her investments.

There exist $K+1$ retradeable assets in the economy, indexed by $k=0,1,\ldots,K$. The first asset $k=0$ is a risk-free bond that pays a constant return $R_f>0$ per unit invested in the previous period. The assets $k=1,\ldots,K$ correspond to risky shares of firms that are traded at prices $p_t=(p_t^{(1)},\ldots,p_t^{(K)})\in\mathbb{R}_+^K$ of period t. Shareholders in period t receive a dividend payment of $d_t^{(k)}$ per unit of the k^{th} share. The vector of all dividend payments in period t is denoted by $d_t=(d_t^{(1)},\ldots,d_t^{(K)})\in\mathbb{R}_+^K$.

In each period t, a group of noisetraders who demand the random quantity $\xi_t\in\mathbb{R}^K$ of shares is active in the asset market.[5] The probabilistic prerequisites on the exogenous noise and the exogenous dividend process are stipulated in the following assumption.

Assumption 6.2. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ an increasing family of sub-σ-algebras of $\mathcal{F}$.*

1. *The dividend payments are described by an $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-adapted stochastic process $\{d_t\}_{t\in\mathbb{N}}$ on $(\Omega,\mathcal{F},\mathbb{P})$ with values in $\mathbb{D}\subset\mathbb{R}_+^K$.*
2. *The noise traders' transactions are governed by a $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-adapted stochastic process $\{\xi_t\}_{t\in\mathbb{N}}$ on $(\Omega,\mathcal{F},\mathbb{P})$ with values in $\mathbb{R}^K$, which is uncorrelated with the dividend process $\{d_t\}_{t\in\mathbb{N}}$ defined in 1.*

There are two professional financial investors $i=1,2$, from now on referred to as *mediators*, who are characterized by subjective beliefs regarding the future gross returns of the assets, which are determined by future cum-dividend prices. In each period young consumers are allowed to select a mediator based on his past performance. For simplicity, we abstract from intermediation costs for consumers and suppress the constants e whenever possible.

For later purposes it is convenient to express all aggregate quantities in per capita of young consumers. Denote by $\eta_t\in[0,1]$ the fraction of consumers who employ Mediator 1 in period t and by $1-\eta_t$ the fraction who employ Mediator 2. Then $W_t^{(1)}=\eta_t e$ and $W_t^{(2)}=(1-\eta_t)e$ are the respective amounts of resources per capita that the mediators receive from young consumers before trading takes place in that period. The mediators' earnings from dividend and interest payments from the per-capita portfolio holding $(x_{t-1}^{(i)},y_{t-1}^{(i)})\in\mathbb{R}^K\times\mathbb{R}$ obtained after trading in period $t-1$ are $r_f y_{t-1}^{(i)}+\langle d_t,x_{t-1}^{(i)}\rangle$. Since aggregate repayment obligations to old consumers are $R_f y_{t-1}^{(i)}+\langle p_t+d_t,x_{t-1}^{(i)}\rangle$, their budget constraints in period t reads $W_t^{(i)}=\langle p_t,x^{(i)}\rangle+y^{(i)}$.

Let $(q_t^{(i)},V_t^{(i)})$ be mediator i's subjective beliefs for the future cum-dividend price $q_{t+1}=p_{t+1}+d_{t+1}$ in period t. Then the per-capita demand of mediator i is

$$x_t^{(i)}=\Phi^{(i)}(a_t^{(i)},q_t^{(i)},V_t^{(i)},p)=a_t^{(i)}V_t^{(i)-1}[q_t^{(i)}-R_f p_t] \qquad \textbf{(6.17)}$$

[5] Noise traders will be thought of as traders whose portfolio decisions are not captured by a standard microeconomic decision model. Alternative interpretations as those of De Long, Shleifer, Summers, and Waldmann (1990, p. 709) apply as well.

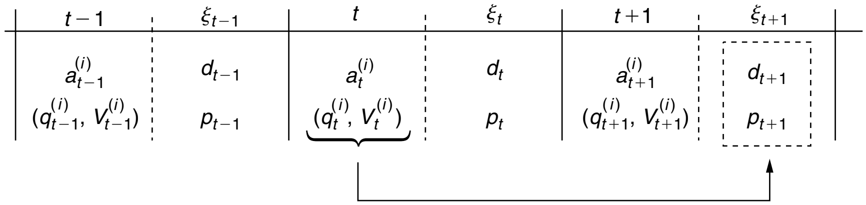

FIGURE 6.3 Timeline of price formation.

with $a_t^{(1)} := \frac{\eta_t}{\alpha}$ and $a_t^{(2)} := \frac{1-\eta_t}{\alpha}$ denoting the mediator's risk-adjusted market shares, respectively. The market-clearing condition of the asset markets in period t reads

$$\sum_{i=1}^{2} \left[\Phi^{(i)}(a_t^{(i)}, q_t^{(i)}, V_t^{(i)}, p_t) - x_{t-1}^{(i)}\right] + \xi_t - \xi_{t-1} = 0 \tag{6.18}$$

where $\xi_t \in \mathbb{R}^K$ denotes the per-capita portfolio holdings of the noise traders after trading in period t.

To solve Eq. 6.18, note that the sum of previous positions $\sum_{i=1}^{2} x_{t-1}^{(i)} + \xi_{t-1}$ must be equal to the total (per-capita) stock of assets $x_m \in \mathbb{R}_+^K$ in the economy. For each t, let

$$a_t = (a_t^{(1)}, a_t^{(2)}) \tag{6.19}$$

be period t's profile of risk-adjusted market shares. Solving the market-clearing condition (Eq. 6.18) for p_t yields a *temporary equilibrium map*, which for arbitrary beliefs $(q_t^{(i)}, V_t^{(i)})$ and market shares a_t takes the form

$$p_t = G\big(\xi_t, a_t, (q_t^{(i)}, V_t^{(i)})_{i=1}^2\big) := A_t^{(1)}\, q_t^{(1)} + A_t^{(2)}\, q_t^{(2)} - A_t(x_m - \xi_t) \tag{6.20}$$

with coefficient matrices

$$A_t := \frac{1}{R_f}\left(a_t^{(1)} V_t^{(1)-1} + a_t^{(2)} V_t^{(2)-1}\right)^{-1} \quad \text{and}$$
$$A_t^{(i)} := a_t^{(i)} A_t V_t^{(i)-1}, \quad i = 1, 2$$

Due to the positive definiteness of all subjective covariance matrices, all coefficient matrices in Eq. 6.20 are well defined and invertible.

The map G determines market-clearing prices in each period, given the beliefs of all investors and the demand of noise traders. It is essentially the same as Eq. 6.16 except that it allows for varying market shares. The crucial assumption here is that there is no a priori dependence of beliefs on current asset prices. Although economically desirable, any such assumption will potentially introduce nonlinearities into the market-clearing condition (Eq. 6.18). This in turn may easily destroy the existence and uniqueness of G. The dependence of beliefs on past prices will be introduced in Section 6.3.2 by adopting the concept of a forecasting rule. It will then be shown that forecasting rules of investors who care about the precision of their forecasts will not depend on asset prices of the current period.

As G does not contain past prices as arguments, the evolution of the asset prices can only be driven by the forecasting technology of the mediators and the way that consumers select among mediators. The dating of the beliefs $(q_t^{(i)}, V_t^{(i)})$ relative to the price p_t in the price law (Eq. 6.20) contains an expectational lead; that is, beliefs are one period ahead of the map G with respect to the realization of prices. This is illustrated in Figure 6.3.

Since the subjective beliefs $(q_t^{(i)}, V_t^{(i)})$ and the market shares a_t must be set prior to trading in period t, they must be based on information observable up to time $t-1$. Mathematically this implies that they are $\mathcal{F}_{t-1}$ measurable. Under Assumption 6.2, the conditional mean values and the conditional covariance matrices of the ex-dividend price[6] are

$$\mathbb{E}_{t-1}[p_t] = A_t^{(1)} q_t^{(1)} + A_t^{(2)} q_t^{(2)} - A_t(x_m - \mathbb{E}_{t-1}[\xi_t]) \tag{6.21}$$

and

$$\mathbb{V}_{t-1}[p_t] = A_t \mathbb{V}_{t-1}[\xi_t] A_t \tag{6.22}$$

respectively. The volatility of ex-dividend prices as well as their correlations between the ex-dividend prices of various risky assets is exclusively generated by the noise-trader behavior, the subjective covariance matrices of the mediators, and the choice behavior of consumers. In particular, this correlation is zero if noise traders are absent so that $\mathbb{V}_{t-1}[\xi_t]$ is the zero matrix. In view of Assumption 6.2, the corresponding second moments of the cum-dividend prices $q_t = p_t + d_t$ are

$$\mathbb{V}_{t-1}[q_t] = \mathbb{V}_{t-1}[p_t] + \mathbb{V}_{t-1}[d_t] \tag{6.23}$$

6.3.2. Perfect Forecasting Rules

A complete description of the evolution of asset prices and portfolios requires a specification of how investors form their expectations. This is done by adopting the concept of a *forecasting rule* in the sense of Grandmont (1982). *A priori*, there is a great degree of freedom in specifying such forecasting rules. Many authors, especially those who contributed to this handbook, use this flexibility in search for a better understanding of how asset markets behave in reality and in search for a better match of empirically observed price patterns.

This section will complement these contributions with a more normative consideration. Consider to this end a benchmark case in which at least one mediator—say, Mediator 2—is able to correctly anticipate future asset prices. Because Mediator 2 is endowed with mean-variance preferences, it suffices to investigate the case in which

[6] We write $\mathbb{E}_{t-1}[p_t]$ for the conditional expectations $\mathbb{E}[p_t|\mathcal{F}_{t-1}]$ and $\mathbb{V}_{t-1}[p_t]$ for the conditional second moments $\mathbb{V}[p_t|\mathcal{F}_{t-1}]$.

the first two moments of his subjective probability distributions coincide with the first two moments of the true distributions. To this end we review the notion of a *perfect forecasting rule for first moments*, also referred to as an *unbiased forecasting rule*, and a *perfect forecasting rule for second moments* and provide a brief outline of conditions guaranteeing existence of such forecasting rules. For brevity we adopt the term *rational expectations* to describe the situation in which Mediator 2 is able to correctly predict the first two moments of the price process, whereas Mediator 1 as well as other market participants may have nonrational beliefs.

The main informational constraint for an investor to apply a perfect forecasting rule is the fact that market shares and expectations of other market participants are generally unobservable. Therefore, we can argue that it is highly unlikely that investors have rational expectations in real situations. However, as the rational expectations paradigm is still very popular in economics and finance, perfect forecasting rules is a concept that allows us to reconcile this traditional view with the recent behavioral approaches in these disciplines. The concept of accurateness adopted here is relaxed in Böhm and Wenzelburger (2002) and Wenzelburger (2006) to ϵ-perfect forecasting rules that allow for deviations of a prescribed magnitude ϵ. Wenzelburger (2006) also demonstrates that in terms of accurate forecasts, successful learning schemes should aim at estimating perfect forecasting rules.

Suppose for simplicity that the exogenous dividend process is known. Let date $t \in \mathbb{N}$ be arbitrary and $p_{t-1}^{(2)}$ denote the forecast for the *ex-dividend prices* p_t and $q_{t-1}^{(2)} = p_{t-1}^{(2)} + \mathbb{E}_{t-1}[d_t]$ be the forecast for the *cum-dividend prices* $q_t = p_t + d_t$, both made at date $t-1$. Using Eq. 6.21, the expected cum-dividend price conditional on information available at date $t-1$ is

$$\mathbb{E}_{t-1}[q_t] = A_t^{(1)} q_t^{(1)} + A_t^{(2)} q_t^{(2)} - A_t(x_m - \mathbb{E}_{t-1}[\xi_t]) + \mathbb{E}_{t-1}[d_t] \qquad \textbf{(6.24)}$$

The condition that Mediator 2's forecast error conditional on information available at date $t-1$ vanishes is

$$\mathbb{E}_{t-1}[q_t - q_{t-1}^{(2)}] = 0 \quad \mathbb{P} - \text{a.s.} \qquad \textbf{(6.25)}$$

By construction, $A_t^{(2)}$ is invertible. Inserting Eq. 6.24 into Eq. 6.25, one may therefore solve for the forecast $q_t^{(2)}$ to obtain

$$q_t^{(2)} = A_t^{(2)-1} \Big(p_{t-1}^{(2)} - A_t^{(1)} q_t^{(1)} + A_t(x_m - \mathbb{E}_{t-1}[\xi_t]) \Big)$$

Rearranging, one obtains a perfect forecasting rule for first moments

$$\begin{aligned} q_t^{(2)} &= \Psi_{1\star}^{(2)}\big(a_t, \mathbb{E}_{t-1}[d_t], \mathbb{E}_{t-1}[\xi_t], q_t^{(1)}, V_t^{(1)}, V_t^{(2)}, q_{t-1}^{(2)}\big) \qquad \textbf{(6.26)} \\ &:= R_f\, p_{t-1}^{(2)} - \frac{1}{a_t^{(2)}} V_t^{(2)} \Big(\Phi^{(1)}(a_t^{(1)}, q_t^{(1)}, V_t^{(1)}, p_{t-1}^{(2)}) + \mathbb{E}_{t-1}[\xi_t] - x_m \Big) \end{aligned}$$

The term in the parentheses of the right side of Eq. 6.26 is the expected excess demand of Mediator 1 and noise traders. By construction, the forecasting rule (Eq. 6.26) chooses period t's forecast $q_t^{(2)}$ such that the previous forecast $q_{t-1}^{(2)}$ is *unbiased* in the sense that Eq. 6.25 is satisfied. Thus Eq. 6.26 provides the best least-squares prediction for q_{t+1} conditional on information available at time t. Since each $A_t^{(2)}$, $t \in \mathbb{N}$ is invertible, Eq. 6.26 is well defined and thus generates correct conditional first moments for Mediator 2 along the whole actual price process. The important insight of the functional form (Eq. 6.26) is that Mediator 2 has to know the expected excess demand of all other market participants to obtain unbiased forecasts.[7]

The construction of *perfect forecasting rules for second moments* is based on the observation that the subjective covariance matrix $V_t^{(2)}$ in the expression for Eq. 6.26 has not been specified yet. The idea is analogous to the idea for perfect forecasting rules for first moments: Choose the matrix $V_t^{(2)}$ such that $V_{t-1}^{(2)}$ is the correct forecast for the true second moments $\mathbb{V}_{t-1}[q_t]$.

Note first that in the absence of noise traders, Eq. 6.22 implies $\mathbb{V}_{t-1}[p_t] \equiv 0$. By Assumption 6.2, Eq. 6.23 then reduces to

$$\mathbb{V}_{t-1}[q_t] = \mathbb{V}_{t-1}[d_t] \quad \text{for all} \ \ t \in \mathbb{N}$$

As a consequence, $V_t^{(2)}$ must be equal to $\mathbb{V}_t[d_{t+1}]$ in order to be correct. This surprising result holds even for nonlinear price laws as in Böhm and Chiarella (2005).

If noise traders are operating in the market, matters change. Recall that by Assumption 6.2 (2) ξ_t and d_t are uncorrelated, so that

$$\mathbb{V}_{t-1}[q_t] = A_t \mathbb{V}_{t-1}[\xi_t] A_t + \mathbb{V}_{t-1}[d_t] \tag{6.27}$$

Using Eq. 6.27, the condition that Mediator 2's forecast errors for second moments of period $t-1$ vanish takes the form

$$A_t \mathbb{V}_{t-1}[\xi_t] A_t + \mathbb{V}_{t-1}[d_t] - V_{t-1}^{(2)} = 0 \tag{6.28}$$

Since $A_t = \frac{1}{R_f}\left(a_t^{(1)} V_t^{(1)-1} + a_t^{(2)} V_t^{(2)-1}\right)^{-1}$, this implies that perfect forecasting rules for second moments are determined by symmetric, positive definite solutions $V_t^{(2)}$ to Eq. 6.28: $V_t^{(2)}$ has to be chosen such that the previously determined forecast $V_{t-1}^{(2)}$ becomes correct.

Equation 6.28 is a matrix polynomial equation. In the one-asset case $K = 1$, it reduces to a scalar quadratic equation that is straightforward to solve. The multi-asset case requires some linear algebra. The following existence result is found in Wenzelburger (2004, 2006). For the sake of simplicity assume that the dividend process is predictable, that is, $\mathbb{E}_{t-1}[d_t] = d_t$ and $\mathbb{V}_{t-1}[d_t] = 0$ for all $t \in \mathbb{N}$. The existence

[7] An unbiased forecasting rule for ex-dividend prices is obtained from Eq. 6.26 by setting $p_{t-1}^{(2)} = q_{t-1}^{(2)} - \mathbb{E}_{t-1}[d_t]$ as a forecast for p_t. Then Eq. 6.25 implies that all forecast errors on ex-dividend prices for Mediator 2 vanish in the mean, that is, $\mathbb{E}_{t-1}[p_t - p_{t-1}^{(2)}] = 0$ for all times t.

proof now consists of two steps. First, find a symmetric positive definite matrix Γ_t such that $\Gamma_t^{-1}\mathbb{V}_{t-1}[\xi_t]\Gamma_t^{-1} = V_{t-1}^{(2)}$. Second, if the matrix

$$\tfrac{1}{R_f}\Gamma_t - a_t^{(1)}V_t^{(1)-1} \tag{6.29}$$

is positive-definite, the desired second moment is found.

Proposition 6.2. *Under the hypotheses of Assumption 6.2, suppose that dividends are predictable and that each $\Lambda_{t-1} := \mathbb{V}_{t-1}[\xi_t]$, $t \in \mathbb{N}$ is positive-definite. Set*[8]

$$\Gamma_t = \sqrt{\Lambda_{t-1}}\sqrt{\left(\sqrt{\Lambda_{t-1}}\,V_{t-1}^{(2)}\,\sqrt{\Lambda_{t-1}}\right)^{-1}}\sqrt{\Lambda_{t-1}}, \quad t \in \mathbb{N}$$

Then the forecasting rule for second moments $\Psi_{2\star}^{(2)}$, given by

$$V_t^{(2)} = \Psi_{2\star}^{(2)}\left(\Lambda_{t-1}, a_t, V_t^{(1)}, V_{t-1}^{(2)}\right) := a_t^{(2)}\left(\tfrac{1}{R_f}\Gamma_t - a_t^{(1)}V_t^{(1)-1}\right)^{-1} \tag{6.30}$$

provides correct second moments of the price process at date t in the sense that Eq. 6.28 holds whenever Eq. 6.29 is positive-definite and $a_t^{(2)} > 0$.

When noise traders induce zero correlation between different assets such that $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi^2 I_K$, Γ_t reduces to $\Gamma_t = \sigma_\xi\sqrt{V_{t-1}^{(2)-1}}$. If Eq. 6.29 is not symmetric and positive-definite, $\Psi_{2\star}^{(2)}$ does not define a covariance matrix and forecasting rules that predict correct second moments for all times t do not exist.

The following two lemmas establish existence and uniqueness in two special cases. In the first one, all mediators agree on subjective second moments. In the second case, Mediator 1 believes in constant covariance matrices. Using Eq. 6.19, note that $a_t^{(1)} + a_t^{(2)} = \frac{1}{\alpha}$ for all $t \in \mathbb{N}$, and set $\overline{a} = \frac{1}{\alpha}$ for the *aggregate risk tolerance* of all consumers.

Lemma 6.1. *Under the hypotheses of Proposition 6.2, suppose $V_t^{(1)} \equiv V_t^{(2)}$ for all times t. Then the forecasting rule (Eq. 6.30) takes the form*

$$V_t^{(2)} = \left(R_f\overline{a}\right)\sqrt{\Lambda_{t-1}^{-1}}\sqrt{\left(\sqrt{\Lambda_{t-1}}\,V_{t-1}^{(2)}\,\sqrt{\Lambda_{t-1}}\right)}\sqrt{\Lambda_{t-1}^{-1}}$$

If, in addition, $\mathbb{V}_{t-1}[\xi_t] \equiv \Lambda$ is constant over time, the constant rule

$$V_t^{(2)} \equiv (R_f\overline{a})^2\Lambda^{-1} \quad \text{for all } t \in \mathbb{N}$$

is perfect for second moments as well.

[8]Recall the definition of a square root of a symmetric positive definite matrix. It is well known that any symmetric and positive definite $K \times K$ matrix B can be diagonalized so that $B = O^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_K)O$ for real eigenvalues $\lambda_1, \ldots, \lambda_K > 0$ and for some orthogonal matrix O, that is, $O^\top O = I_K$. The square root $\sqrt{B}$ of B is a symmetric positive definite matrix, defined by $\sqrt{B} := O^\top \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})O$, such that $B = \sqrt{B}\sqrt{B}$.

Lemma 6.2. *Under the hypotheses of Proposition 6.2, let* $\mathbb{V}_{t-1}[\xi_t] \equiv \sigma_\xi^2 I_K$ *with* $\sigma_\xi > 0$. *Assume that the following hypotheses are satisfied.*

1. *Mediator 1 uses constant covariance matrices* $V_t^{(1)} \equiv V^{(1)}$ *for all times t, where* $V^{(1)-1} = O^\top \operatorname{diag}(\lambda^{(11)}, \ldots, \lambda^{(1K)})\ O$ *for some orthogonal* $K \times K$ *matrix O. The eigenvalues satisfy*

$$\max\left\{ \lambda^{(1k)} \ :\ k = 1, \ldots, K \right\} < \frac{1}{4} \left(\frac{\sigma_\xi}{R_f \bar{a}} \right)^2$$

2. *The initial forecast in period* 0 *is* $V_0^{(2)} := O^\top \operatorname{diag}(\lambda_0^{(21)-1}, \ldots, \lambda_0^{(2K)-1}) O$*, where O is given in 1 and* $\lambda_0^{(21)}, \ldots, \lambda_0^{(2K)} > 0$.

If the initial eigenvalues $\lambda_0^{(21)}, \ldots, \lambda_0^{(2K)}$ *are suitably large, the forecasting rule (Eq. 6.30) is perfect for second moments whenever* $a_t^{(2)} > 0$.

Observe that perfect forecasting rules for second moments are not necessarily uniquely determined. Mathematically this is due to the natural existence of multiple solutions to matrix polynomial equations, see Gohberg, Lancaster, and Rodman (1982). Of course, this multiplicity arises already in the one-asset case.

6.3.3. Systematic and Nonsystematic Risk

In a mean-variance framework it is natural to evaluate investment decisions of boundedly rational investors by comparing the Sharpe ratios associated with their portfolios. It follows from the separation principle discussed in Section 6.2 that the higher the Sharpe ratio of a portfolio, the higher the expected utility of an investment. Any portfolio $x_t^{(i)}$ of an investor i with mean-variance preferences is, of course, efficient with respect to subjective beliefs $(q_t^{(i)}, V_t^{(i)})$. However, such a portfolio will not necessarily be efficient with respect to the true moments of the actual price process. Moreover, since investors in an environment with heterogeneous beliefs will most likely not hold a proportion of the market portfolio, it is *a priori* unclear which portfolio will attain the highest Sharpe ratio. A *modified market portfolio*, which in the sense of Markowitz is *efficient* with respect to the true first moments and which attains the highest possible Sharpe ratio, has been introduced in Wenzelburger (2004). Adopting this concept, the purpose of this section is to introduce a security market line for heterogeneous beliefs, and to adapt the CAPM concepts of *systematic* and *nonsystematic* risk to the model under consideration.

Assume from now on that all covariance matrices $\mathbb{V}_{t-1}[\xi_t]$, $t \in \mathbb{N}$ are positive-definite so that all $\mathbb{V}_t[q_{t+1}]$, $t \in \mathbb{N}$ are invertible and let $\pi_t := \mathbb{E}_t[q_{t+1}] - R_f p_t$ denote the vector of expected excess returns in period t. A *reference portfolio* of period t may then be defined by

$$x_t^{\text{ref}} := \mathbb{V}_t[q_{t+1}]^{-1} \pi_t \qquad \textbf{(6.31)}$$

The reference portfolio x_t^{ref} of period t is a "fictitious" portfolio because it is not necessarily held by a trader. It is a quantity that can be associated with any price process with non-degenerate second moments. From the discussion in Section 6.2.3 it is not hard to see that the reference portfolio x_t^{ref} is collinear to the market portfolio x_m if beliefs are homogeneous and noise trader are absent. It is therefore justified to refer to x_t^{ref} as the *modified market portfolio*, which accounts for the diversity in investors' beliefs.

Since the value of this reference portfolio at prices p_t is $\langle p_t, x_t^{\mathrm{ref}}\rangle$, its return is

$$R_{t+1}^{\mathrm{ref}} = \frac{\langle q_{t+1}, x_t^{\mathrm{ref}}\rangle}{\langle p_t, x_t^{\mathrm{ref}}\rangle} - 1$$

Its conditional variance $\mathbb{V}\mathrm{ar}_t[R_{t+1}^{\mathrm{ref}}]$ computes as

$$\mathbb{V}\mathrm{ar}_t[R_{t+1}^{\mathrm{ref}}] = \frac{\langle x_t^{\mathrm{ref}}, \mathbb{V}_t[q_{t+1}]x_t^{\mathrm{ref}}\rangle}{\langle p_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle^2} = \frac{\langle \pi_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle}{\langle p_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle^2} \tag{6.32}$$

and hence is proportional to the squared market price of risk $\langle \pi_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle$. This implies that the *(conditional) risk premium* of the reference portfolio satisfies

$$\mathbb{E}_t[R_{t+1}^{\mathrm{ref}}] - r_f = \langle \pi_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle^{\frac{1}{2}}\, \mathbb{V}\mathrm{ar}_t[R_{t+1}^{\mathrm{ref}}]^{\frac{1}{2}} \tag{6.33}$$

so that its risk-return characteristics lie on the capital-market line of period t.

Let $x_t^{(i)}$ denote the portfolio of risky assets held by mediator i in period t after investing $W_t^{(i)}$. Then the realized return $R_{t+1}^{(i)}$ on i's portfolio in period $t+1$ is

$$R_{t+1}^{(i)} = r_f + \tfrac{1}{W_t^{(i)}}\langle \pi_t, x_t^{(i)}\rangle \tag{6.34}$$

The conditional covariance $\mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}]$ between $R_{t+1}^{(i)}$ and the return of the reference portfolio x_t^{ref} is

$$\mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}] = \tfrac{1}{W_t^{(i)}}\langle x_t^{(i)}, \mathbb{V}_t[q_{t+1}]x_t^{\mathrm{ref}}\rangle \tag{6.35}$$

Inserting Eq. 6.31 into Eq. 6.35, we see that the risk premium associated with the portfolio $x_t^{(i)}$ becomes

$$\mathbb{E}_t[R_{t+1}^{(i)}] - r_f = \langle p_t, \mathbb{V}[q_{t+1}]^{-1}\pi_t\rangle\, \mathbb{C}\mathrm{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\mathrm{ref}}] \tag{6.36}$$

Thus the risk premium of mediator i's portfolio is proportional to the covariance between its return and the return of the reference portfolio. Combining Eq. 6.36 with Eqs. 6.32 and 6.33 yields the following theorem (see Wenzelburger, 2004).

Theorem 6.3. *Let i be an arbitrary mediator and assume that preferences and beliefs of all market participants are such that asset markets clear in each period $t \in \mathbb{N}$ at* *prices p_t. Moreover, assume that $\mathbb{V}_t[q_{t+1}]$ is positive definite for all $t \in \mathbb{N}$ such that the reference portfolio is well defined. Then for each $t \in \mathbb{N}$,*

$$\mathbb{E}_t[R_{t+1}^{(i)}] - r_f = \beta_t^{(i)} \left[\mathbb{E}_t[R_{t+1}^{\text{ref}}] - r_f\right]$$

where

$$\beta_t^{(i)} := \frac{\mathbb{C}\text{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\text{ref}}]}{\mathbb{V}\text{ar}_t[R_{t+1}^{\text{ref}}]}$$

Theorem 6.3 is a generalization of the *security market line* (e.g., see LeRoy and Werner, 2001, or Luenberger, 1998) to asset markets with heterogeneous beliefs. It states that the risk premium of a portfolio i depends linearly as its covariance with the reference portfolio; that is, its $\beta_t^{(i)}$. One important observation is that its proof required no assumption on the preferences of investors and, apart from the covariance structure, no assumption on the nature of the price process.

Theorem 6.3 provides further insights as to why the conditional beta $\beta_t^{(i)}$ of a portfolio $x_t^{(i)}$ is an important measure of risk. First, as in the static case, we may always choose a random variable $\varepsilon_{t+1}^{(i)}$ such that the random return of $x_t^{(i)}$ may be written as

$$R_{t+1}^{(i)} = r_f + \beta_t^{(i)}[R_{t+1}^{\text{ref}} - r_f] + \varepsilon_{t+1}^{(i)} \tag{6.37}$$

Taking expectations, Theorem 6.3 implies $\mathbb{E}_t[\varepsilon_{t+1}^{(i)}] = 0$. Taking the conditional covariance between R_{t+1}^{ref} and (6.37) and using the definition of the conditional beta, we see that $\mathbb{C}\text{ov}_t[\varepsilon_{t+1}^{(i)}, R_{t+1}^{\text{ref}}] = 0$. Thus, the variance of $R_{t+1}^{(i)}$ becomes the sum of two parts:

$$\mathbb{V}\text{ar}_t[R_{t+1}^{(i)}] = (\beta_t^{(i)})^2 \mathbb{V}\text{ar}_t[R_{t+1}^{\text{ref}}] + \mathbb{V}\text{ar}_t[\varepsilon_{t+1}^{(i)}]$$

The first part $(\beta_t^{(i)})^2\, \mathbb{V}\text{ar}_t[R_{t+1}^{\text{ref}}]$ may be termed *systematic risk*. This risk is associated with the market in period t as a whole. If $x_t^{(i)}$ consisted of a single asset, this risk could not be diversified away, because every asset with a nonzero beta contains this risk. The second part may be termed *nonsystematic* or *idiosyncratic risk*. This risk is uncorrelated with the market and could be reduced by diversification, namely, by choosing a combination of the risk-free asset and the reference portfolio.

Second, Theorem 6.3 states that the risk premium of a portfolio $x_t^{(i)}$ can only be higher than the risk premium of the reference portfolio x_t^{ref} at the expense of higher risk, that is, $\mathbb{V}\text{ar}_t[R_{t+1}^{(i)}] \geq \mathbb{V}\text{ar}_t[R_{t+1}^{\text{ref}}]$. Using the well-known inequality

$$\mathbb{C}\text{ov}_t[R_{t+1}^{(i)}, R_{t+1}^{\text{ref}}] \leq \sqrt{\mathbb{V}\text{ar}_t[R_{t+1}^{(i)}]}\,\sqrt{\mathbb{V}\text{ar}_t[R_{t+1}^{\text{ref}}]}$$

Theorem 6.3 implies that the *Sharpe ratio* conditional on information at date t of a portfolio $x_t^{(i)}$ is bounded from above by

$$\frac{\mathbb{E}_t[R_{t+1}^{(i)}] - r_f}{\sqrt{\mathbb{V}\mathrm{ar}_t[R_{t+1}^{(i)}]}} \leq \frac{\mathbb{E}_t[R_{t+1}^{\mathrm{ref}}] - r_f}{\sqrt{\mathbb{V}\mathrm{ar}_t[R_{t+1}^{\mathrm{ref}}]}} = \langle \pi_t, \mathbb{V}_t[q_{t+1}]^{-1}\pi_t\rangle^{\frac{1}{2}} \qquad \textbf{(6.38)}$$

The upper bounds of the Sharpe ratios (Eq. 6.38) are the market prices of risk that are generated by the endogenous price process. Since any portfolio of risky assets $\tilde{x}_t = \lambda\, x_t$ with $\lambda > 0$ will have the same Sharpe ratio as $x_t \in \mathbb{R}^K$, only the mix of a portfolio has an influence on its Sharpe ratio.[9] Hence, in a world of heterogeneous investors, perfect forecasting rules for first and second moments, if they exist, allow us to form portfolios that are efficient in the the original sense of Markowitz and thus attain the highest possible conditional Sharpe ratios.

6.3.4. Selecting Mediators

Let us next review a probabilistic framework that allows us to model the decision making of consumers who are assumed to be boundedly rational in the sense of Simon (1982). Adopting a standard assumption in the social interaction literature (e.g., see Blume, 1993; Brock and Durlauf, 2001; Horst and Scheinkman, 2006) assume that mediators are randomly selected, where the choice probabilities depend on a certain performance measure. The general idea is that this performance measure and thus a mediator's market shares depend on the empirical distribution of observable quantities such as asset prices, wealth positions, and returns. There are numerous ways to specify such performance measures. Since the work of Brock and Hommes (1997a,b, 1998), most approaches build on the fact that consumers are generally unable to make accurate forecasts about the future and therefore have to rely on the *empirical performance* of an investment strategy. In this section, we follow the intuition that consumers take into account realized returns of mediators in deciding on where to invest their endowments.

Let $X_t \in \mathbb{R}^d$ denote the vector of observables at time t and ϱ_t its empirical distribution, that is,

$$\varrho_t := \frac{1}{t}\sum_{s=0}^{t-1}\delta_{X_s} \quad \text{with} \quad \varrho_t(f) := \int f\,\mathrm{d}\varrho_t = \frac{1}{t}\sum_{s=0}^{t-1} f(X_s) \qquad \textbf{(6.39)}$$

where δ_x denotes the Dirac measure that puts all mass on x, and $f : \mathbb{R}^d \to \mathbb{R}$ is a bounded, measurable map. A consumer's willingness to invest through a specific mediator may now be linked to a performance measure of a mediator's investment strategy as follows.

[9]Note that the Sharpe ratio is only a meaningful concept for multiple risky assets, $K > 1$.

Definition 6.1. Let $f^{(l)} : \mathbb{R}^d \to \mathbb{R}$, $l = 1, 2, \ldots, L$ be a fixed list of bounded, measurable functions. For any $t \in \mathbb{N}$, denote by

$$z_t := \left(\varrho_t(f^{(1)}), \ldots, \varrho_t(f^{(L)})\right)$$

the list of *empirical averages* that is relevant in period t. A *performance measure* for mediator i is a Lipschitz continuous function $\Pi^{(i)} : \mathbb{R}^L \to \mathbb{R}$ such that mediator i's *performance* in period t is characterized by $\pi_t^{(i)} = \Pi^{(i)}(z_t)$.

In this setting the performance of mediator i at time t depends on the entire history X_s, $s < t$ of the process excluding the realization X_t. The list of performance measures $\pi_t = (\pi_t^{(1)}, \pi_t^{(2)})$ is also referred to as a *fitness measure*, which consumers observe. Since the functions $f^{(l)}$, $l = 1, \ldots, L$ are fixed and the same for all mediators, a performance measure may be interpreted as a real-valued mapping from the set of probability distributions on $\mathbb{R}^d$ so that $\pi_t^{(i)} \equiv \Pi^{(i)}(\varrho_t)$. In a mean-variance frame work, the following two examples are natural: empirical averages and empirical Sharpe ratios.

EXAMPLE 6.1

Suppose that the performance of mediator i is measured by empirical averages of historically realized returns $\{R_s^{(i)}\}_{s=0}^t$ on investment as given by Eq. 6.34. Having invested the amount $W_{t-1}^{(i)}$, her return from selling the portfolio $x_{t-1}^{(i)}$ as given in Eq. 6.17 in period t is

$$R_t^{(i)} = f^{(i)}(X_t) := r_f + \frac{1}{\alpha\, e}\langle [q_t - R_f p_{t-1}], V_{t-1}^{(i)-1}[q_t^{(i)} - R_f p_{t-1}]\rangle$$

Here $X_t \in \mathbb{R}^d$ is a suitably defined vector, $f^{(i)} : \mathbb{R}^d \to \mathbb{R}$ is a suitably defined continuous function,[10] and $L = 2$. Based on empirical averages of returns, the performance of mediator i in period t is measured by

$$\pi_t^{(i)} := \varrho_t(f^{(i)}) = \frac{1}{t}\sum_{s=0}^{t-1} R_s^{(i)}$$

where $\Pi^{(i)}$ is the identity mapping.

EXAMPLE 6.2

For each $i = 1, 2$, set $f^{(i+2)} := f^{(i)2}$ with $f^{(i)}$ as in Example 6.1 to get

$$\varrho_t(f^{(i+2)}) = \tfrac{1}{t}\sum_{s=0}^{t-1} R_s^{(i)2}$$

[10]A priori, there is no reason to assume that returns can be represented by bounded functions. Assuming that consumers do not trust in the validity of extraordinarily high or low returns, one could modify the performance measure by replacing $f^{(i)}(X)$ with $\tilde{f}^{(i)}(X) := \max\{R_{\min}, \min\{f^{(i)}(X), R_{\max}\}\}$.

with $L = 4$. Based on *empirical Sharpe ratios*, the performance of mediator i in period t is measured by

$$\pi_t^{(i)} = \Pi^{(i)}(z_t) := \frac{\varrho_t(f^{(i)}) - r_f}{\left(\varrho_t(f^{(i+2)}) - \varrho_t(f^{(i)})^2\right)^{\frac{1}{2}}}$$

noting that the denumerator of $\pi_t^{(i)}$ is the empirical standard deviation of mediator i's realized returns.

The central assumption in modeling consumers' choice is that their choice probabilities depend on the current performance measures π_t or, equivalently, on the list of empirical averages $z_t \in \mathbb{R}^L$. Specifically, let consumers act conditionally independent of each other given z_t so that an individual consumer at time t chooses Mediator 1 with probability $F(z_t)$ and Mediator 2 with probability $1 - F(z_t)$, where $F : \mathbb{R}^L \to \mathbb{R}$ is a uniformly continuous "choice function." In the limit of an infinite number of consumers, the law of large numbers implies that the fraction of consumers who in fact choose Mediator 1 is $F(z_t)$. Thus Mediator 1's market share at time t is deterministic and given by

$$\eta_t = F(z_t), \quad \text{with} \quad z_t = \left(\varrho_t(f^{(1)}), \ldots, \varrho_t(f^{(L)})\right) \tag{6.40}$$

whereas Mediator 2's market share is $1 - F(z_t)$. A convenient parameterization of a choice function is provided by a LOGIT model (e.g., see Anderson, de Palma, and Thisse, 1992), as the following example illustrates.

EXAMPLE 6.3

Suppose the (modified) logit function

$$\varphi([\pi_t^{(2)} - \pi_t^{(1)}], \beta) := \frac{\overline{\eta} - \underline{\eta}}{\exp(\beta[\pi_t^{(2)} - \pi_t^{(1)}]) + 1} + \underline{\eta} \tag{6.41}$$

describes the probability with which a consumer chooses Mediator 1. The values $0 \leq \underline{\eta} \leq \overline{\eta} \leq 1$ are upper and lower bounds for the choice probabilities, respectively. The *intensity of choice* $\beta > 0$ specifies the strength of the impact of the mediators' performance on the choice decision. Based on empirical averages, a choice function (Eq. 6.40) for Mediator 1 is now defined by setting

$$\eta_t = F\big(z_t\big) := \varphi\big(\Delta\Pi(z_t), \beta\big), \quad z_t \in \mathbb{R}^L \tag{6.42}$$

where $\Delta\Pi(z_t) := \Pi^{(2)}(z_t) - \Pi^{(1)}(z_t)$.

6.3.5. Dynamic Stability with Rational Expectations

Let us discuss the evolution of asset prices when, under the hypotheses of Section 6.3.2, Mediator 2 has rational expectations in the sense that the first moments of cum-dividend prices are correctly predicted for all times t and second moments whenever possible. Let $\Psi_{1\star}^{(2)}$ be the unbiased forecasting rule defined in Eq. 6.26 and $\Psi_{2\star}^{(2)}$ be some forecasting rule for second moments that coincides with Eq. 6.30 whenever Eq. 6.30 is well defined. Let Mediator 1 use some adaptive forecasting rules $\Psi_1^{(1)}$ and $\Psi_2^{(1)}$ to determine her beliefs regarding the evolution of future asset prices. Inserting all forecasting rules into the temporary equilibrium map (Eq. 6.20) and assuming the dividend process to be predictable, we obtain the system of stochastic difference equations:

$$\begin{cases} q_t &= q_{t-1}^{(2)} + A_t(\xi_t - \mathbb{E}_{t-1}[\xi_t]) \\ q_t^{(1)} &= \Psi_1^{(1)}(q_{t-1}, \ldots, q_{t-N}) \\ V_t^{(1)} &= \Psi_2^{(1)}(q_{t-1}, \ldots, q_{t-N}) \\ q_t^{(2)} &= \Psi_{1\star}^{(2)}\left(a_t, d_t, \mathbb{E}_{t-1}[\xi_t], q_t^{(1)}, V_t^{(1)}, V_t^{(2)}, q_{t-1}^{(2)}\right) \\ V_t^{(2)} &= \Psi_{2\star}^{(2)}\left(a_t, \mathbb{V}_{t-1}[\xi_t], V_t^{(1)}, V_{t-1}^{(2)}\right) \end{cases} \tag{6.43}$$

Recalling that the risk-adjusted market shares are $a_t = (\frac{\eta_t}{\alpha}, \frac{1-\eta_t}{\alpha})$, Eqs. 6.43 together with some choice function Eq. 6.40 define a time-one map of a dynamical system in a random environment that describes the evolution of prices and forecasts. The map (Eq. 6.43) will, in general, be nonlinear. However, as the following discussion will show, this nonlinearity is of a very specific form.

Suppose that Mediator 1 is a chartist who uses the simple *technical trading rule*

$$q_t^{(1)} = \Psi_1^{(1)}(q_{t-1}, \ldots, q_{t-N}) := \sum_{n=1}^{N} D^{(n)}\, q_{t-n} \tag{6.44}$$

as a forecasting rule, where $D^{(n)}$, $n = 1, \ldots, N$ are $K \times K$ matrices that describe the anticipated impact of the past N asset prices. Let her subjective covariance matrix be constant over time, that is, $V_t^{(1)} \equiv V^{(1)}$. Then, under the hypotheses of Lemma 6.2, Mediator 2 may have correct second moments that are constant over time. If $X_t = (q_t^{(2)}, q_{t-1}^{(2)}, q_t, \ldots, q_{t-N-1}) \in \mathbb{R}_+^d$ with $d = K(N+4)$ denotes the vector consisting of the last two price forecasts of Mediator 2 and past N realized cum-dividend prices, Eq. 6.43 takes the form

$$X_t = \mathbf{A}(a_t^{(2)}, V_t^{(2)})X_{t-1} + \mathbf{B}(a_t^{(2)}, \xi_t, d_t, V_t^{(2)}) \tag{6.45}$$

with a coefficient block matrix $\mathbf{A}(a_t^{(2)}, V_t)$ and a vector $\mathbf{B}(a_t^{(2)}, \xi_t, V_t)$. Both are random and defined as follows. Setting

$$\mathbf{A}_t^{(0)} = \mathbf{A}^{(0)}(a_t^{(2)}, V_t^{(2)}) := R_f \left[I_K + \frac{a - a_t^{(2)}}{a_t^{(2)}} V_t^{(2)} V^{(1)-1} \right]$$

$$\mathbf{A}_t^{(n)} = \mathbf{A}^{(n)}(a_t^{(2)}, V_t^{(2)}) := -\frac{(a - a_t^{(2)})}{a_t^{(2)}} V_t^{(2)} V^{(1)-1} D^{(n)}, \quad n = 1, \ldots, N$$

yields the $d \times d$ matrix

$$\mathbf{A}(a_t^{(2)}, V_t^{(2)}) := \begin{pmatrix} \mathbf{A}_t^{(0)} & 0 & \mathbf{A}_t^{(1)} & \cdots & \mathbf{A}_t^{(N)} & 0 & 0 \\ I_K & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ I_K & 0 & \ddots & & & & \vdots \\ 0 & 0 & I_K & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & I_K & 0 \end{pmatrix} \tag{6.46}$$

and the vector

$$\mathbf{B}(a_t^{(2)}, \xi_t, d_t, V_t^{(2)}) := \begin{pmatrix} \frac{1}{a_t^{(2)}} V_t^{(2)} \Big(x_m - \mathbb{E}_{t-1}[\xi_t] \Big) - \mathbf{A}^{(2)}(a_t^{(2)}, V_t^{(2)}) d_t \\ 0 \\ A_t(\xi_t - \mathbb{E}_{t-1}[\xi_t]) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^d$$

The important observation now is that the system of stochastic difference equations (Eq. 6.45) is (affine-) linear if the market shares and the subjective covariance matrices are constant over time such that $a_t^{(2)} = a$ and $V_t^{(2)} = V$ for all times t. In this case, the stochastic process generated by Eq. 6.45 is Markovian if the exogenous noise is Markovian and represents an asset price dynamic with market shares frozen at level a. The long-run behavior of such sequences is well understood. The process is asymptotically stable under the condition that all eigenvalues of $\mathbf{A}(a, V)$ lie within the unit circle (e.g., see Brandt, 1986, and Arnold, 1998, Corollary 5.6.6). For the nonlinear system (Eq. 6.45) with varying market shares, a slightly stronger condition is needed.

Assumption 6.3.

1. *The eigenvalues of all the matrices* $\mathbf{A}(a, V)$ *with* $(a, V) \in [\underline{\mathbf{a}}, \overline{\mathbf{a}}] \times \mathcal{K}$ *lie uniformly within the unit circle.*
2. *The map*

$$\mathbf{A} : [\underline{\mathbf{a}}, \overline{\mathbf{a}}] \times \mathcal{K} \to [\underline{\mathbf{a}}, \overline{\mathbf{a}}] \times \mathcal{K}, \quad (a, V) \mapsto \mathbf{A}(a, V)$$

is Lipschitz continuous.

3. *The map* $\mathbf{B}(\cdot)$ *is bounded and Lipschitz continuous uniformly in* ξ *and* d*, that is, there exists a constant* c_B *independently of* ξ *and* d *such that*

$$\sup_{\xi} \|\mathbf{B}(a, \xi, d, V) - \mathbf{B}(a', \xi, d, V')\| \leq c_B \big(|a - a'| + \|V - V'\| \big)$$

for all $a, a' \in [\underline{\mathbf{a}}, \overline{\mathbf{a}}]$ *and all* $V, V' \in \mathcal{K}$*.*

Assumption 6.3 (1) essentially states that all eigenvalues of each coefficient matrix (Eq. 6.46) lie uniformly within the unit circle. The other two assumptions are continuity conditions. The following result, given as Proposition 3.2 in Horst and Wenzelburger (2008), demonstrates that Assumption 6.3 guarantees boundedness of any sequence $X = \{X_t\}_{t\in\mathbb{N}}$ generated by Eq. 6.45.

Proposition 6.3. *Under Assumption 6.3, the sequence $\{X_t\}_{t\in\mathbb{N}}$ is almost surely bounded. Specifically, for any initial state x there exists a constant M_x such that*

$$\mathrm{Prob}_x\left[\sup_t \|X_t\| \le M_x\right] = 1$$

Here Prob_x *denotes the probability measure on the canonical path space induced by the process* $\{X_t\}_{t\in\mathbb{N}}$ *with initial state x.*

Summarizing, Eq. 6.45 describes scenarios with a stable price process for any coefficient matrix **A** that satisfies Assumption 6.3. Note that the only crucial assumption here is the linearity of forecasting rules for first moments. As soon as the coefficient matrix **A** has eigenvalues outside the unit disk, the system (Eq. 6.45) may become unstable. Under the particular case under consideration, it follows immediately from Böhm and Chiarella (2005, Thm. 3.2) that instability occurs for $R_f > 1$ and $a^{(2)}$ sufficiently close to unity. Hence market shares of the nonrational Mediator 1 have to be sufficiently high to obtain a stable price process with rational expectations for Mediator 2. A nonlinear stochastic scenario in which consumers switch between mediators will be discussed in Section 6.5.

6.4. MULTIPERIOD PLANNING HORIZONS

Investors in financial markets will typically have different planning horizons. Investors with long planning horizons are likely to invest more wealth into risky assets than those with short planning horizons. Institutional investors, for example, will pursue long-term strategies rather than trying to follow a momentary trend. Hence, the length of a planning horizon should have a significant influence on an investor's risk-taking behavior and thus their portfolio decisions. The traditional CAPM was first generalized by Stapleton and Subrahmanyam (1978) to a static model in which investors face a multiperiod planning horizon. As discussed in the introduction, several extensions to setups with infinitely lived agents have since been developed (e.g., see Magill and Quinzii, 2000, or Angeletos and Calvet, 2005, 2006). These models have in common that beliefs of all agents are homogeneous and rational. The behavior of agents is described by an essentially static one-shot optimization problem. All agents have the *same* infinite planning horizon, which makes it difficult to describe empirically observed trading activities (see Judd, Kübler, and Schmedders, 2003). The impact of *distinct* multiperiod planning horizons on the dynamics of asset prices, asset returns, and portfolio holdings has been investigated in Hillebrand and Wenzelburger (2006a). Their model is a natural extension of the approach presented in Section 6.2.4. It allows describing scenarios in which investors update subjective beliefs repeatedly and revise previously made portfolio plans accordingly. Hence, an investigation of how *distinct* planning horizons with heterogeneous beliefs affect individual portfolio decisions and asset prices is possible.

6.4.1. Overlapping Cohorts of Investors

To facilitate the exposition, we abstract from social interactions between consumers. As in Section 6.3, these can be incorporated by replacing the risk tolerances with risk-adjusted market shares. Further details are found in Hillebrand and Wenzelburger (2006b). Throughout this section, assume that the set of investors is composed of $J + 1$ different cohorts or generations.[11] In each trading period $t \in \mathbb{N}$, a new young cohort enters the market and trades for $J + 1$ consecutive periods before its members exit the market to consume terminal wealth in period $t + J$. Each cohort is identified by its index $j = 0, 1, \ldots, J$ which determines the number of remaining periods in the market. In particular, $j = J$ refers to the young and $j = 0$ to the old cohort. Each cohort j consists of two types of investors characterized by risk preferences and subjective beliefs regarding the future evolution of the market. A single investor in an arbitrary period is thus identified by the pair (i, j) describing her type $i \in \{1, 2\}$ and her cohort $j \in \{0, 1, \ldots, J\}$.

There is a single consumption good in the economy that serves as numéraire for all prices and payments. At the beginning of each period, any young investor (i, J) receives an initial endowment of $e^{(i)} > 0$ units of the consumption good that for the sake of simplicity, are assumed to be constant. Investors (i, j) with $j < J$ receive no additional endowments. Assuming that the consumption good cannot be stored by consumers, each investor faces the problem of transferring wealth from the first to the last period of life in which she consumes the proceeds of her investments. The investment opportunities are the same as in Section 6.2.4 and consist of K retradeable risky assets and a risk-free bond that pays a constant return $R_f > 0$ per unit. For simplicity, we abstract from dividend payments and assume that the only source of noise in the markets comes from noise traders, as characterized in Assumption 6.2 (2).

The portfolio choice problem of an investor (i, j) in an arbitrary period t is the following: At the beginning of each period t any investor forms beliefs regarding the future prices that are relevant for his planning horizon j. These beliefs are given by a subjective joint probability distribution for the random variables $p_{t+1}, \ldots, p_{t+j}$. Given his beliefs, the investor's portfolio decision will depend on current prices as well as on his wealth position in period t. As in the one-period case, it is assumed that the portfolio problem in period t is solved *prior* to trading, that is, before the actual price p_t and the noise traders' demand ξ_t has been observed. Current prices will therefore enter the decision problem as a parameter $p \in \mathbb{R}^K$. Let $y_{t-1}^{(i,j+1)} \in \mathbb{R}$ and $x_{t-1}^{(i,j+1)} \in \mathbb{R}^K$ denote the investment in the risk-free asset $k = 0$ and in risky assets $k = 1, \ldots, K$, respectively. Whereas each young investor's wealth is equal to his initial endowment $e^{(i)}$, the wealth $w_t^{(ij)}$ of any

[11] The reader may think of a multiperiod OLG model. As pointed out earlier, this is not essential in the present context.

nonyoung investor (i, j) with $j < J$ at time t is determined by the value of her previous portfolio holding $(y_{t-1}^{(i,j+1)}, x_{t-1}^{(i,j+1)})$ at current prices of period t. Thus

$$w_t^{(ij)} = \begin{cases} e^{(i)} & \text{for } j = J \\ R_f y_{t-1}^{(i,j+1)} + \langle p_t, x_{t-1}^{(i,j+1)} \rangle & \text{for } j = 0, \ldots, J-1 \end{cases} \tag{6.47}$$

To obtain explicit demand functions, the following standard assumptions regarding investors' preferences and beliefs are made (e.g., see Stapleton and Subrahmanyam, 1978).

Assumption 6.4. *Preferences and beliefs of investors are characterized by the following assumptions:*

1. *Each investor of type i has preferences for terminal wealth described by a CARA utility function*

$$u^{(i)}(w) := -\exp\left(-\tfrac{w}{a^{(i)}}\right), \quad w \in \mathbb{R}$$

   where $a^{(i)} > 0$ denotes risk tolerance.
2. *The subjective beliefs of investor (i, j) at time t regarding prices $p_{t+1}, \ldots, p_{t+j}$ are given by a normal distribution on $\mathbb{R}^{Kj}$, described by the first two moments*

$$\mu_t^{(ij)} := \begin{pmatrix} \mu_{t,t+1}^{(i)} \\ \vdots \\ \mu_{t,t+j}^{(i)} \end{pmatrix} \in \mathbb{R}^{Kj}, \qquad \Sigma_t^{(ij)} := \begin{bmatrix} \Sigma_{t,11}^{(i)} & \ldots & \Sigma_{t,1j}^{(i)} \\ \vdots & \ddots & \vdots \\ \Sigma_{t,j1}^{(i)} & \ldots & \Sigma_{t,jj}^{(i)} \end{bmatrix} \in \mathcal{M}_{Kj} \tag{6.48}$$

   and subjective covariance matrices $\Sigma_t^{(ij)} \in \mathcal{M}_{Kj}$ for future prices. Here, $\mu_{t,t+s}^{(ij)} := \mathbb{E}_t^{(ij)}[p_{t+s}]$ denotes investor (i, j)'s subjective mean value for price vectors p_{t+s}, $s = 1, \ldots, j$ conditional on information available at time t. The matrix

$$\Sigma_{t,ss'}^{(ij)} := \mathbb{E}_t^{(ij)}\left[\left(p_{t+s} - \mathbb{E}_t^{(ij)}[p_{t+s}]\right)\left(p_{t+s'} - \mathbb{E}_t^{(ij)}[p_{t+s'}]\right)^\top\right]$$

   denotes investor (i, j)'s subjective conditional covariance matrix between the two price vectors p_{t+s} and $p_{t+s'}$, where $s, s' = 1, \ldots, j$.

Investors (i, j) in Assumption 6.4 are essentially characterized by their risk aversion and their subjective beliefs, parameterized by subjective means and subjective second moments. For simplicity, any nonyoung investor (i, j) with a planning horizon $j < J$ is assumed to hold the same expectations for prices $p_{t+1}, \ldots, p_{t+j}$ as the young investor (i, J). This means that her beliefs are given by the marginal distributions of the respective young investor (i, J), which is known to be normal. Economically, this assumption may be justified by presuming that all investors of type i employ the same financial mediator. Notice that all subjective moments $(\mu_t^{(ij)}, \Sigma_t^{(ij)})$ are based on observables up to period $t - 1$. Mathematically this implies that they are $\mathcal{F}_{t-1}$ measurable.

Assuming that each investor (i, j) maximizes subjectively expected utility of future terminal wealth using self-financing portfolio strategies, Hillebrand and Wenzelburger (2006a) show that Assumption 6.4 is sufficient to obtain explicit demand functions for risky assets.

Theorem 6.4. *Let Assumption 6.4 be satisfied. Then investor (i, j)'s asset demand function for risky assets, given her beliefs $(\mu_t^{(ij)}, \Sigma_t^{(ij)}) \in \mathbb{R}^{Kj} \times \mathcal{M}_{Kj}$ in period t, takes the form:*

$$\Phi^{(ij)}(p, \mu_t^{(ij)}, \Sigma_t^{(ij)}) := \tfrac{a^{(i)}}{R_f^j} \Pi_j^\top \Sigma_t^{(ij)-1} \left(\mu_t^{(ij)} - \Pi_j p \right), \quad p \in \mathbb{R}^K \qquad \textbf{(6.49)}$$

where $\Pi_j := \left[R_f I_K, \ldots, R_f^j I_K \right]^\top \in \mathbb{R}^{Kj \times K}$ *and* $j = 1, \ldots, J$.

Note that each investor is allowed to update beliefs and reoptimize previously planned portfolio decisions throughout her entire life. As in the one-period case, the demand for risky assets (Eq. 6.49) is independent of the investor's initial wealth (Eq. 6.47) of period t. For a two-period planning horizon $j = 2$ and beliefs $(\mu_t^{(i)}, \Sigma_t^{(i)}) \in \mathbb{R}^{2K} \times \mathcal{M}_{2K}$, the demand function (Eq. 6.49) is

$$\Phi^{(i2)}(p, \mu_t^{(i)}, \Sigma_t^{(i)}) = \tfrac{a^{(i)}}{R_f^2} \left[R_f I_K, R_f^2 I_K \right] \begin{bmatrix} \Sigma_{t,11}^{(i)} & \Sigma_{t,12}^{(i)} \\ \Sigma_{t,21}^{(i)} & \Sigma_{t,22}^{(i)} \end{bmatrix}^{-1} \begin{pmatrix} \mu_{t,t+1}^{(i)} - R_f p \\ \mu_{t,t+2}^{(i)} - R_f^2 p \end{pmatrix}$$

If investors assume future prices to be uncorrelated over time; that is, $\Sigma_{t,ss'}^{(i)} = 0$ for all $s \neq s'$, then the demand function is

$$\Phi^{(ij)}(p, \mu_t^{(i)}, \Sigma_t^{(i)}) = a^{(i)} \sum_{s=1}^{j} \tfrac{1}{R_f^{j-s}} \Sigma_{t,ss}^{(i)-1} (\mu_{t,t+s}^{(i)} - R_f^s p) \qquad \textbf{(6.50)}$$

In this case, investor (i, j)'s asset demand function in period t is composed of j mutual funds

$$\Sigma_{t,ss}^{(i)-1} (\mu_{t,t+s}^{(i)} - R_f^s p_t), \quad s = 1, \ldots, j$$

which are discounted by the factors $\frac{1}{R_f^{j-s}}$, respectively. The number of funds depends on the length of the planning horizon j. Investors with the same beliefs will use the same funds, but the total amount invested depends on her risk tolerance. Hence, for uncorrelated beliefs, Theorem 6.4 is a *multifund separation theorem*. For a one-period planning horizon $j = 1$, the demand function (Eq. 6.13) is retained. Setting

$$B_t^{(ij)} = \left[B_{t,1}^{(ij)}, \ldots, B_{t,j}^{(ij)} \right] := \Pi_j^\top \Sigma_t^{(ij)-1} \in \mathbb{R}^{K \times Kj} \qquad \textbf{(6.51)}$$

with $B_{t,s}^{(ij)} \in \mathbb{R}^{K \times K}$ and

$$C_t^{(ij)} := \Pi_j^\top \Sigma_t^{(ij)-1} \Pi_j \in \mathbb{R}^{K \times K} \qquad \textbf{(6.52)}$$

investor (i, j)'s demand function for risky assets at time t takes the more convenient form

$$\Phi^{(ij)}(p, \mu_t^{(ij)}, \Sigma_t^{(ij)}) = \tfrac{a^{(i)}}{R_f^j}\Big(\sum_{s=1}^{j} B_{t,s}^{(ij)}\mu_{t,t+s}^{(i)} - C_t^{(ij)} p\Big), \quad p \in \mathbb{R}^K \tag{6.53}$$

6.4.2. Temporary Equilibria

To determine market-clearing prices, let $x_m \in \mathbb{R}_{++}^K$ denote the total stock of risky assets. Market clearing in period t requires the existence of a price vector $p_t \in \mathbb{R}^K$ such that aggregate demand equals the total stock of risky assets. Given the individual demand functions (Eq. 6.53) for risky assets and the quantity ξ_t demanded by noise traders, the market-clearing condition of period t reads

$$\sum_{i=1}^{2}\sum_{j=1}^{J}\Phi^{(ij)}(p, \mu_t^{(ij)}, \Sigma_t^{(ij)}) + \xi_t = x_m \tag{6.54}$$

Solving Eq. 6.53 for p yields the following *temporary equilibrium map* G, which determines market-clearing prices at time t from the list of subjective beliefs $(\mu_t^{(iJ)}, \Sigma_t^{(iJ)})$ and the noise traders' demand ξ_t as

$$p_t = G\Big(\xi_t, \Big(\mu_t^{(iJ)}, \Sigma_t^{(iJ)}\Big)_{i=1}^{2}\Big) := \sum_{i=1}^{2}\sum_{j=1}^{J} A_t^{(ij)}\mu_{t,t+j}^{(i)} - A_t(x_m - \xi_t) \tag{6.55}$$

where[12]

$$A_t := \Big[\sum_{i=1}^{2}\sum_{j=1}^{J}\tfrac{a^{(i)}}{R_f^j} C_t^{(ij)}\Big]^{-1} \quad \text{and} \quad A_t^{(ij)} := \tfrac{a^{(i)}}{R_f^j} A_t \sum_{n=j}^{J} B_{t,j}^{(in)} \tag{6.56}$$

The temporary equilibrium map (Eq. 6.55) defines an *economic law* in the sense of Böhm and Wenzelburger (2002) for a multiperiod version of the CAPM, which determines market-clearing prices in each trading period as a function of agents' expectations for future prices. The map G is of the *cobweb type* since it contains essentially price forecasts as arguments. Since these expectations refer to future periods $t+1, \ldots, t+J$, the law contains an *expectational lead* of length J. Notice that all coefficient matrices $A_t^{(ij)}$ and A_t are $\mathcal{F}_{t-1}$ measurable such that the uncertainty of the price p_t rests solely with the noise trader demand ξ_t.

Apart from the timing of subjective beliefs, the temporary equilibrium map with a multiperiod planning horizon is structurally quite similar to the case with one-period planning horizons. Here, heterogeneity consists of possibly diverse beliefs *as well as*

[12]Since all Π_j have rank K, all $C_t^{(ij)} = \Pi_j^\top \Sigma_t^{(ij)-1}\Pi_j$ are positive definite and hence invertible. Since the sum of positive-definite matrices is again positive definite, A_t and thus all $A_t^{(ij)}$ are well defined.

different planning horizons of investors. In Section 6.5.4, we will provide evidence that interesting implications arise from multiperiod planning horizons.

6.4.3. Perfect Forecasting Rules

As in the one-period planning case, stochastic difference equations describing the evolution of asset prices are now obtained by specifying the forecasting rules according to which investors update their beliefs. Following the reasoning of Section 6.3.2, we review the case in which at least one type of investor, say of Type 2, is able to correctly anticipate future asset prices. Since by Assumption 6.4 subjective beliefs are characterized by their corresponding first two moments, the analysis is restricted to the case in which the conditional mean values and the conditional covariance matrices of the price process induced by Eq. 6.55 are correctly predicted. To this end, the notion of a perfect forecasting rule introduced in Section 6.3.2 is generalized to the multiperiod case. For brevity we again adopt the term *rational expectations* to describe the situation in which Type-2 investors are able to correctly predict the first two moments of the price process, whereas other market participants may have erroneous beliefs.

Perfect Forecasting Rules for First Moments

Following Wenzelburger (2006) we assume that investors of Type 2 use a *no-updating* forecasting rule. The key characteristics of such a forecasting rule is that in any period t, the first $J-1$ forecasts will not be updated, so that

$$\mu^{(2)}_{t,t+j} = \mu^{(2)}_{t-1,t+j}, \quad j = 1, \ldots, J-1 \tag{6.57}$$

The intuition of no-updating forecasting rules is as follows: Let $\mathbb{E}_{t-1}[\cdot]$ denote the expectations operator conditional on information available at date $t-1$. Since the coefficient matrices $A_t^{(ij)}$ and A_t in the price law (Eq. 6.55) are based on information known at date $t-1$, the conditional expectations of p_t is

$$\mathbb{E}_{t-1}[p_t] = \sum_{i=1}^{2} \sum_{j=1}^{J} A_t^{(ij)} \mu^{(i)}_{t,t+j} - A_t\big(x_m - \mathbb{E}_{t-1}[\xi_t]\big)$$

As in the one-period case, the idea now is to choose the most recent forecast $\mu^{(2)}_{t,t+J}$ such that

$$\mathbb{E}_{t-1}[p_t] - \mu^{(2)}_{t-1,t+1} = 0 \quad \mathbb{P}-\text{a.s.} \tag{6.58}$$

Suppose for a moment that $\mu^{(2)}_{t,t+J}$ can be chosen such that Eq. 6.58 holds. Then the no-updating condition (Eq. 6.57) implies that the conditional forecast errors of all forecasts $\mu^{(2)}_{t-j,t}$, $j = 1, \dots, J$ for p_t vanish, that is,

$$\mathbb{E}_{t-1}[p_t] - \mu^{(2)}_{t-j,t} = 0, \quad j = 1, \dots, J, \quad \mathbb{P} - \text{a.s.} \tag{6.59}$$

By the law of iterated expectations, for each $j = 1, \dots, J$,

$$\mathbb{E}_{t-j}[p_t - \mu^{(2)}_{t-j,t}] = \mathbb{E}_{t-j}\left[\mathbb{E}_{t-1}[p_t - \mu^{(2)}_{t-j,t}]\right] = 0 \quad \mathbb{P} - \text{a.s.}$$

Hence all forecasts for p_t provide unbiased least-squares predictions conditional on the information of the respective periods as well.

The problem of obtaining unbiased forecasts is thus reduced to solving the linear Eq. 6.58 for $\mu^{(2)}_{t,t+J}$. Assuming $A^{(2J)}_t$ to be invertible, we obtain a perfect forecasting rule for first momonts:

$$\mu^{(2)}_{t,t+J} = A^{(2J)-1}_t \left[\mu^{(2)}_{t-1,t} - \sum_{j=1}^{J} A^{(1j)}_t \mu^{(1)}_{t,t+j} - \sum_{j=1}^{J-1} A^{(2j)}_t \mu^{(2)}_{t,t+j} + A_t\left(x_m - \mathbb{E}_{t-1}[\xi_t]\right)\right]$$

Define now the *expected excess demand* of all investors in period t except investor $(2, J)$ as

$$\begin{aligned}\Phi_{\text{ex}}\left(\mu^{(2)}_{t-1,t}, \left(\mu^{(iJ)}_t, \Sigma^{(iJ)}_t\right)_{i=1}^{2}\right) := &\sum_{j=1}^{J} \Phi^{(1j)}\left(\mu^{(2)}_{t-1,t}, \mu^{(1j)}_t, \Sigma^{(1j)}_t\right) \\ &+ \sum_{j=1}^{J-1} \Phi^{(2j)}\left(\mu^{(2)}_{t-1,t}, \mu^{(2j)}_t, \Sigma^{(2j)}_t\right) + \mathbb{E}_{t-1}[\xi_t] - x_m\end{aligned} \tag{6.60}$$

Using the fact that $A^{(2J)}_t = \frac{a^{(2)}}{R^J_f} A_t B^{(2J)}_{t,J}$, the preceding perfect forecasting rule for first moments takes the following form.

Proposition 6.4. *Let Assumptions 6.2 (2) and 6.4 be satisfied. If $B^{(2J)}_{t,J}$ as given in Eq. 6.51 is nonsingular for all $t \in \mathbb{N}$, a perfect forecasting rule for first moments $\Psi^{(2)}_{1\star}$ of Type-2 investors is given by a function*

$$\mu^{(2J)}_t = \Psi^{(2)}_{1\star}\left(\mu^{(2J)}_{t-1}, \Sigma^{(2J)}_t, \zeta_t\right) \tag{6.61}$$

such that

$$\begin{cases} \mu^{(2)}_{t,t+j} = \mu^{(2)}_{t-1,t+j}, \quad j = 1, \dots, J-1 \\ \mu^{(2)}_{t,t+J} := B^{(2J)-1}_{t,J}\left[C^{(2J)}_t \mu^{(2)}_{t-1,t} - \sum_{s=1}^{J-1} B^{(2J)}_{t,s} \mu^{(2)}_{t-1,t+s} - \frac{R^J_f}{a^{(2)}}\zeta_t\right] \end{cases}$$

where $\zeta_t = \Phi_{\text{ex}}\left(\mu^{(2)}_{t-1,t}, \left(\mu^{(iJ)}_t, \Sigma^{(iJ)}_t\right)^2_{i=1}\right)$ *is the expected excess demand of all investors except investor* $(2, \boldsymbol{J})$.

The forecasting rule (Eq. 6.61) is also referred to as an *unbiased forecasting rule*. It is a linear function of previously made forecasts and is *a priori* independent of previous prices. Since the matrices $B^{(2J)}_{t,J}$, $t \in \mathbb{N}$ depend essentially on second-moment beliefs, they can be chosen to be invertible. Hence, unbiased forecasting rules exist generically. From the perspective of an investor of Type 2, the unknown quantity for applying Eq. 6.61 is essentially the expected excess demand (Eq. 6.60). Inserting Eq. 6.61 into the price law (Eq. 6.55), the system of equations

$$\begin{cases} p_t &= \mu^{(2)}_{t-1,t} + A_t\left(\xi_t - \mathbb{E}_{t-1}[\xi_t]\right) \\ \mu^{(2J)}_t &= \Psi^{(2)}_{1\star}(\mu^{(2J)}_{t-1}, \Sigma^{(2J)}_t, \zeta_t) \\ \zeta_t &= \Phi_{\text{ex}}\left(\mu^{(2)}_{t-1,t}, \left(\mu^{(iJ)}_t, \Sigma^{(iJ)}_t\right)^2_{i=1}\right) \end{cases} \tag{6.62}$$

determines the asset prices of period t under unbiased expectations for investors of Type 2, given arbitrary beliefs of Type-1 investors.

A noticeable fact of an unbiased no-updating forecasting rule is that by Eq. 6.59 *all* forecasts $\mu^{(2)}_{t-j,t}$, $j = 1, \ldots, J$ for p_t are unbiased least-squares predictions conditional on information available at date $t-1$. In this sense, an unbiased no-updating rule yields the most precise forecasts as forecast errors vanish conditional on information that is not available at the stage in which they have been made. This property holds only for forecasts that feed back into the temporary equilibrium map in a nontrivial manner (see Wenzelburger, 2006).

Perfect Forecasting Rules for Second Moments

Though perfect forecasting rules for first moments exist generically, more restrictions are required to ensure the existence of a perfect forecasting rule for second moments. Let $\mathbb{V}_t[\cdot]$ and $\mathbb{C}\text{ov}_t[\cdot]$ denote the objective variance and covariance operator conditional on the σ-algebra $\mathcal{F}_t$. The correctness of second-moment beliefs requires the subjective (block matrix) entries of the matrices $\Sigma^{(2J)}_t$ defined in Eq. 6.48 to coincide with the corresponding objective moments, that is,

$$\Sigma^{(2J)}_{t,jj'} = \mathbb{C}\text{ov}_t\left[p_{t+j}, p_{t+j'}\right], \quad j, j' = 1, \ldots, \boldsymbol{J}, \quad \mathbb{P}-\text{a.s.} \tag{6.63}$$

for all times t. Under the assumption that investors of Type 2 use the unbiased no-updating forecasting rule (Eq. 6.61), Hillebrand and Wenzelburger (2006a) show that the covariance structure of the asset prices as given in Eq. 6.62 takes a particularly simple form.

Lemma 6.3. *Under the hypotheses of Proposition 6.4, assume that agents of Type* 2 *use the unbiased no-updating forecasting rule (Eq. 6.61). Then, for each* $t \in \mathbb{N}$ *and* *each $j, j' = 1, \ldots, J$, the covariance matrices of the associated asset prices (Eq. 6.62) are given by*

$$\mathbb{C}\mathrm{ov}_t\left[p_{t+j}, p_{t+j'}\right] = \begin{cases} \mathbb{E}_t\left[A_{t+j}\mathbb{V}_{t+j-1}[\xi_{t+j}]A_{t+j}\right] & \textit{for } j = j' \\ 0 & \textit{for } j \neq j' \end{cases}$$

Lemma 6.3 reveals that under unbiased no-updating, correlations between prices of distinct periods are zero. Furthermore, it implies that the covariance matrix of prices within each period $t \in \mathbb{N}$ satisfies

$$\mathbb{V}_t[p_{t+1}] = A_{t+1}\mathbb{V}_t[\xi_{t+1}]A_{t+1}$$

and

$$\mathbb{E}_{t-1}\left[\mathbb{V}_t[p_{t+j}]\right] = \mathbb{V}_{t-1}[p_{t+j}], \quad j = 1, \ldots, J-1 \tag{6.64}$$

As a consequence, under no-updating, the correct covariance matrices for young investors of Type 2 must be of the form

$$\Sigma_t^{(2J)} = \begin{bmatrix} \Sigma_{t,11}^{(2)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{t,JJ}^{(2)} \end{bmatrix}, \quad t \in \mathbb{N} \tag{6.65}$$

with $\Sigma_{t,jj}^{(2)} \in \mathcal{M}_k$. It follows from Eq. 6.64 that the block matrix entries in Eq. 6.65 have to satisfy the *consistency conditions*

$$\mathbb{E}_{t-1}\left[\Sigma_{t,jj}^{(2)}\right] = \Sigma_{t-1,(j+1)(j+1)}^{(2)}, \quad j = 1, \ldots, J-1, \quad t \in \mathbb{N} \tag{6.66}$$

Under no-updating, a perfect forecasting rule for second moments is now constructed as follows. In each period t, choose Eq. 6.65 so that

$$\Sigma_{t,jj'}^{(2)} = \begin{cases} \Sigma_{t-1,(j+1)(j+1)}^{(2)} & \text{for } j = j' < J \\ 0 & \text{for } j \neq j' \end{cases} \tag{6.67}$$

implying that the consistency conditions (Eq. 6.66) hold automatically. In view of Eq. 6.56, the remaining matrix $\Sigma_{t,JJ}^{(2)}$ has then to be specified in such a way that the second-moment belief $\Sigma_{t-1,11}^{(2)}$ formed in period $t-1$ is correct, that is,

$$\mathbb{V}_{t-1}[p_t] - \Sigma_{t-1,11}^{(2)} = A_t\mathbb{V}_{t-1}[\xi_t]A_t - \Sigma_{t-1,11}^{(2)} = 0 \tag{6.68}$$

Clearly, if Eq. 6.68 holds for all times t, it follows from Eq. 6.64, together with the no-updating condition (Eq. 6.67), that for each $j = 1, \ldots, J$,

$$\mathbb{V}_{t-j}[p_t] = \mathbb{E}_{t-j}\left[\mathbb{V}_{t-1}[p_t]\right] = \Sigma_{t-1,11}^{(2)} = \Sigma_{t-j,jj}^{(2)}, \quad t \in \mathbb{N} \tag{6.69}$$

such that all covariance matrices are correct. Thus the existence problem of perfect forecasting rules for second moments is reduced to finding symmetric and positive definite solutions $\Sigma^{(2)}_{t,JJ}$ to Eq. 6.68.

As in the one-period case, the existence proof is divided into two steps. First, find a symmetric positive definite matrix Γ_t such that $\Gamma_t^{-1} V_{t-1}[\xi_t]\Gamma_t^{-1} = \Sigma^{(2)}_{t-1,11}$. Second, setting

$$C_t := \left[\sum_{j=1}^{J} \frac{a^{(1)}}{R_f^j} C_t^{(1j)}\right]^{-1} \tag{6.70}$$

for notational ease, the desired second moment and hence a perfect forecasting rule for second moments is found if

$$\Gamma_t - C_t - a^{(2)} \sum_{j=1}^{J-1} \sum_{n=1}^{j} R_f^{2n-j} \Sigma^{(2)-1}_{t,nn} \tag{6.71}$$

is symmetric and positive definite. This result is summarized as follows.

Proposition 6.5. *Under the hypotheses of Proposition 6.4, assume that investors of Type 2 use the unbiased no-updating forecasting rule (6.61). Suppose that each $\Lambda_{t-1} := \mathbb{V}_{t-1}[\xi_t]$, $t \in \mathbb{N}$ is positive, set*

$$\Gamma_t = \sqrt{\Lambda_{t-1}} \sqrt{\left(\sqrt{\Lambda_{t-1}}\, \Sigma^{(2)}_{t-1,11} \sqrt{\Lambda_{t-1}}\right)^{-1}} \sqrt{\Lambda_{t-1}}, \quad t \in \mathbb{N} \tag{6.72}$$

and C_t, $t \in \mathbb{N}$ as defined in Eq. 6.70. Then the no-updating forecasting rule for second moments $\Psi^{(2)}_{2\star}$, given by

$$\Sigma_t^{(2J)} = \Psi^{(2)}_{2\star}\left(\Lambda_{t-1}, \Sigma_t^{(1J)}, \Sigma_{t-1}^{(2J)}\right) \tag{6.73}$$

such that

$$\Sigma^{(2)}_{t,jj'} := \begin{cases} 0 & \text{for } j \neq j' \\ \Sigma^{(2)}_{t-1,(j+1)(j+1)} & \text{for } j = j' < J \\ a^{(2)} R_f^J \left[\Gamma_t - C_t - a^{(2)} \displaystyle\sum_{j=1}^{J-1} \sum_{n=1}^{j} R_f^{2n-j} \Sigma^{(2)-1}_{t-1,(n+1)(n+1)}\right]^{-1} & \text{for } j = j' = J \end{cases}$$

provides correct second moments of the price process at date t in the sense that, Eq. 6.68 holds whenever Eq. 6.71 is positive definite.

As an immediate consequence of Proposition 6.5, Corollary 1 in Hillebrand and Wenzelburger (2006a) shows that perfect second-moment beliefs under no-updating generate zero autocorrelations between all asset prices of different periods within the planning horizon J. The covariance structure of the resulting asset price process is

$$\mathbb{C}\mathrm{ov}_t\left[p_{t+j}, p_{t+j'}\right] = \begin{cases} \mathbb{C}\mathrm{ov}_{t+j-1}[p_{t+j}, p_{t+j}], & \text{if } j = j' \\ 0, & \text{if } j \neq j' \end{cases}$$

for each $j, j' = 1, \ldots, J$ and all times $t \in \mathbb{N}$.

Note that the matrix Γ_t as defined previously is the same as for the one-period case in Proposition 6.2. The key problem again is that *a priori* the right side of Eq. 6.73 need neither be well defined nor positive definite such that a perfect forecasting rule for second moments may not exist at all. However, if either all investors agree on constant subjective second moments or all Type-1 investors believe in constant covariance matrices, existence obtains. Corollary 6.2 establishes existence in the first case, Corollary 6.3 in the second case.

Corollary 6.2. *Let the hypotheses of Proposition 6.5 be satisfied and assume, in addition, that $\mathbb{V}_{t-1}[\xi_t] \equiv \Lambda$ is constant over time. Suppose that all investors have homogeneous and constant second-moment beliefs. Then the constant forecasting rule*

$$\Sigma^{(2)}_{t,jj'} := \begin{cases} (\overline{a}\rho)^2\Lambda^{-1}, & \textit{if } j = j' \\ 0, & \textit{if } j \neq j' \end{cases} \qquad \textit{for all } j, j' = 1, \ldots, J,\ t \in \mathbb{N}$$

with

$$\overline{a} := a^{(1)} + a^{(2)} \quad \textit{and} \quad \rho := \sum_{j=1}^{J}\sum_{s=1}^{j} R_f^{2s-j} \tag{6.74}$$

provides correct second moments for all times $t \in \mathbb{N}$.

By Corollary 6.2, perfect homogeneous constant second-moment beliefs are proportional to the inverse of the covariance matrix of the noise-trader transactions, where $\overline{a}$ is the aggregate risk tolerance of the economy.

Corollary 6.3. *Let the hypotheses of Proposition 6.5 be satisfied and assume, in addition, that the following holds:*

1. *The covariance matrix of the noise-trader portfolios is constant over time and of the form $\mathbb{V}_{t-1}[\xi_t] \equiv \Lambda$ for all times $t \in \mathbb{N}$.*
2. *All Type-1 investors use constant second-moment beliefs and the matrix ΛC with $C \equiv C_t$ as given in Eq. 6.70 is symmetric positive definite such that*

$$\Lambda C = O^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_K) O$$

where all eigenvalues $\lambda_1, \ldots, \lambda_K$ are sufficiently large such that $\lambda_k > \frac{2\rho}{a^{(2)}}$ with ρ as defined in (6.74) and $O^\top O = I_K$.

Then any constant forecasting rule of the form

$$\Sigma^{(2)}_{t,jj'} := \begin{cases} C\,O^{\top}\mathrm{diag}(\kappa_1^{\pm},\ldots,\kappa_K^{\pm})O & \text{for } j = j' \\ 0 & \text{for } j \neq j' \end{cases}$$
$$j, j' = 1, \ldots, J,\ t \in \mathbb{N}$$

with eigenvalues $\kappa_k^{\pm}$ *given by*

$$\kappa_k^{\pm} = \left(\tfrac{\lambda_k}{2} - \rho a^{(2)}\right) \pm \sqrt{\left(\tfrac{\lambda_k}{2} - \rho a^{(2)}\right)^2 - \left(\rho a^{(2)}\right)^2}, \quad k = 1, \ldots, K$$

provides correct second moment beliefs for investors of Type 2 *for all times* t.

Observe again that by Corollary 6.3, perfect forecasting rules for second moments are not necessarily uniquely determined. The main informational constraint for applying the perfect forecasting rules given in Eqs. 6.61 and 6.71 is again the fact that the beliefs of investors are, in general, unobservable quantities.

6.4.4. Portfolio Holdings

The impact of different planning horizons on the portfolio holdings of investors is as follows. The first result concerns the portfolios of "rational" investors of Type 2 who are able to correctly predict the first two moments of the price process. Using the representation (Eq. 6.50) of the asset demand function, an immediate consequence of Propositions 6.4 and 6.5 and their corollaries is the following theorem.

Theorem 6.5. *Let the hypotheses of Proposition 6.2 be satisfied. Then the portfolio holding in period* t *of a rational investor of Type* 2 *with a planning horizon of length* j *is*

$$\Phi^{(2j)}(p_t, \mu_t^{(2j)}, \Sigma_t^{(2j)}) = a^{(2)} \sum_{s=1}^{j} \frac{1}{R_f^{j-s}} \Sigma_{t,ss}^{(2)-1} (\mu_{t,t+s}^{(2)} - R_f^s p_t)$$

where the beliefs $(\mu_t^{(2j)}, \Sigma_t^{(2j)}) \in \mathbb{R}^{Kj} \times \mathcal{M}_{Kj}$ *are given by the no-updating forecasting rules from Propositions 6.4 and 6.5.*

Theorem 6.5 states that the period-t portfolios of rational Type-2 investors are composed of $j+1$ mutual funds, j risky funds, and the risk-free asset.

The second result concerns the case in which beliefs of investors are homogeneous. This finding extracts the impact of different planning horizons on portfolios and prices as subjective beliefs depend only on the length j of a planning horizon. In particular, all investors within one cohort hold identical beliefs. Dropping the index i, one may write $(\mu_t^{(j)}, \Sigma_t^{(j)}) \in \mathbb{R}^{Kj} \times \mathcal{M}_{Kj}$ for the beliefs of cohort j, in period t instead of $(\mu_t^{(ij)}, \Sigma_t^{(ij)})$. If p_t is the market-clearing price of period t determined by Eq. 6.55, Hillebrand and

Wenzelburger (2006a) refer to the portfolio

$$x_t^{(j)} := \sum_{i=1}^{2} \Phi^{(ij)}(p_t, \mu_t^{(j)}, \Sigma_t^{(j)}) = \frac{\overline{a}}{R_f^j} \Pi_j^\top \Sigma_t^{(j)-1} (\mu_t^{(j)} - \Pi_j p_t) \qquad \textbf{(6.75)}$$

as the *aggregate generational portfolio* of cohort j, where as before $\overline{a}$ is the aggregate risk tolerance of the economy. It is straightforward to obtain the following $(J+1)$-*fund separation theorem*.

Theorem 6.6. *Let beliefs be homogeneous. Then the risky portfolio $x_t^{(ij)}$ held by an investor (i, j) after trading in period $t \in \mathbb{N}$ is given by a constant share of the aggregate generational portfolio (Eq. 6.75) of cohort j, such that*

$$x_t^{(ij)} = \frac{a^{(i)}}{\overline{a}} x_t^{(j)}$$

This share is determined by the individual risk tolerance $a^{(i)}$ relative to the aggregate risk tolerance $\overline{a}$.

Theorem 6.6 is a generalization of the two-fund separation theorem discussed in Section 6.2 to the multiperiod planning-horizon case. Under homogeneous beliefs, each investor will hold a combination of risk-free assets and a portfolio (or mutual fund) of risky assets that depend on the length of their planning horizon. The portfolio of risky assets is a multiple of the corresponding generational portfolio rather than of the market portfolio x_m. In view of Theorem 6.4, the J generational portfolios corresponding to different planning horizons will, in general, not be collinear. Therefore, even under homogeneous beliefs, planning horizons of distinct lengths will lead to structurally distinct portfolio holdings and hence to trade among different cohorts.

In light of the preceding results, the evolution of generational portfolios is illustrated in Figure 6.4 by means of an example. The reader is referred to Hillebrand and Wenzelburger (2006a, b) for further details. Suppose that only one risky asset ($K = 1$) is traded between two investors who live for three consecutive periods ($J = 2$).

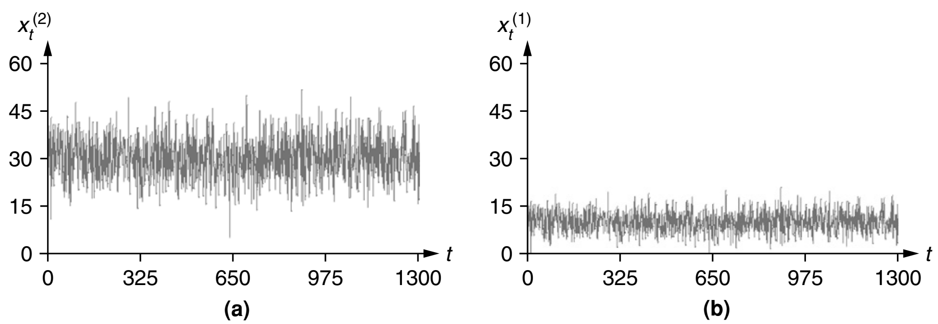

FIGURE 6.4 Risky portfolios of rational investors, $J = 2$, $K = 1$: (a) young investors; (b) middle-aged investors.

Investors have homogeneous rational expectations in the earlier sense. The evolution of the generational portfolios illustrated in Figure 6.4 shows the risky portfolios $x_t^{(j)}$ of young ($j = 2$) and middle-aged ($j = 1$) generations. Throughout the whole time span displayed in the figure, young investors hold more of the risky assets than middle-aged investors do, confirming the intuition stated earlier in this section that a longer planning horizon increases the investor's willingness to take on risk. Moreover, young investors' portfolios fluctuate more than those of middle-aged ones, indicating that fluctuations in asset prices are to a large extent absorbed by the members of the young cohort. Although these phenomena appear to be numerically robust, a theoretical explanation is still missing.

6.5. NONERGODIC ASSET PRICES

An old conjecture that dates back to Alchian (1950) and Friedman (1953) states that agents who do not learn to make accurate predictions about the future will be driven out of the market. This conjecture is a main pillar of the rational expectations paradigm but has been repeatedly challenged in the recent literature. In a series of papers (e.g., De Long, Shleifer, Summers, and Waldmann, 1990, 1991) the capability of noise traders to survive in financial markets has been analyzed. These results cover the static case with noisy errors only. In the present context of a dynamic CAPM, Alchian and Friedman's conjecture suggests that only those investors who hold efficient portfolios in the sense of Section 6.3.3 will survive in the long run as these attain the highest conditional Sharpe ratios. A first simulation analysis in Böhm and Wenzelburger (2005), however, suggests that an expert trader holding efficient portfolios may not be identified by consumers who select investment strategies on the basis of simple performance indicators.

This section reports on results of Horst and Wenzelburger (2008) regarding the asset-price dynamics of a variant of Eq. 6.43 in which investors have one-period planning horizons. They demonstrate that the stochastic difference equation (Eq. 6.43) may display a *path-dependent* and hence *nonergodic* asymptotic behavior, which can be described by a combination of analytical and numerical methods. The analysis has been inspired by Blume and Easley (1992, 2002, 2006) and Sandroni (2000), who analyze the question of whether markets favor accurate predictions. It also relates to a series of investigations by Evstigneev, Hens, and Schenk-Hoppé (2002, 2006, 2008), who investigate investment strategies that prevail in a market. In all these models, however, the portfolio choices of agents do not create a random environment. This feedback is one of the key features of the following example.

For tractability, the following simplifications are made. There are no dividend payments, and the noise-trader portfolios $\{\xi_t\}_{t\in\mathbb{N}}$ are governed by an *exogenous* IID process with mean $\bar{\xi}$ and a nondegenerate covariance matrix Λ. There is one type of consumer with initial endowment $e > 0$, risk aversion $\alpha > 0$, and a choice behavior as described in Example 6.3. The market is operated by two mediators, a *chartist* $i = 1$ and an *expert trader* $i = 2$, with rational expectations in the sense of Section 6.3.2. The profile of risk-adjusted market shares $a_t = \left(\frac{\eta_t}{\alpha}, \frac{1-\eta_t}{\alpha}\right)$ at any date t is determined by the chartist's market share $\eta_t \in [0, 1]$.

The chartist is a trend chaser who applies a simple *technical trading rule* of the form given in Eq. 6.44. The forecasts of the *expert trader* are unbiased and determined by an *unbiased forecasting rule* $\Psi^{(2)}_{1\star}$ as given in Eq. 6.26. For simplicity, both mediators are assumed to have correct beliefs on second moments. Since the covariance matrix Λ is constant over time, these correct second-moment beliefs are given by

$$V_t^{(i)} \equiv \left(\frac{1+r_f}{\alpha}\right)^2 \Lambda^{-1}, \quad i = 1, 2 \tag{6.76}$$

(see Section 6.3.2). Since cum-dividend and ex-dividend prices coincide, the system of stochastic difference equations (Eq. 6.43) that describes the evolution of asset prices takes the form

$$\begin{cases} p_t &= q^{(2)}_{t-1} + \frac{1+r_f}{\alpha}\Lambda^{-1}(\xi_t - \bar{\xi}) \\ q^{(1)}_t &= \sum_{n=1}^{N} D^{(n)} p_{t-n} \\ q^{(2)}_t &= \frac{1+r_f}{1-\eta_t} q^{(2)}_{t-1} - \frac{\eta_t}{1-\eta_t} q^{(1)}_t + \frac{(1+r_f)^2}{(1-\eta_t)\alpha} \Lambda^{-1}(x_m - \bar{\xi}) \end{cases} \tag{6.77}$$

where the last equation is the unbiased forecasting rule (Eq. 6.26). Setting

$$X_t := (q^{(2)}_t, q^{(2)}_{t-1}, p_t, \ldots, p_{t-N-1}) \in \mathbb{R}^d$$

with $d = K(N+4)$, the system of difference equations (Eq. 6.77) takes the form

$$X_t = \mathbf{A}(\eta_t) X_{t-1} + \mathbf{B}(\eta_t, \xi_t), \quad t \in \mathbb{N}, \tag{6.78}$$

where, by abuse of notation, the $d \times d$ coefficient matrix $\mathbf{A}(\eta_t)$ and the vector $\mathbf{B}(\eta_t, \xi_t) \in \mathbb{R}^d$ corresponding to the representation (Eq. 6.45) are given by

$$\begin{aligned} \mathbf{A}^{(0)}(\eta) &:= \frac{1+r_f}{1-\eta} I_K & \mathbf{A}^{(n)}(\eta) &:= -\frac{\eta}{1-\eta} D^{(n)}, \ n = 1, \ldots, N \\ \mathbf{B}^{(0)}(\eta) &:= \frac{(1+r_f)^2}{(1-\eta)\alpha} \Lambda^{-1} (x_m - \bar{\xi}) & \mathbf{B}^{(1)}(\xi) &:= \frac{1+r_f}{\alpha} \Lambda^{-1} (\xi - \bar{\xi}) \end{aligned}$$

To guarantee long-run stability of Eq. 6.78, the chartist's market share is restricted to a compact interval $[\underline{\eta}, \overline{\eta}]$. For any fixed market share $\eta \in [\underline{\eta}, \overline{\eta}]$, the corresponding process $\{X^\eta_t\}_{t \in \mathbb{N}}$ with market shares frozen at η is Markovian and defined by the linear recursive relation

$$X^\eta_t = \mathbf{A}(\eta) X^\eta_{t-1} + \mathbf{B}(\eta, \xi_t), \quad t \in \mathbb{N} \tag{6.79}$$

If Assumption 6.3 (1) is satisfied, each sequence $\{X^\eta_t\}_{t \in \mathbb{N}}$, $\eta \in [\underline{\eta}, \overline{\eta}]$ is bounded and the difference equation Eq. 6.79 admits a unique stationary solution, that is, a unique stationary and ergodic process $\{x^\eta_t\}_{t \in \mathbb{N}}$ that satisfies Eq. 6.79. For any starting point x, the distributions τ^η_t of X^η_t converge weakly to the distribution τ^η of $\{x^\eta_t\}_{t \in \mathbb{N}}$ as $t \to \infty$. The distribution τ^η is the unique stationary distribution of the Markov process $\{X^\eta_t\}_{t \in \mathbb{N}}$.

6.5.1. Characterization of Long-Run Equilibria

It turns out that the possible long-run equilibria of the market, that is, the possible limiting distributions of the sequence $\{X_t\}_{t\in\mathbb{N}}$ generated by Eq. 6.78 can be characterized by a fixed-point property. Numerical simulations suggest that the long-run market shares settle down to constant values that depend on initial conditions, the intensity of choice, and on the chosen random environment. For this reason, it is sensible to start the analysis from the assumption that the process of market shares $\{\eta_t\}_{t\in\mathbb{N}}$ converges almost surely as $t \to \infty$. To this end, let $\tau^\eta(f)$ denote the integral of a bounded function f with respect to the distribution τ^η.

Theorem 6.7. *Suppose that Assumption 6.3 is satisfied and that the process of market shares $\{\eta_t\}_{t\in\mathbb{N}}$ converges almost surely to some random variable η_*. Then the sequence of empirical averages $\{\varrho_t\}_{t\in\mathbb{N}}$ as defined in Eq. 6.39 converges almost surely weakly to a random limiting distribution. Specifically,*

$$\text{Prob}\left[\lim_{t\to\infty} \varrho_t(f) = \tau^{\eta_*}(f)\right] = 1$$

for all bounded, continuous functions f, where τ^{η_*} is the stationary distribution of the Markov process $\{X_t^{\eta_*}\}_{t\in\mathbb{N}}$.

Theorem 6.7 implies that the distributions of $\{X_t\}_{t\in\mathbb{N}}$ converge weakly to a *random* limiting distribution, provided that the sequence of market shares settles down to a random limit in the long run. This result imposes the following consistency condition between long-run market shares and limiting distributions of $\{X_t\}_{t\in\mathbb{N}}$. Define a map $\zeta : [\underline{\eta}, \overline{\eta}\,] \to \mathbb{R}^L$ by

$$\zeta(\eta) := \left(\tau^\eta(f^{(1)}), \ldots, \tau^\eta(f^{(L)})\right) \tag{6.80}$$

which assigns to any fixed market share η the *long-run empirical average* of the corresponding Markov process $\{X_t^\eta\}_{t\in\mathbb{N}}$. The question of existence and uniqueness of limiting distributions of $\{X_t\}_{t\in\mathbb{N}}$ can now be reduced to two equivalent fixed point conditions, as follows.

Corollary 6.4. *Under the assumptions of Theorem 6.7, the random variable η_* takes values in the set of fixed points*

$$\mathcal{E} := \left\{\eta \in [\underline{\eta}, \overline{\eta}] \;:\; 0 = F \circ \zeta(\eta) - \eta\right\}$$

where F is the choice function defined in Eq. 6.42. The long-run empirical averages of the process $\{X_t\}_{t\in\mathbb{N}}$ are given by $\zeta(\eta_*)$ and take values in the set of fixed points

$$\mathcal{S} := \left\{z \in \mathbb{R}^L \;:\; 0 = \zeta \circ F(z) - z\right\}$$

Notice that $\zeta(\eta_*)$ is a random variable so that the possible long-run empirical averages of $\{X_t\}_{t\in\mathbb{N}}$ are random. It is well known that the map $\zeta : [\underline{\eta}, \overline{\eta}\,] \to \mathbb{R}^L$ is continuous, but typically no analytical expressions are available (see Brandt, 1986). However, good numerical approximations of ζ may easily be obtained by simulating the benchmark processes $\{X_t^\eta\}_{t\in\mathbb{N}}$, $\eta \in [\underline{\eta}, \overline{\eta}]$ as defined in Eq. 6.79. As the choice function F is analytically specified, an accurate numerical approximation of the sets $\mathcal{E}$ and $\mathcal{S}$ and thus of the possible long-run equilibria of Eq. 6.78 is available.

6.5.2. Convergence to Long-Run Equilibria

Because the characterization result stated in Corollary 6.4 rests on the assumption that the sequence of market shares $\{\eta_t\}_{t\in\mathbb{N}}$ converges to some random constant η_*, it remains to state conditions under which this assumption is satisfied. If the intensity of choice β is sufficiently small so that consumers are little inclined to switch between mediators, it is intuitively clear that market shares settle down to a constant limit. It turns out, however, that this property may be lost and the limit is random if β is sufficiently large, reflecting strong social interaction between consumers. Social interactions may thus constitute an *endogenous* source of randomness.

The evolution of prices, beliefs, and market shares can be analyzed by means of an approximating ordinary differential equation. Recall that

$$z_t = \left(\varrho_t(f^{(1)}), \ldots, \varrho_t(f^{(L)})\right) \in \mathbb{R}^L$$

denotes the vector of empirical averages at time t. Using the choice function (Eq. 6.42), the stochastic difference equation (Eq. 6.78) takes the form

$$\begin{cases} X_t = \mathbf{A}\big(F(z_t)\big)X_{t-1} + \mathbf{B}\big(F(z_t), \xi_t\big) \\ z_t = \frac{t-1}{t} z_{t-1} + \frac{1}{t}\left(f^{(1)}(X_{t-1}), \ldots, f^{(L)}(X_{t-1})\right) \end{cases} \tag{6.81}$$

Define a map $g : \mathbb{R}^L \to \mathbb{R}^L$ by setting

$$g(z) := \zeta \circ F(z) - z$$

The zeros of the map g are given by the set $\mathcal{S}$ defined in Corollary 6.4, whereas continuity of F and ζ implies continuity of g. With this notation, a well-known stochastic approximation result found in Chapter 5 of Kushner and Yin (2003) can be applied to yield the following result.

Theorem 6.8. *Suppose that the map g is Lipschitz continuous such that the* ordinary differential equation

$$\dot{z} = g(z) \tag{6.82}$$

admits a unique solution. Let $\mathcal{S}^* := \{\bar{z}_1, \ldots, \bar{z}_N\} \subset \mathcal{S}$ be the set of asymptotically stable steady states with corresponding basins of attraction $DA(\bar{z}_i)$.

If the sequence $\{z_t\}_{t\in\mathbb{N}}$ of empirical averages visits a compact subset of some basin $DA(\bar{z}_i)$ infinitely often with probability $\mathbf{p} > 0$, then the following holds:

1. *The sequence $\{z_t\}_{t\in\mathbb{N}}$ converges to $\bar{z}_i$ with at least probability $\mathbf{p}$, that is,*

$$\lim_{t\to\infty} |z_t - \bar{z}_i| = 0 \ \textit{ with at least probability } \mathbf{p}$$

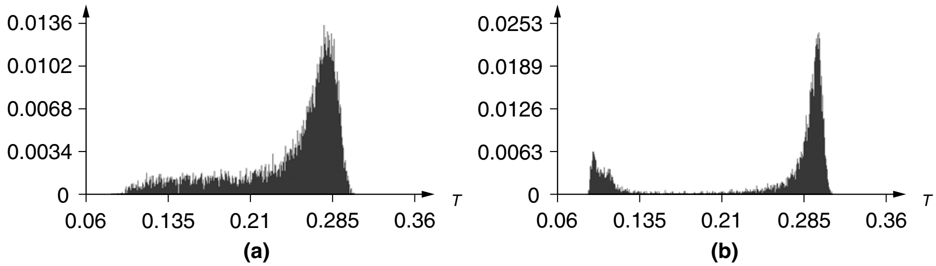

FIGURE 6.5 Empirical distributions of market shares for $\beta = 2$ and $\eta_0 = 0.355$: (a) distribution for $T = 500$; (b) distribution for $T = 1{,}000$.

2. *The process of market shares* $\{\eta_t\}_{t\in\mathbb{N}}$ *converges with at least probability* $\mathbf{p}$ *to a stationary value* $F(\bar{z}_i) \in \mathcal{E}$ *with* $\mathcal{E}$ *as defined in Corollary 6.4.*

Theorem 6.8 reduces the convergence properties of the asset price process to the asymptotic behavior of a deterministic ordinary differential equation (Eq. 6.82) on the level of empirical averages. In particular, market shares converge almost surely to some constant if Eq. 6.82 has a unique, globally asymptotically stable steady state. In this case the corresponding asset price process is ergodic. If, on the contrary, Eq. 6.82 admits multiple stable steady states, then ergodicity breaks down and market shares along with asset prices converge to a random limit, thus endogenously creating uncertainty. This phenomenon is illustrated in the following section.

6.5.3. Performance of Efficient Portfolios

Common folklore suggests that a rational expert trader who has correct beliefs and hence holds efficient portfolios as introduced in Section 6.3.3 will attain larger market shares than a nonrational investor with incorrect beliefs such as a chartist. However, it is intuitively clear that the expert trader will attract more consumers only if the performance measures point at efficient portfolios. Otherwise, portfolios other than the efficient portfolio appear to perform better and the nonrational investor attracts more consumers. One reason that such a scenario could indeed occur is that any performance measure involves empirical estimators, which need not be consistent. The following simulation exercise demonstrates that the chartist and the rational expert trader often coexist in the market.

Setting the risk-free rate to $r_f = 1\%$, the stability condition of Eq. 6.78 is satisfied if the chartist's market share η_t is bounded from below by $\underline{\eta} = 6\%$ and from above by $\overline{\eta} = 36\%$. All empirical densities displayed here are calculated using a sample of $N = 100{,}000$ independent repetitions. All simulation results were generated by the software package Macrodyn.

Notice first that empirical distributions converge rather slowly. Figure 6.5 displays the empirical densities of market shares after $T = 500$ and $T = 1000$ periods, respectively. Hence the application of Theorem 6.8 is worthwhile.

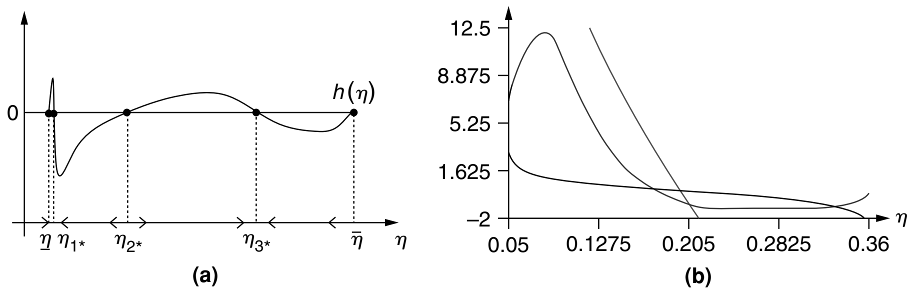

FIGURE 6.6 Long-run market shares: empirical average returns, $\dot{\eta} = h(\eta)$: (a) Sharpe ratios; (b) $\beta = 0.5$ (*steep line*) and $\beta = 2$ (*flat line*).

Empirical Returns as Performance Measures

In comparing the performance of the two mediators using empirical averages of returns, it is best to use the difference $z_t = \varrho_t(f^{(1)} - f^{(2)})$ with $f^{(i)}$, $i = 1, 2$ as defined in Example 6.1. The ODE defining map g is then one-dimensional. It is shown in Horst and Wenzelburger (2008) that in this case the ODE (Eq. 6.82) is topologically conjugate to an ODE $\dot{\eta} = h(\eta)$ in terms of market shares, where $h : [\underline{\eta}, \overline{\eta}] \to \mathbb{R}$ is a continuous function satisfying $h(\underline{\eta}) = h(\overline{\eta}) = 0$. The conjugacy of the two ODEs means that their qualitative behavior is the same (see Arrowsmith and Place, 1994). Thus, the long-run behavior of empirical averages is precisely described by the long-run behavior of market shares. The function h can be approximated using a numerical approximation of ζ, which is obtained by simulating the benchmark models (Eq. 6.79). As indicated in Figure 6.6a, h has *five* steady states for $\beta = 2$: three in the interior of the interval $[\underline{\eta}, \overline{\eta}]$ along with both of its boundary points.

Using standard arguments of the theory of ODEs, the phase diagram of the ODE in Figure 6.6a shows that all solutions converge to one of the two asymptotically stable steady states, which are η_{1*} and η_{3*}. Their respective basins of attraction are simply separated by the unstable steady state η_{2*} in the middle. Translated into the stochastic behavior of Eq. 6.81, Theorems 6.7 and 6.8 guarantee convergence of both market shares and asset prices. The respective limits, however, are not necessarily unique and will depend on the random noise-trader transactions as well as on the initial market share η_0. Since the lowest and the highest possible market share of the chartists are unstable under the dynamics of the ODE, the long-run market share of the chartists will always lie in the open interval $(\underline{\eta}, \overline{\eta})$. Setting aside that $\underline{\eta}$ and $\overline{\eta}$ have been set by the modelers to ensure boundedness of the price processes, this may be interpreted as a situation in which both mediators will prevail in the market.

A simulation study of Eq. 6.81 confirms this intuition, demonstrating that the long-run market shares are indeed random. The probability with which they converge to the possible steady states depends both on the initial condition η_0 and the intensity of choice β. Figure 6.7 shows the empirical distribution of market shares after $T = 10{,}000$ periods for $N = 100{,}000$ independent samples of η_T when chartists initially have a market share

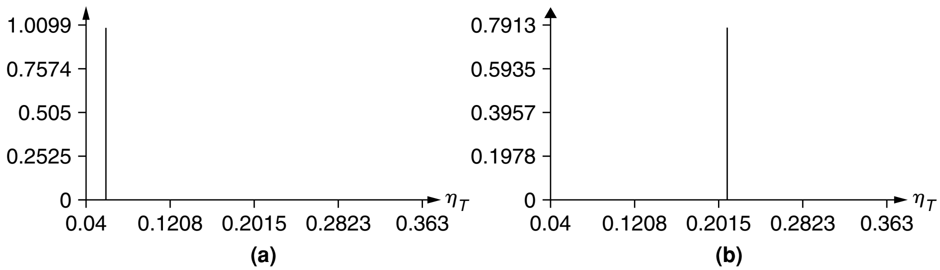

FIGURE 6.7 Empirical distributions of market shares for average returns, $\beta = 2$: (a) initial market share $\eta_0 = 0.065$; (b) initial market share $\eta_0 = 0.15$.

of 6.5% and 15%, respectively. The densities will be concentrated around the two stable steady states of the ODE in Figure 6.6a, respectively.

Empirical Sharpe Ratios as Performance Measures

The analysis of asymptotic market shares becomes more involved when the performance of mediators is judged by comparing historical Sharpe ratios. In this case the choice function (6.42) is no longer invertible, because the performance measure as defined in Example 6.2 is based on a four-dimensional vector $z_t \in \mathbb{R}^4$. As a consequence, the evolution of the system is described by a four-dimensional ODE. The set of asymptotic market shares as given in Corollary 6.4 allows the representation

$$\mathcal{E} = \left\{ \eta \in [\underline{\eta}, \overline{\eta}\,] : \varphi^{-1}(\eta, \beta) = \Delta\Pi \circ \zeta(\eta) \right\} \qquad \textbf{(6.83)}$$

where $\varphi^{-1}(\cdot, \beta)$ is the inverse of the logit function (Eq. 6.41) and $\pi^{(2)} - \pi^{(1)} = \Delta\Pi \circ \zeta(\eta)$ describes the stationary difference in Sharpe ratios of the two mediators. Although $\varphi^{-1}(\cdot, \beta)$ is analytically available, the map $\Delta\Pi \circ \zeta$ on the right side must be obtained by simulating the benchmark models (Eq. 6.79).

The hump-shaped curve in Figure 6.6a represents a numerical approximation of $\Delta \circ \zeta$. Figure 6.6b indicates that the two functions in Eq. 6.83 have three intersection points, provided that the intensity of choice β is sufficiently large. The leftmost intersection point ($\beta = 2$) is hardly visible but exists because $\varphi^{-1}(\cdot, 2)$ has two vertical asymptotes at $\underline{\eta}$ and $\overline{\eta}$, respectively. These intersection points characterize the possible long-run market shares of the chartist for $\beta = 2$. If, however, the impact of the mediators' performances is weak, that is, if β is low enough such that $\varphi^{-1}(\cdot, \beta)$ is sufficiently steep as for $\beta = 0.5$, the possible limit is uniquely determined.

Simulations of the nonlinear model Eq. 6.81 reveal again that the asymptotic behavior of market shares is random. The long-run market share depends strongly on the intensity of choice and is sensitive to initial conditions. For $\beta = 1$, Figure 6.8 depicts the empirical densities after $T = 10{,}000$ periods, which correspond to an initial market share of $\eta_0 = 6.5\%$ in panel (a) and for $\eta_0 = 35.5\%$ in panel (b), respectively.

Similar observations are made for $\beta = 2$. In this case, however, the distribution of market shares may be bimodal. Figure 6.8 shows the empirical distribution for

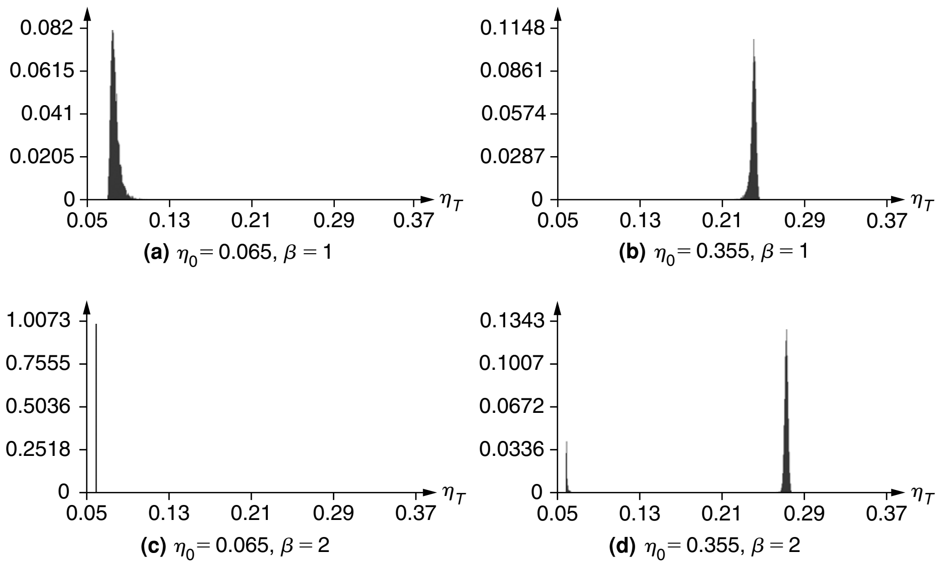

FIGURE 6.8 Empirical distributions of market shares for Sharpe ratios.

$\eta_0 = 6.5\%$ in panel (c) and for $\eta_0 = 35.5\%$ in panel (d). If $\eta_0 = 6.5\%$, the chartist almost "dies out" in the sense that her long-run market share is pushed to the leftmost fixed point. For $\eta_0 = 35.5\%$ the empirical distribution of asymptotic market shares is bimodal, with two peaks that are approximately located at the outer intersection points of the respective functions depicted in Figure 6.6b. If the chartist's initial market share is sufficiently high, she will "survive" with positive probability. The bimodality demonstrates that the long-run market shares depend also on the specific history generated by the noise-trader transactions.

These two examples show that the long-run market shares of two financial mediators may strongly depend on the random environment of the market that is created by the choice behavior of consumers. As a consequence, asset prices may become nonergodic as the price process converges in distribution, with the limiting distribution being path dependent. Economically, this result implies that social interaction among consumers may endogenously create a risk that leads to inefficient portfolio holdings.

6.5.4. A Boom-and-Bust Scenario

The previous scenario can be extended to one in which consumers have a two-period planning horizon $J = 2$ and whose investment opportunities consist of one risky ($K = 1$) and one risk-free asset, as introduced in Section 6.4. Assume, in addition, that young consumers may select between two mediators, a chartist ($i = 1$) and an expert trader ($i = 2$), who carry out their investment decision. To ensure boundedness of the price process, let the minimal and maximal market shares of the chartist be $\underline{\eta} = 0$ and $\overline{\eta} = 0.6$,

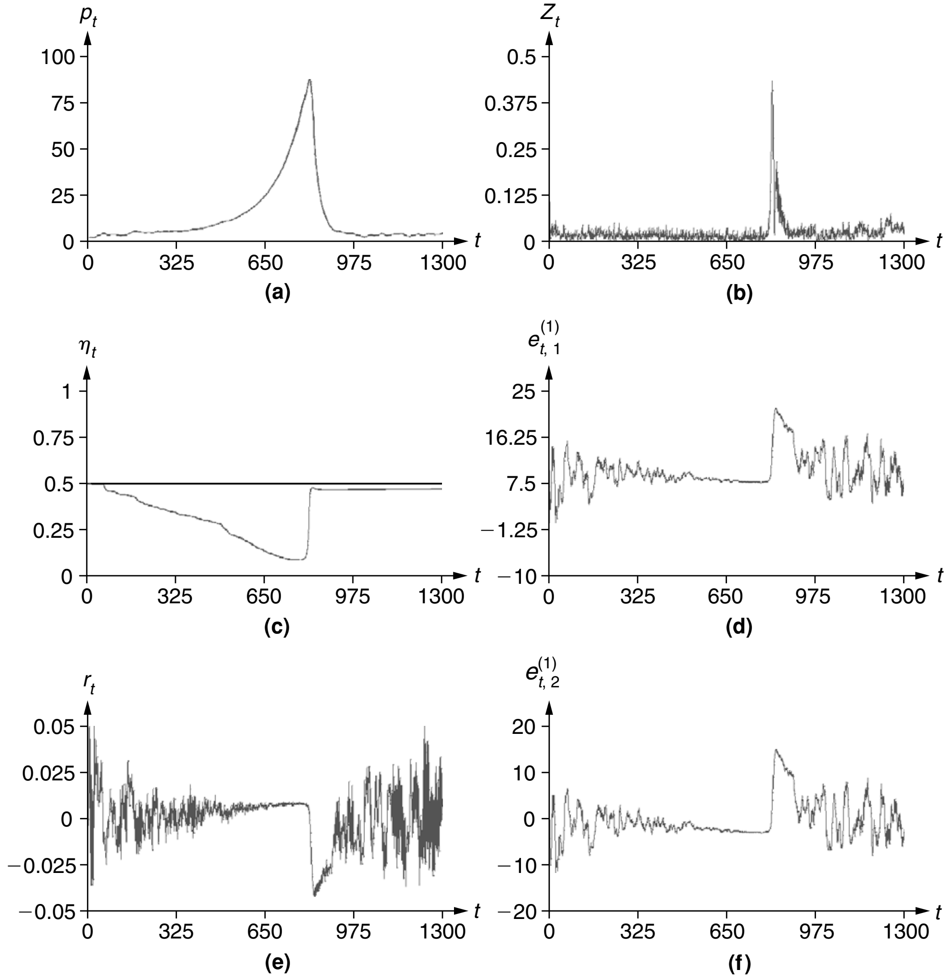

FIGURE 6.9 Market shares, trading volume, prices, and forecast errors in a multiperiod case: (a) asset prices; (b) trading volume; (c) chartist's market shares; (d) chartist's one-period-ahead forecast errors in %; (e) asset returns; and (f) chartist's two-period-ahead forecast errors in %.

respectively. The intensity of choice parameter is $\beta = 0.3$. For further details, we refer to Hillebrand and Wenzelburger (2006b).

Figure 6.9 portrays time windows of the chartist's market share, the mediators' forecast errors, asset prices, and asset returns. The most prominent phenomenon observed in Figure 6.9a is a sudden boom in asset prices within the time window [600, 900], followed by a bust, after which the price process continues to fluctuate about its initial level (≈ 5). The time window displayed in Figure 6.9c shows that the boom in asset prices is accompanied by a rapid decrease in the chartist's market share, which reaches

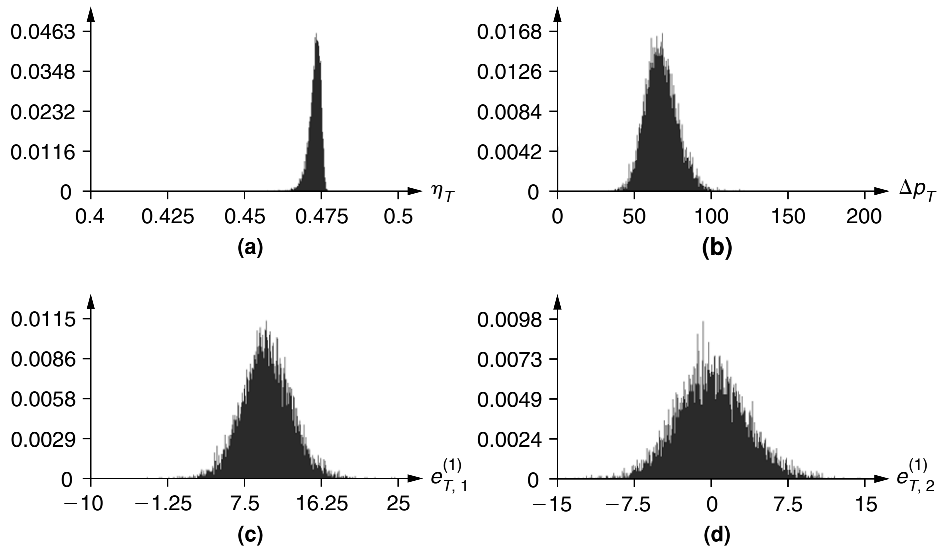

FIGURE 6.10 Empirical distributions at $T = 1500$, $\beta = 0.3$: (a) chartist's long-run market shares; (b) price range; (c) chartist's one-period-ahead forecast errors in %; and (d) chartist's two-period-ahead forecast errors in %.

a minimum at $\approx 10\%$. After that, her market share increases again and finally returns to a constant level of ≈ 45–48%. Throughout the entire time window in Figure 6.9c, the chartist's market share takes values below the 0.5-line and is thus lower than the expert's share. Asset returns in Figure 6.9e exhibit the phenomenon of volatility clustering. The trading volume per capita is depicted in Figure 6.9b, showing that the boom is accompanied by a sudden increase in the trading volume, where the trading activity is slightly delayed with respect to the price boom.

During the boom phase the chartist's relative forecast errors as displayed in Figures 6.9d and 6.9f converge rapidly to about 10% and 0%, respectively, which can be attributed to the increased level of asset prices during the boom. In the bust phase after the price peak, the forecast errors increase again and eventually fluctuate within their initial ranges.

Due to the nonlinearity of the system, the long-run behavior of generated processes is not necessarily ergodic. Hence one has to be careful when inferring results from time averages rather than from averages over alternative sample paths of the noise process. Empirical distributions in which the variable of interest is evaluated for a random sample of 10^4 paths of the noise process provide some confirmation. Figure 6.10a shows that the long-run market shares of the chartist in period $T = 1500$ are below 50% with positive probability.

Numerical evidence for the boom/bust scenario is obtained by evaluating the random variable

$$\Delta p_T := \max_{0 \le t \le T} \left\{ |p_t - p_T| \right\}$$

which is an estimate of the price range until period T. Its empirical distribution for $T = 1500$ and 10^4 sample paths are displayed in Figure 6.10b, where the initial asset price has been set to $p_0 = 10$. From the shape of the distribution it can be inferred that the price peak as displayed in Figure 6.10b occurs with positive probability. Finally, Figures 6.10c and 6.10d display densities of the relative forecast errors of the chartist. These fluctuate between ±7%, with a 10% bias of the one-period-ahead forecasts.

Hillebrand and Wenzelburger (2006b) provide numerical evidence that market shares freeze to constant values in the long run, but the relative extreme boom/bust scenario has not yet been reported for the one-period case. Given the relatively low intensity of choice and the linearity of the forecasting rules, this leads to the conclusion that heterogenous planning horizons seem to have a significant potential to amplify booms and busts in models in which consumers are allowed to switch between mediators. Whether price booms and busts follow nonergodic patterns remains an interesting research topic for the future.

6.6. CONCLUSION

Choosing the traditional CAPM as a starting point, we reviewed a fully explicit dynamic CAPM with interacting heterogeneous agents. We carefully distinguished between a *temporary equilibrium map* describing the basic market mechanism of an asset market and a *forecasting rule*, which models the way in which an investor forms expectations. The combination of both components with a model for exogenous perturbations stipulates an explicitly defined *deterministic dynamical system* in a *random environment*. This provides a simple but general framework for the investigation of endogenous processes of asset prices and allocations of multiple risky assets that are driven by agents' characteristics.

Much emphasis was placed on a benchmark case in which one type of investor has rational expectations while other market participants may have erroneous beliefs. This was achieved adopting the notion of a *perfect forecasting rule*, which allowed for a generalization of some of the fundamental concepts in traditional finance. The most central one is the notion of a *modified market portfolio* that accounts for diversity of beliefs. This portfolio is efficient in the sense of Markowitz and attains the highest possible conditional Sharpe ratios along any price path. Introducing the concept of a *generational portfolio* for a multiperiod setting, we showed that under homogeneous beliefs investors will hold a constant proportion of distinct generational portfolios that correspond to the length of their planning horizon. Finally, we also reviewed a multifund separation theorem.

When agents are allowed to switch between financial mediators, asset prices may behave in a nonergodic manner if their interactive complementarities are too strong. In such cases price processes converge in distribution, but the limiting distribution is path dependent. It was shown that a nonrational investor such as a chartist may attain a larger market share than a rational expert, who may even be driven out of the market. In this sense social interaction between agents may endogenously generate the risk to hold inefficient portfolios as the efficient ones may not be identified. Long-run market shares of financial mediators depend significantly on the random environment generated by the social interaction. Finally, numerical evidence was provided that distinct planning horizons may be responsible for volatility clustering in time series of returns (as well as prices) and that they may have an amplifying effect on booms and busts.

While a vast amount of ongoing research explores the dynamic behavior of asset prices, the insights of this survey point at several potentially fruitful research topics. The first and perhaps most ambitious one is to extend the setup with multiperiod planning horizons to a macroeconomic model by incorporating a real sector. This direction is pursued in Hillebrand (2008), who compares alternative pension systems with production in a stochastic multiperiod overlapping generations model. Hillebrand provides the framework appropriate to address the question of how to safeguard old-age consumption. His approach also allows for utility functions with constant relative risk aversion and thus for an investigation into how endogenous wealth processes and intertemporal consumption streams feed back into an asset price process. The sequential nature of the approach described in this survey should facilitate analytical investigations and help reduce the complexity of numerical computations that seem to become more and more popular (e.g., see Kübler and Schmedders, 2005).

A second research issue is to incorporate richer structures of securities such as futures, derivatives, and securities with collateral as, for example, in Magill and Quinzii (1996, 2000) or Kübler and Schmedders (2003). Detemple and Selden (1991) demonstrate that in a static general equilibrium, primary and derivative asset markets, generically, interact. This result stands in contrast to the arbitrage pricing theory and seems to have been overlooked by the literature so far. An extension to a dynamic setup that allows for an investigation of the effects of option markets on asset prices and allocations remains to be developed. Moreover, research into future markets with applications to resource markets such as the oil market might prove fruitful. The third research issue is to incorporate strategic interaction into dynamic models with interacting agents. This should contribute to a better understanding of how asset prices can be influenced by large investors. An interesting topic, for example, is the question of which extent central banks are capable of controlling exchange rates. Finally, an increasing amount of research is now being undertaken to explore the stability of banking systems. Pyle (1971) is the first who has linked Markowitz's portfolio theory to the theory of financial intermediation. However, the models developed from Pyle's approach are—as most of the models in the banking theory—static (see Chapter 8 of Freixas and Rochet, 1997). Dynamic models that describe the interaction of commercial banks remain to be established.

References

Alchian, A., Uncertainty, evolution and economic theory, *Journal of Political Economy*, 1950, 58, 211–221.

Allingham, M., Existence theorems in the CAPM, *Econometrica*, 1991, 59, 1169–1174.

Anderson, S. P., A. de Palma, and J.-F., Thisse, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.

Angeletos, G.-M., and L.-E. Calvet, Incomplete-markets dynamics in a neoclassical production economy, *Journal of Mathematical Economics*, 2005, 41(4), 407–438.

Angeletos, G.-M., and L.-E. Calvet, Idiosyncratic risk, growth and the business cycle, *Journal of Monetary Economics*, 2006, 53, 1095–1115.

Arnold, L., *Random Dynamical Systems*, Springer-Verlag, 1998.

Arrowsmith, D. K., and C. M. Place, *Introduction to Dynamical Systems*, Cambridge University Press, 1994.

Black, F., Capital market equilibrium with restricted borrowing, *Journal of Business*, 1972, 45, 444–455.

Blume, L. E.,, The statistical mechanics of strategic interaction, *Games and Economic Behavior*, 1993, 5, 387–424.

Blume, L. E., and D. Easley, Evolution and market behavior, *Journal of Economic Theory*, 1992, 58, 9–40.

Blume, L. E., and D. Easley, Optimality and natural selection in markets, *Journal of Economic Theory*, 2002, 107, 95–135.

Blume, L. E., and D. Easley, If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets, *Econometrica*, 2006, 74, 929–966.

Böhm, V., CAPM basics, Discussion Paper 495, Department of Economics, Bielefeld University, 2002.

Böhm, V., and C. Chiarella, Mean variance preferences, expectations formations, and the dynamics of random asset prices, *Mathematical Finance*, 2005, 15, 61–97.

Böhm, V., N. Deutscher, and J. Wenzelburger, Endogenous random asset prices in overlapping generations economies, *Mathematical Finance*, 2000, 10, 23–38.

Böhm, V., and J. Wenzelburger, Expectations, forecasting and perfect foresight: A dynamical systems approach, *Macroeconomic Dynamics*, 1999, 3, 167–186.

Böhm, V., and J. Wenzelburger, Perfect predictions in economic dynamical systems with random perturbations, *Macroeconomic Dynamics*, 2002, 6, 687–712.

Böhm, V., and J. Wenzelburger, On the performance of efficient portfolios, *Journal of Economic Dynamics and Control*, 2005, 29, 721–740.

Brandt, A., The stochastic equation $Y_{t+1} = A_t Y_t + B_t$ with stationary coefficients, *Advances of Applied Probability*, 1986, 18, 211–220.

Brock, W. A., and S. N. Durlauf, Discrete choice with social interactions, *Review of Economic Studies*, 2001, 68: 235–260.

Brock, W. A., and C. H. Hommes, Models of complexity in economics and finance, in B. Hanzon, C. Heij, J. Schumacher and C. Praagman (eds.), *Systems Dynamics in Economic and Finance Models*, Chap. 1, 3–44. Wiley, 1997a.

Brock, W. A., and C. H. Hommes, A rational route to randomness, *Econometrica*, 1997b, 65, 1059–1095.

Brock, W. A., and C. H. Hommes, Heterogeneous beliefs and routes to chaos in a simple asset-pricing model, *Journal of Economic Dynamics and Control*, 1998, 22, 1235–1274.

Chamberlain, G., A characterization of the distributions that imply mean-variance utility functions, *Journal of Economic Theory*, 1983, 29, 185–201.

Chiarella, C., R. Dieci, and X. He, Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework, *Journal of Behavior and Organization*, 2007, 62, 408–427.

Chiarella, C., and X.-Z. He, Heterogenous beliefs, risk and learning in a simple asset-pricing model, *Computational Economics*, 2002, 19, 95–132.

Chiarella, C., and X.-Z. He, Heterogenous beliefs, risk and learning in a simple asset-pricing model with a market maker, *Macroeconomic Dynamics*, 2003, 7, 503–536.

Cochrane, J. H., *Asset Pricing,* revised edition, Princeton University Press, 2005.

Copeland, T. E., J. F. Weston, and K. Shastri, *Financial Theory and Corporate Policy*, 4th ed., Pearson Addison Wesley, 2005.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, An intertemporal general equilibrium model of asset prices, *Econometrica*, 1985, 53, 363–384.

Cuthbertson, K., *Quantitative Financial Economics*, John Wiley and Sons, 1996.

Dana, R. A., Existence and uniqueness of equilibria when preferences are additively separable, *Econometrica*, 1993, 61, 953–957.

Dana, R. A., Existence, uniqueness and determinacy of equilibrium in CAPM with a riskless asset, *Journal of Mathematical Economics*, 1999, 32, 167–175.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, Noise traders risk in financial markets, *Journal of Political Economy*, 1990, 98, 703–738.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, The survival of noise traders, *Journal of Business*, 1991, 64, 1–19.

Deimling, K., *Nonlinear Functional Analysis*, Springer-Verlag, 1980.

Detemple, J., and L. Selden, A general equilibrium analysis of option and stock market interactions, *International Economic Review*, 1991, 32, 279–303.

Duffie, D., *Dynamic Asset Pricing Theory*, 2nd ed., Princeton University Press, 1996.

Eckwert, B., The risk and price volatility of real assets in an overlapping generations economy with fiat money, *Finance*, 1992, 13, 7–21.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Market selection of financial trading strategies: Global stability, *Mathematical Finance*, 2002, 12, 329–339.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Evolutionary stable stock markets, *Economic Theory*, 2006, 27, 449–468.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Globally evolutionarily stable portfolio rules, *Journal of Economic Theory*, 2008, 140, 192–228.

Föllmer, H., U. Horst, and A. P. Kirman, Equilibria in financial markets with heterogeneous agents: A probabilistic perspective, *Journal of Mathematical Economics*, 2005, 41, 123–155.

Föllmer, H., and M. Schweizer, A microeconomic approach to diffusion models for stock prices, *Mathematical Finance*, 1993, 3, 1–23.

Freixas, X., and J.-C. Rochet, *Microeconomics of Banking*, 5th ed., MIT Press, 1997.

Friedman, M., *Essays in Positive Economics*, University of Chicago Press, 1953.

Gale, D., and H. Nikaido, The Jacobian matrix and global univalence of mappings, *Mathematische Annalen*, 1965, 159, 68–80.

Gohberg, I., P. Lancaster, and L. Rodman, *Matrix Polynomials*, Academic Press, 1982.

Grandmont, J.-M., Temporary general equilibrium theory, in *Handbook of Mathematical Economics*, K. Arrow and M. Intrilligator (eds.), North-Holland Publishing Co. 1982.

Hens, T., J. Laitenberger, and A. Löffler, Two remarks on the uniqueness of equilibria in the CAPM, *Journal of Mathematical Economics*, 2002, 37, 123–132.

Hillebrand, M., *Pension Systems, Demographic Change, and the Stock Market*, Vol. 610 of *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 2008.

Hillebrand, M., and J. Wenzelburger, The impact of multiperiod planning horizons on portfolios and asset prices in a dynamic CAPM, *Journal of Mathematical Economics*, 2006a, 42, 565–593.

Hillebrand M., and J. Wenzelburger, On the dynamics of asset prices and portfolios in a multiperiod CAPM, *Chaos, Solitons and Fractals*, 2006b, 29, 578–594.

Horst, U., Financial price fluctuations in a stock market model with many interacting agents, *Economic Theory*, 2005, 25, 917–932.

Horst, U., and C. Rothe, Queuing, social interactions and the microstructure of financial markets, *Macroeconomic Dynamics*, 2008, 12, 211–233.

Horst, U., and J. Scheinkman, Equilibria in systems of social interaction, *Journal of Economic Theory*, 2006, 27, 44–77.

Horst, U., and J. Scheinkman, A limit theorem for systems of social interaction, *Journal of Mathematical Economics*, in press, 2008.

Horst, U., and J. Wenzelburger, On non-ergodic asset prices, *Economic Theory*, 2008, 34, 207–234.

Huberman, G., Capital asset pricing in an overlapping generations model, *Journal of Economic Theory*, 1984, 33, 232–248.

Huffman, G., Adjustment cost and capital asset pricing, *Journal of Finance*, 1985, 11, 691–709.

Huffman, G., Asset pricing with capital accumulation, *International Economic Review*, 1986, 27, 565–582.

Judd, K. L., F. Kübler, and K. H. Schmedders, Asset trading volume with dynamically complete markets and heterogeneous agents, *Journal of Finance*, 2003, 58, 2203–2217.

Kübler, F., and H. Polemarchakis, Stationary Markov equilibria for overlapping generations, *Economic Theory*, 2004, 24, 623–643.

Kübler, F., and K. Schmedders, Stationary equilibria in asset-pricing models with incomplete markets and collateral, *Econometrica*, 2003, 71, 1767–1793.

Kübler, F., and K. Schmedders, Approximate versus exact equilibria in dynamic economies, *Econometrica*, 2005, 73, 1205–1235.

Kushner, H. J., and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications: Stochastic Modelling and Applied Probability*, 2nd ed., Springer-Verlag, 2003.

Lajeri, F., and L. T. Nielsen, Parametric characterizations of risk aversion and prudence, *Economic Theory*, 2000, 15, 469–476.

LeRoy, S., and J. Werner, *Principles of Financial Economics*, Cambridge University Press, 2001.

Lintner, J., The valuation of risky assets and selection of risky investment in stock portfolios and capital budgets, *Review of Economics and Statistics*, 1965, 47, 13–37.

Lucas, R. E., Asset prices in an exchange economy, *Econometrica*, 1978, 46, 1426–1446.

Luenberger, D. G., *Investment Science*, Oxford University Press, 1998.

Magill, M., and M. Quinzii, Incomplete markets over an infinite horizon: Long-lived securities and speculative bubbles, *Journal of Mathematical Economics*, 1996, 26, 133–170.

Magill, M., and M. Quinzii, Infinite horizon CAPM equilibrium, *Economic Theory*, 2000, 15, 103–138.

Markowitz, H., Portfolio selection, *Journal of Finance*, 1952, 7, 77–91.

Mehra, R., and E. C. Prescott, Recursive competitive equilibrium: The case of homogenous households, *Econometrica*, 1980, 48, 1365–1379.

Mehra R., and E. C. Prescott, The equity premium: A puzzle, *Journal of Monetary Economics*, 1985, 15, 145–161.

Merton, R. C., An analytic derivation of the efficient portfolio frontier, *Journal of Financial and Quantitative Analysis*, 1972, 7, 1851–1872.

Merton R. C., An intertemporal capital asset pricing model, *Econometrica*, 1973, 41, 867–888.

Mossin, J., Equilibrium in a capital asset market, *Econometrica*, 1966, 34, 268–290.

Nielsen, L. T., Portfolio selection in the mean-variance model: A note, *Journal of Finance*, 1987, 42, 1371–1376.

Nielsen, L. T., Uniqueness of equilibrium in the classical capital asset pricing model, *Journal of Financial and Quantitative Analysis*, 1988, 23, 329–336.

Nielsen, L. T., Existence of equilibrium in CAPM, *Journal of Economic Theory*, 1990a, 52, 223–231.

Nielsen, L. T., Existence of equilibrium in CAPM without a riskless asset, *Review of Economic Studies*, 1990b, 57, 315–324.

Orosel, G. O., Informational efficiency and welfare in the stock market, *European Economic Review*, 1996, 40, 1379–1411.

Orosel, G. O., Market participation and share prices, *Mathematical Finance*, 1997, 7, 375–398.

Orosel, G. O., Participation costs, trend chasing, and volatility of stock prices, *Review of Financial Studies*, 1998, 11, 521–557.

Owen, J., and R. Rabinovitch, On the class of elliptical distributions and their applications to the theory of portfolio choice, *Journal of Finance*, 1983, 38, 745–752.

Pyle, D., On the theory of financial intermediation, *Journal of Finance*, 1971, 26, 737–747.

Rochet, J.-C., Capital requirements and the behaviour of commercial banks, *European Economic Review*, 1992, 36, 1137–1170.

Rockafellar, T. R., *Convex Analysis*, Princeton University Press, 1970.

Sandroni, A., Do markets favor agents able to make accurate predictions?, *Econometrica*, 2000, 68, 1303–1341.

Sargent, T. J., *Bounded Rationality in Macroeconomics*, Clarendon Press, 1993.

Sharpe, W. F., Capital asset pricing: A theory of market equilibrium under conditions of risk, *Journal of Finance*, 1964, 19, 425–442.

Simon, H. A., *Models of Bounded Rationality*, MIT Press, 1982.

Stapleton, R., and M. Subrahmanyam, A multiperiod equilibrium asset pricing model, *Econometrica*, 1978, 46, 1077–1096.

Tobin, J., Liquidity preference as behavior towards risk, *Review of Economic Studies*, 1958, 25, 65–86.

Wenzelburger, J., Learning to predict rationally when beliefs are heterogeneous, *Journal of Economic Dynamics and Control*, 2004, 28, 2075–2104.

Wenzelburger, J., *Learning in Economic Systems with Expectations Feedback*, Vol. 555, *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 2006.

Wenzelburger, J., A note on the two-fund separation theorem, Keele Economic Research Papers 1, Keele University, 2008.

CHAPTER 7

Market Selection and Asset Pricing

Lawrence Blume
Cornell University, IHS, and Santa Fe Institute

David Easley
Department of Economics
Cornell University

7.1. Introduction 405
7.1.1. Evolution in Biology and Economics 405
7.1.2. A Short History of the Market Selection Hypothesis 407
7.1.3. Scope of This Chapter 409
7.2. The Economy 411
7.2.1. Traders 411
7.2.2. Beliefs 412
7.3. Equilibrium Allocations and Prices 413
7.3.1. Pareto Optimality 413
7.3.2. Competitive Equilibrium 414
7.4. Selection 416
7.4.1. Literature 416
7.4.2. A Leading Example 416
7.4.3. Selection in Complete IID Markets 419
7.4.4. The Basic Equations 420
7.4.5. Who Survives? Necessity 421
7.4.6. Selection and Market Equilibrium 421
7.4.7. More General Stochastic Processes 422
7.5. Multiple Survivors 423
7.5.1. Who Survives? Sufficiency 423
7.6. The Life and Death of Noise Traders 426
7.6.1. The Importance of Market Structure 426
7.6.2. Laws of Large Numbers 428

Note: The authors are grateful to the NYU Department of Economics and to the Santa Fe Institute, where some of the research surveyed here was undertaken, and to the National Science Foundation for research support under grant ITR-0325453.

7.7. Robustness 431
7.7.1. Unbounded Economies 431
7.7.2. Incomplete Markets 432
7.7.3. Differential Information 432
7.7.4. Selection over Non-EU Traders 432
7.7.5. Selection over Rules 434
7.8. Conclusion 435
References 436

Abstract

This chapter surveys asset pricing in dynamic economies with heterogeneous, rational traders. By *rational* we mean traders whose decisions can be described by preference maximization, where preferences are restricted to those that have an subjective expected utility (SEU) representation. By *heterogeneous* we mean SEU traders with different and distinct payoff functions, discount factors, and beliefs about future prices that are not necessarily correct. We examine whether the market favors traders with particular characteristics through the redistribution of wealth and the implications of wealth redistribution for asset pricing.

The arguments we discuss on the issues of market selection and asset pricing in this somewhat limited domain have a broader applicability. We discuss selection dynamics on Gilboa-Schmeidler preferences and on arbitrarily specified investment and savings rules to see what discipline, if any, the market wealth redistribution dynamic brings to this environment. We also clarify the relationship between the competing claims of the market selection analysis and the noise-trader literature.

Keywords: market selection, subjective expected utility, rational traders, Bayesian learning, heterogeneous consumers, asset pricing

7.1. INTRODUCTION

In this chapter we survey asset pricing in dynamic economies with heterogeneous, rational traders. *Rational* traders' decisions can be described by preference maximization, where preferences are restricted to those that have an SEU representation. However, the arguments we bring to bear on the issues of market selection and asset pricing in this somewhat limited domain have a broader applicability, on which we will briefly touch.

7.1.1. Evolution in Biology and Economics

Evolutionary finance is one of the main themes of this book. Our research is not an attempt to bring models from evolutionary biology to bear on market phenomena. It is distinctly economic rather than biological. At the outset (which for us was 1981, when we first began to think about these issues) we observed that the invisible hand works by steering resource flows away from some choice behaviors and toward others. This makes a nice analogy with natural selection in natural, that is, biological, systems, and so we coined the phrase *market selection hypothesis* to frame the question: Do markets redirect resources toward "rational" decision makers? And though it is convenient to exploit the analogy further by using such terms as *fitness* or *survival index* (as we do here), our models are distinctly not biological. The metaphor becomes clouded when we ask after the analogues of species, genes, and other objects of the evolutionary model landscape. It is simply opaque when we look for the analogue of the genetic transmission mechanism.

The eminent biologist George C. Williams offers what amounts to a critique of excessively biological thinking when considering cultural evolution (1992, Chapter 2). Cultural evolution involves the proliferation of packets of information, *codices* in Williams' terminology. Biological analogies for evolution, he suggests, are more likely to be found in epidemiology than in population genetics, which he defines as "that branch of epidemiology that deals with infectious elements transmitted exclusively from parent to offspring." All but explicit in his discussion is the idea that proliferation is a process of direct social contact, which economists call *social learning*. However, markets provide other transmission mechanisms that make the population genetics metaphor even less successful.

We have chosen to ignore information sharing and instead to uncover the mechanism of wealth reallocation through the market toward decision rules that generate higher return. We do this in part because the implications of wealth redistribution stand out most clearly when they are studied in isolation, but this choice is not simply a modeling tactic. First, as an empirical fact, we note that social learning often happens on time scales different than the transacting time scale. In small, slow markets, such as the Ithaca, New York, housing market, transactions occur at a slow rate, and the rate of social learning may exceed the transaction rate. In this case we expect the long-run market behavior to be driven in large part by the kinds of epidemiological behavior to which Williams alludes. In well-developed financial markets, learning may take place at a rate much slower than the transaction rate, and so we expect wealth-share dynamics to be the principle driver.

A more important reason for ignoring social learning in markets is that at this point we believe that there is little to say. There are two modes of exchanging asymmetric information in markets: learning from prices, the subject of rational expectations; and learning directly from others, social learning. Blume and Easley (2006) and Sandroni (2000) ask how traders with rational expectations fare against traders with other beliefs. This is an obvious question but also an odd one. To understand why, consider what a rational expectations inference rule must be in an economy with heterogeneous traders and state variables other than beliefs. The rational expectations inference rule will condition on the state variable as well as on prices. Alternatively, it could use the entire past history of prices to forecast the current state. The rational expectations traders we and Sandroni theorize about hold inference rules for state distributions that depend on prices and the distribution of wealth. Rational expectations traders need either to observe wealth or have some way of forecasting the wealth distribution from private information and the past history of prices. This is a bit of a stretch. One could ask a weaker question. Suppose rational expectations traders were those who held price-state inference rules that would be correct were they the only traders in the market but were not necessarily correct for other wealth distributions where traders with other inference rules had significant sway. Would the other traders be driven out? Can wealth redistribution through trading drive the economy to a rational expectations equilibrium(REE)? We do not know the answer to this question, but in 1982 we asked a similar question about learning dynamics. Then we learned that rational expectations were locally stable under learning dynamics, but not globally stable, and that the limit behavior of the economy could be very different from the predictions of the REE. We do not expect the answer to be different here.

Another tack is to propose a population of traders with asymmetric information and different rules and ask which rules and which information sets are selected for. This path is taken by Mailath and Sandroni (2003), who build a model containing traders who receive signals about the true state of the world. They ask a simple question: Will better-informed traders drive out those less well informed? They construct a model in which equilibrium is a sequence of partially revealing rational expectations equilibria to answer this question. This is a very difficult task, and the best they (or anyone else so far) can do is to construct a very elaborate example. The end result is that, under some conditions, better-informed traders drive out less well-informed traders. This is a good first step; it is also the state of the art. We look forward to future progress on this question. However, at the moment there is simply little to say about market selection in environments with asymmetric information.

The literature that melds social learning to market selection is in much worse condition. A typical paper will build a single-period market model wherein one's gain depends on one's trading rule and the rule of others, and drive the whole thing by replicator dynamics. Why replicator dynamics, one might ask? At the very best it is a simple special case, perhaps having something to do with imitation under some very special and not very likely assumptions about what is observable; perhaps more likely, it is chosen simply because it is there and has a tradition in game theory. We believe, along with Williams, that models of social learning should be tuned to the social processes at work. On the other hand, we see great promise for social learning models that are empirically well founded and that are applied to economic environments in which physical stocks or wealth are state variables along with beliefs.

7.1.2. A Short History of the Market Selection Hypothesis

The link between economic forces and natural selection was established at the outset of the Darwinian revolution. Darwin claimed inspiration from Malthus's *Essay on Population* (see Barlow, 1969). Perhaps surprisingly, however, the link was only one way for more than a century. Although Marshall (1961, p. 772), claimed that economics "is a branch of biology broadly interpreted," his analytical work shows scant evidence of biological thinking. Though we can find traces of evolutionary reasoning in Knight and Schumpeter, natural selection did not emerge as an analytical argument until after the war.[1]

[1] See Schumpeter (1934) and the rather odd paper of Knight (1923).

The first sustained application of Darwinian arguments in economics was to the theory of the firm.[2] Critics of the neoclassical theory of the firm argued that managers did not have the information or tools to undertake the marginal calculations profit-maximization requires. In response, Chicago economists argued that although firms' behaviors need not be the product of an optimization exercise, market forces favored firms that acted most like profit maximizers. Thus, wrote (Enke 1951), "In these instances the economist can make aggregate predictions *as if* each and every firm knew how to secure maximum long-run profits."[3] This view is most notably associated with Friedman (1953, p. 22):

> Whenever this determinant (of business behavior) happens to lead to behavior consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not the business will tend to lose resources and can be kept in existence only by the addition of resources from the outside. The process of natural selection thus helps to validate the hypothesis (of profit maximization) or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.

The intuition offered by Alchian, Enke, and Friedman is that eventually capital reallocation will drive out firms that do not maximize profits. The first claim comes in two varieties. One is that nonmaximizing firms will make losses and be driven out of the market because they cannot continue to fund their operations out of retained earnings and will not attract investors. Winter (1971) attempts to make the retained-earnings argument precise by modeling explicitly the a process of innovation, growth, and the death of firms.

This argument is not without its critics. Koopmans (1957, p. 140) argues that this is bad modeling strategy. "But if this [natural selection] is the basis for our belief in profit maximization, then we should postulate that basis itself and not the profit maximization which it implies in certain circumstances." Blume and Easley (2002) build a dynamic equilibrium model wherein firms invest from retained earnings and show that although only profit-maximizing firms survive, the long-run state of the evolutionary process is inefficient, so that the conclusions of static welfare analysis are wrong when applied to the stable steady state. Dutta and Radner (1999) challenge even the survival of profit maximizers. In a world of uncertainty, they take the normative firm behavior rule to be one of maximizing the expected discounted sum of dividends. They show that firms pursuing this strategy will almost surely fail in finite time, whereas other rules offer a positive probability of long-run survival. These long-run survivors are able

[2]Samuelson (1985, p. 166) writes, "Reactions within economics against highfalutin borrowings of the methodology of mathematical physics led Alfred Marshall and a host of later writers to hanker for a 'biological' approach to political economy." In contrast, the evolutionary approaches we are about to describe are simply models of how markets allocate resources among economic actors with diverse tastes, goals, and behaviors. Interest in these models is independent of their biological connotations.

[3]Enke (1951, p. 567), italics in the original. See also Alchian (1950), whose more nuanced view referenced the problem that profit maximization is ill-defined in a world of uncertainty, and that it was not one's absolute profits but relative profitability that determines long-run survival.

to attract investment funds, and so, if new entrants' behaviors are sufficiently diverse, *"after a long time practically all of the surviving firms will not have been maximizing profits"* (1999, p. 769; italics in original).

Natural selection arguments have also been offered as a rationale for the (informationally) efficient pricing of financial assets. It has long been believed, without any apparent justification, that in dynamic economies in which traders have heterogeneous beliefs, structure on asset prices arises in the long run from evolutionary forces akin to natural selection across traders in markets. Fama (1965, p. 38), for instance, argues that

> ... dependence in the noise generating process would tend to produce "bubbles" in the price series ... If there are many sophisticated traders in the market, however, they will be able to recognize situations where the price of a common stock is beginning to run up above its intrinsic value. If there are enough of these sophisticated traders, they may tend to prevent these "bubbles" from ever occurring.

According to Fama (p. 40), "A superior analyst is one whose gains over many periods of time are *consistently* greater than those of the market." We call this the *market selection* argument for the informational efficiency of prices. Cootner (1964) was an early, clear proponent of this argument: "Given the uncertainty of the real world, the many actual and virtual traders will have many, perhaps equally many, forecast ... If any group of traders was consistently better than average in forecasting stock prices, they would accumulate wealth and give their forecasts greater and greater weight. In this process, they would bring the present price closer to the true value."

A contrarian view is offered by the so-called *noise-trader* literature. The term *noise trader* was first used in print by Kyle (1985) to refer to uninformed traders who traded randomly.[4] Black (1986) uses the term to refer to traders whose trades are based on uninformative signals—noise—and argues for their importance to financial markets. Two related questions are: Can noise traders influence prices, and can noise traders survive? In two surprisingly influential papers, DeLong et al. (1990, 1991) address this question. The first paper builds an overlapping generations equilibrium model with noise and rational traders to show that noise traders can both affect prices and receive higher expected returns than do the rational traders. The survival question cannot be asked in this model because the supply of noise traders is exogenously fixed. High expected returns say nothing about the possibility of survival. Of course, the authors recognize this, and so the second paper attempts to describe "the survival of noise traders in financial markets." We discuss their effort in Section 7.6 and explain its relation to our own research (Blume and Easley 1992, Blume and Easley 2006).

7.1.3. Scope of This Chapter

We set the stage for our investigation of dynamics by first delineating the restrictions on asset prices that arise from rationality alone. Rationality has a variety of

[4] Kyle attributes the phrase to Sanford Grossman; see Dow and Gorton (2008).

meanings in the literature. Here we take rationality to mean that traders' preferences satisfy the Savage (1951) axioms. Thus traders are subjective expected utility maximizers. The existence of a SEU representation implies that each trader is a Bayesian, but it does not otherwise restrict traders' beliefs. In particular, SEU rationality does not imply that traders have correct beliefs. The restriction to SEU rationality is unfortunate. The dominance of SEU rationality in economics today is somewhat of a historical anomaly. The explosion of research in decision theory over the past 20 years has provided us with a rich class of plausible preferences and choice functions, and the classification of theoretical economic findings into the continuum spanned by robust insights versus artifacts of expected utility is a compelling research program that is proceeding at best fitfully. Not surprisingly, the literature on survival in asset markets is largely dominated by SEU decision makers, although Blume and Easley (1992) treat decision rules directly. Section 7.7.4 describes Condie's (2008) recent work on selection with maximin expected utility traders.

The economies we analyze have traders who live forever and discount streams of SEU payoffs (except in Sections 7.7.4 and 7.7.5), who have stochastic endowments of a single consumption good and trade a complete set of Arrow securities in each period. We do not explicitly consider richer sets of securities, but more complex assets can be priced by arbitrage from the prices of the Arrow securities. For these economies, any map from partial histories of the economy into prices of the Arrow securities with imputed interest rates that assign finite present discounted values to constant wealth streams is consistent with an equilibrium in which all traders are subjective expected utility maximizers. So structure on asset prices comes not from the hypothesis of SEU maximization but from restrictions on traders beliefs and discount factors. These restrictions are ancillary to the hypothesis of trader rationality. One such belief restriction is that of rational expectations—the assumption that beliefs are "correct." This belief restriction constrains the set of SEU preferences that can appear in the market, and since it imposes restrictions on market observables, it requires both theoretical and empirical justification.

It is important for the rational expectations equilibrium conclusions that belief restrictions are satisfied by all traders who influence prices. This is surely problematic. How is it that all these traders know the truth or even place positive probability on it? Where does this knowledge or prior restriction come from? It cannot be derived from learning, as the rationality model with a prior restriction is itself supposed to be a model of the learning process. What happens if more realistically we assume that initially some, but not all, traders know the truth or are able to learn it? Are there forces that provide structure to long-run asset prices? This requires an analysis of a dynamic economy with heterogeneous traders.

The market selection argument sounds compelling, but until recently there has been no formal investigation of its validity. In this article we describe some recent results on market selection. These results address both the long-run composition of the trader pool and the consequences of selection for asset prices in the long run. We begin by describing a class of infinite-horizon heterogeneous trader economies.

7.2. THE ECONOMY

We suppose that a finite number of infinitely lived agents trade claims on a single consumption good across dates and states of nature. Time is discrete and is indexed by $t \in \{0, 1 \ldots, \infty\}$. The possible states at each date form a finite set $S = \{1, \ldots, \mathbf{S}\}$, with cardinality $\mathbf{S} = |S|$. A *path* is an infinite sequence of states, one for each date. The set of all paths is denoted by Σ, with members $\sigma = (\sigma_0, \ldots)$. The value of path σ at date t is denoted σ_t. The *partial history* through date t of σ is $\sigma^t = (\sigma_0, \ldots, \sigma_t)$, and H_t denotes the set of all partial histories through date t.

The set Σ together with its product sigma-field $\mathcal{F}$ is the measurable space on which everything will be built. Let p denote the "true" probability measure on Σ. Expectation operators without subscripts intend the expectation to be taken with respect to the measure p. For any probability measure q on Σ, $q_t(\sigma)$ is the (marginal) probability of the partial history $(\sigma_0, \ldots, \sigma_t)$. That is, $q_t(\sigma) = q(\{\sigma_0 \times \cdots \times \sigma_t\} \times S \times S \times \cdots)$.

In the next few paragraphs we introduce a number of stochastic processes $x : \Sigma \to \prod_{t=0}^{\infty} \mathbf{R}_{++}$. These are sequences of random variables such that each x_t is date-t measurable; that is, its value depends only on the realization of states through date t. Formally, $\mathcal{F}_t$ is the σ-field of events measurable through date t, and each $x_t(\sigma)$ is assumed to be $\mathcal{F}_t$-measurable. The assumption of $\mathcal{F}_t$-measurability means that we can take x_t to be defined on H_t, and we will sometimes write $x_t(\sigma^t)$ to emphasize this.

7.2.1. Traders

An economy contains I traders, each with consumption set $\mathbf{R}_{++}$. A *consumption plan* is a stochastic process $c : \Sigma \to \prod_{t=0}^{\infty} \mathbf{R}_{++}$. Each trader is endowed with a particular consumption plan e^i, called the *endowment stream.* The *continuation plan* at t for the consumption plan c is the process $(c_{t+1}, c_{t+2}, \ldots)$.

We assume that each trader's preferences over consumption plans satisfy the Savage (1951) axioms and thus each trader is a subjective expected utility maximizer. We also assume that each traders' payoff function is time separable and exhibits geometric discounting. Specifically, trader i has beliefs about the evolution of states, which are represented by a probability distribution p^i on Σ. She also has a payoff function $u_i : \mathbf{R}_{++} \to \mathbf{R}$ on consumptions and a discount factor β_i strictly between 0 and 1, so the utility of a consumption plan is

$$U_i(c) = E_{p^i}\Big\{\sum_{t=0}^{\infty} \beta_i^t u_i\big(c_t(\sigma)\big)\Big\} \qquad \textbf{(7.1)}$$

We will assume throughout the following properties of payoff functions:

A.7.1. *The payoff functions u_i are C^1, strictly concave, and strictly monotonic and satisfy an Inada condition at 0.*

We also assume that endowments are strictly positive and that the aggregate endowment is uniformly bounded. Let $e_t(\sigma) = \sum_i e_t^i(\sigma)$ denote the aggregate endowment at date t on path σ.

A.7.2. *For all traders i, all dates t and all paths* σ, $e_t^i(\sigma) > 0$. *Furthermore, there are numbers* $F \geq f > 0$ *such that* $f \leq \inf_{t,\sigma} e_t(\sigma) \leq \sup_{t,\sigma} e_t(\sigma) < F < \infty$.

The bounds are important to the derivation of our results. Our conclusions hold when F grows slowly enough or f converges to 0 slowly enough but may fail when F grows too quickly or f converges to 0 too quickly. We will explore this idea in Section 7.7.1.

The following assumption about beliefs will be maintained for convenience throughout this discussion. Any trader who violates this axiom would not survive, so there is no cost to discarding them now.

A.7.3. *For all traders i, all dates t and all paths* σ, $p_t(\sigma) > 0$ *and* $p_t^i(\sigma^t) > 0$.

7.2.2. Beliefs

Our assumption that traders are subjective expected utility maximizers provides beliefs p over paths. The representation places no restrictions on beliefs other than the obvious requirement that they are a probability on $(\Sigma, \mathcal{F})$. One important special case is that of beliefs generated by IID forecasts. If trader i believes that all the σ_t are IID draws from a common distribution ρ, then p^i is the corresponding distribution on infinite sequences. In this case, $p_t^i(\sigma) = \prod_{\tau=0}^{t} \rho^i(\sigma_\tau)$. Our belief framework also allows for Bayesian learners confronting model uncertainty. Suppose that models are parametrized by $\theta \in \Theta$; that is, corresponding to every $\theta \in \Theta$ is a probability distribution p^θ on $(\Sigma, \mathcal{F})$. Suppose that $\mathcal{T}$ is a σ-field on Θ with respect to which the map $\theta \to p^\theta(A)$ is measurable for all $A \in \mathcal{F}$. Finally, suppose our trader has prior beliefs on models represented by a probability distribution μ on $(\Theta, \mathcal{T})$. Then $p(A) = \int p^\theta(A) d\mu$, the predictive distribution, is the belief that a Bayesian trader would use to evaluate a consumption plan.

SEU maximizers are, by definition, Bayesian learners. Traders revise their beliefs about future values of σ_t in light of what they have already seen. To be clear on this, write the expected utility of a plan c as

$$\begin{aligned} U_i(c) &= E_{p^i}\left\{ \sum_{t=0}^{\infty} \beta_i^t u_i\big(c_t(\sigma)\big) \right\} \\ &= E_{p^i}\left\{ \sum_{t=0}^{T} \beta_i^t u_i\big(c_t(\sigma)\big) + \sum_{t=T+1}^{\infty} \beta_i^t u_i\big(c_t(\sigma)\big) \right\} \\ &= E_{p_T^i}\left\{ \sum_{t=0}^{T} \beta_i^t u_i\big(c_t(\sigma)\big) \right\} + E_{p^i|\sigma^T}\left\{ \sum_{t=T+1}^{\infty} \beta_i^t u_i\big(c_t(\sigma)\big) \right\} \end{aligned}$$

Thus continuation plans at $t+1$ given partial history σ^t are evaluated according to the conditional beliefs $p^i(\cdot|\sigma^t)$. We denote by $p^i(s|\sigma^t)$ the conditional probability $p^i(\sigma_{t+1} = s|\sigma^t)$.

Although SEU traders must act as though they update their beliefs about the future given the past using Bayes rule; this is not in fact restrictive. Suppose that a trader has some initial distribution of beliefs $\tilde{p}_0$ about σ_0. Suppose too that the trader's beliefs on states at Time 1 conditional on the realization of the Time 0 state, σ_0, are given by a learning rule $\tilde{p}_1(\sigma_0, \cdot)$. Similarly, for each partial history σ^t, the trader's beliefs on states at Time $t+1$ are given by a learning rule $\tilde{p}_{t+1}(\sigma^t, \cdot)$. A trader who follows this procedure uses a *belief-based learning rule*.

It follows immediately from the Kolmogorov Extension Theorem that any belief-based learning rule can be represented by a belief.

Theorem 7.1. *If $\{\tilde{p}_t\}_{t=0}^{\infty}$ is a belief-based learning rule, there is a subjective belief p on Σ such that*

1. *For all $A \subset S$, $\tilde{p}_0(A) = p(\sigma_0 \in A)$, and*
2. *For all partial histories σ^t and $A \subset S$, $\tilde{p}_{t+1}(\sigma^t, A) = p(\sigma_{t+1} \in A | \sigma^t)$.*

Thus requiring a trader to satisfy the Savage axioms places no restrictions on his sequence of one-period forecasts. Restrictions on these forecasts are typically obtained by placing restrictions on some set of models for the stochastic process that the individual considers and by restricting his prior on the model set. It is worth emphasizing that even if we can observe a trader's entire sequence of one-period-ahead forecasts, observations that contradict Bayesian behavior, and thus a subjective expected utility representation, are not possible unless the observer has some prior knowledge about the trader's beliefs.

7.3. EQUILIBRIUM ALLOCATIONS AND PRICES

We characterize equilibrium allocations and prices by examining Pareto optimal consumption paths and the prices that support them. The first welfare theorem applies to the economies we study, so every competitive path is Pareto optimal. Thus any property of all optimal paths is a property of any competitive path.

7.3.1. Pareto Optimality

Standard arguments show that in this economy, Pareto optimal consumption allocations can be characterized as maxima of weighted-average social welfare functions. If $c^* = (c^{1*}, \ldots, c^{I*})$ is a Pareto optimal allocation of resources, then there is a nonnegative vector of welfare weights $(\lambda^1, \ldots, \lambda^I) \neq 0$ such that c^* solves the problem

$$\begin{aligned} \max_{(c^1,\ldots,c^I)} \quad & \sum_i \lambda^i U_i(c^i) \\ \text{such that} \quad & \sum_i c^i - e \leq 0 \\ & c_t^i(\sigma) \geq 0 \; \forall t, \sigma \end{aligned} \tag{7.2}$$

where $e_t = \sum_i e_t^i$. The first-order conditions for the optimization problem (7.2) are: For all t there is a positive $\mathcal{F}_t$-measurable random variable η_t such that

$$\lambda^i \beta_i^t u_i'\big(c_t^i(\sigma)\big) p_t^i(\sigma) - \eta_t(\sigma) = 0 \tag{7.3}$$

almost surely, and

$$\sum_i c_t^i(\sigma) = e_t(\sigma) \tag{7.4}$$

These equations will be used to characterize the long-run behavior of consumption plans for individuals with different utility functions, discount factors, and beliefs.

7.3.2. Competitive Equilibrium

A price system is a price for consumption in each state at each date such that the value of each trader's endowment is finite.

Definition 7.1. An $\mathbf{R}_{++}$-valued stochastic process π is a *present-value price system* if and only if, for all traders i, $\sum_t \sum_{\sigma^t \in H_t} \pi_t(\sigma^t) \cdot e_t^i(\sigma^t) < \infty$.

As is usual, a competitive equilibrium is a price system and a consumption plan for each trader that is affordable and preference maximal on the budget set such that all the plans are mutually feasible. The existence of competitive equilibrium price systems and consumption plans is straightforward to prove (see Peleg and Yarri, 1970).

The standard interpretation of a competitive equilibrium price system is that there are complete markets in which trade of consumption plans occurs at Date 0. Then as history unfolds, the mutually feasible consumption plans are realized. An alternative interpretation is that markets are dynamically complete. According to this interpretation, a limited collection of assets is traded at each date, but the collection of assets is rich enough that traders can use them to construct competitive equilibrium consumption plans. The simplest set of assets that are sufficient are called *Arrow securities*. We assume that at each partial history σ^t and for each state s, there is an Arrow security that trades at partial history σ^t and that pays off one unit of account in partial history (σ^t, s) and zero otherwise. The price of the state s Arrow security in units of consumption at partial history σ^t is the price of consumption at partial history (σ^t, s) in terms of consumption at partial history σ^t, which is $\tilde{q}_t^s(\sigma) \equiv \pi_{t+1}(\sigma^t, s)/\pi_t(\sigma^t)$. Under our assumptions, every equilibrium present-value price system will be strictly positive (because every partial history is believed to have positive probability and because conditional preferences for consumption in each possible state are nonsatiated), and so all current value prices are well defined. We will be particularly interested in normalized current-value prices: $q_t^s(\sigma) = \tilde{q}_t^s(\sigma)/\sum_v \tilde{q}_t^v(\sigma)$.

Arrow security prices are sometimes called *state prices* because the current value price of Arrow security s at time t is the price in partial history σ^t of one unit account in state s at time $t+1$. Once we have prices for Arrow securities, all other assets can be priced by arbitrage. So, correct asset pricing reduces to correct pricing for Arrow securities.

It is not obvious what it means to price an Arrow security (or any other asset) correctly. For long-lived assets, it is often asserted that prices should equal the present discounted value of the dividend stream. But in a world in which traders' discount factors are not all identical, it is not intuitively obvious what the discount rate should be; and to say that the "correct" discount rate is the "market" discount rate is to beg the question, Is the market discount rate, after all, correct? With Arrow securities, it seems that prices should be related to the likelihood of the states. But in a market with endowment risk in which attitudes to risk are not all identical, risk premia should matter too, and again in a market in which not all traders have the same attitude to risk, it is not obvious what the correct risk premium is. So that we can meaningfully talk about correct prices, we make the following assumption:

A.7.4. *There is an $e > 0$ such that for all paths σ and dates t, $e_t(\sigma) \equiv e$.*

That is, there is no aggregate risk. The only risk in this economy is who gets what, not how much is to be gotten. The reason for this assumption is the following result:

Theorem 7.2. *Assume A.7.1. through A.7.4.*

1. *If all traders have identical beliefs p', then for all dates t and paths σ, and all s, $q_t^s(\sigma) = p'(s|\sigma^t)$.*
2. *On each path σ at each date t and for all $\epsilon > 0$ there is a $\delta > 0$ such that if $|c_t^i(\sigma) - e| < \delta$, then $||q_t(\sigma) - p^i(\cdot|\sigma^t)|| < \epsilon$.*

We know from Theorem 7.1 that the beliefs are arbitrary. So, even in the simple economy of Theorem 7.2, equilibrium Arrow security prices are arbitrary. Of course, there are restrictions on prices of securities that can be represented as bundles, over time or over states, of Arrow securities. Although these redundant securities are not present in our model, we could easily include them and price them by arbitrage. Inclusion of such securities would lead to falsifiable restrictions on security prices. But note that rejecting these restrictions would do far more than reject SEU—it would reject any decision theory in which individuals recognize and take advantage of arbitrage opportunities.

To obtain restrictions on asset prices in the simple economy of Theorem 7.2 we would need to have restrictions on traders' beliefs. A consequence of the first point is that in a rational expectations equilibrium, the Arrow securities spot prices will be $p(s|\sigma^t)$, the true probabilities of the state realizations given partial history σ^t. Thus we now know what it means for assets to be "correctly" priced. The second point asserts that when one trader is dominant in the sense that her demand is very large relative to that of the other traders, the equilibrium will primarily reflect her beliefs. The proof of both points is elementary, in the first case from a calculation and in the second from a calculation and the upper hemi-continuity of the equilibrium correspondence.

What happens if some traders know the truth, have rational expectations, and others do not? Will the rational traders drive out the incorrect ones and force prices to converge to their "correct" values? These questions are addressed in the next section.

7.4. SELECTION

By *selection* we mean the idea that markets identify those traders with the most accurate information, and the market prices come to reflect their beliefs. In this section we discuss the literature that focuses on selection over SEU traders, and we provide an example showing how selection works in complete markets economies.

7.4.1. Literature

Among the first to formally analyze the selection question were DeLong, Shleifer, Summers, and Waldmann (1990, 1991). The first paper shows in an overlapping generations model that traders with incorrect beliefs can earn higher expected returns than those earned by traders with incorrect beliefs. They do so because they take on extra risk. But survival is not determined by expected returns, and so this result says nothing about selection. The later paper argues that traders whose beliefs reflect irrational overconfidence can eventually dominate an asset market in which prices are set exogenously. But, as prices are exogenous, these traders are not really trading with each other; if they were, then were traders with incorrect beliefs to dominate the market, prices would reflect their beliefs and rational traders might be able to take advantage of them.

In an economy with complete markets and traders with a common discount factor, the market does select for traders with correct beliefs.[5] Sandroni (2000) shows that in a Lucas trees economy with some traders who have correct expectations, controlling for discount factors, all traders who survive have rational expectations. Blume and Easley (2006) show that this result holds in any Pareto optimal allocation in any bounded classical economy and thus for any complete markets equilibrium. For bounded complete markets economies there is a survival index that determines which traders survive and which traders vanish. This index depends only on discount factors, the actual stochastic process of states, and traders' beliefs about this stochastic process. Most important, for these economies, attitudes toward risk do not matter for survival.

The literature also provides various results demonstrating how the market selects among learning rules. The market selects for traders who learn the true process over those who do not learn the truth, for Bayesians with the truth in support of their prior over comparable non-Bayesians, and among Bayesians according to the dimension of the support of their prior (assuming that the truth is in the support). In the next section we illustrate how these complete market selection results work in a simple economy.

7.4.2. A Leading Example

Suppose there are two states of the world, $S = \{A, B\}$. States are IID draws, and the probability of state A at any date t is ρ. Arrow securities are traded for each state at each date, so markets are dynamically complete. Traders have logarithmic utility and have

[5] In economies with incomplete markets, the market selection hypothesis can fail to be true. See Section 7.7.2 and Blume and Easley (2006).

identical discount factors, $0 < \beta < 1$. Trader i knows that the state process is IID and believes that A will occur in any given period with probability ρ_i. This is basically just a big Cobb-Douglas economy, and equilibrium is easy to compute.

Since the processes and beliefs are IID, counts will be important. Let the number of occurrences of state s on path σ by date t be $n_t^s(\sigma^t) = |\{\tau \leq t : \sigma_\tau = s\}|$.

Let w_0^i denote the present discounted value of trader i's endowment stream, and let $w_t^i(\sigma)$ denote the amount of wealth that i transfers to partial history σ^t, measured in current units. The optimal consumption plan for trader i is to spend fraction $(1-\beta)\beta^t \rho_i^{n_t^A(\sigma)}(1-\rho_i)^{n_t^B(\sigma)}$ of w_0^i on consumption at date-event σ^t. This can be described recursively as follows: In each period, eat fraction $1-\beta$ of beginning wealth, w_t^i, and invest the residual, βw_t^i, in such a manner that the fraction α_t^i of date-t investment that is allocated to the asset that pays off in state A is ρ_i. Let q_t^A denote the price of the security that pays out 1 in state A at date t and 0 otherwise; let q_t^B denote the corresponding price for the other date-t Arrow security. Given the beginning-of-period wealth and the market price, trader i's end-of-period wealth is determined only by that period's state:

$$w_{t+1}^i(A) = \frac{\beta \rho_i w_t^i}{q_t^A}$$

$$w_{t+1}^i(B) = \frac{\beta(1-\rho_i) w_t^i}{q_t^B}$$

Each unit of Arrow security pays off 1 in its state, and the total payoff in that state must be the total wealth invested in that asset. Thus in equilibrium,

$$\sum_i \frac{\beta \rho^i w_t^i}{q_t^s} = \sum_j \beta w_t^j$$

and so the price of Arrow security s at date t is

$$q_t^A = \sum_i \rho^i \frac{w_t^i}{\sum_j w_t^j} = \sum_i \rho^i r_t^i, \text{ and}$$

$$q_t^B = \sum_i (1-\rho^i) r_t^i$$

where r_t^i is the *share* of date t wealth belonging to trader i.

That is, the price of Arrow security s at date t is the *wealth share weighted* average of beliefs. So at any date, the market price is stated by averaging traders' beliefs. Of course, there is no reason for this average to be correct, since the initial distribution of wealth was arbitrary. But the process of allocating the assets and then paying them off reallocates wealth. The distribution of wealth evolves through time, and the limit distribution of wealth determines prices in the long run. We can work this out to see how the market "learns." In this example it should be clear what "correct" asset pricing means. If all traders had rational expectations, the price of the A Arrow security at any point in the date-event tree would be ρ, and the price of the B Arrow security would be $1 - \rho$.

Let $1_A(s)$ and $1_B(s)$ denote the indicator functions on S for states A and B, respectively. Along any path σ of states,

$$w^i_{t+1}(\sigma) = \beta\left(\frac{\rho^i}{q^A_t(\sigma)}\right)^{1_A(\sigma_{t+1})}\left(\frac{1-\rho^i}{q^B_t(\sigma)}\right)^{1_B(\sigma_{t+1})} w^i_t(\sigma) \tag{7.5}$$

and so the ratio of i's wealth share to j's evolves as follows:

$$\frac{r^i_{t+1}(\sigma)}{r^j_{t+1}(\sigma)} = \left(\frac{\rho^i}{\rho^j}\right)^{1_A(\sigma_{t+1})}\left(\frac{1-\rho^i}{1-\rho^j}\right)^{1_B(\sigma_{t+1})}\frac{r^i_t(\sigma)}{r^j_t(\sigma)}$$

This evolution is more readily analyzed in its log form:

$$\log\frac{r^i_{t+1}(\sigma)}{r^j_{t+1}(\sigma)} = 1_A(\sigma_{t+1})\log\left(\frac{\rho^i}{\rho^j}\right) + 1_B(\sigma_{t+1})\log\left(\frac{1-\rho^i}{1-\rho^j}\right) + \log\frac{r^i_t(\sigma)}{r^j_t(\sigma)} \tag{7.6}$$

To understand how the market can learn, consider a Bayesian whose prior beliefs about state evolution contain I IID models in his support, $\{\rho^1, \ldots, \rho^I\}$, and let r^i_t denote the probability he assigns to model i posterior to the first $t-1$ observations. The Bayesian rule for posterior revision is exactly that of Eq. 7.6. Thus, the market is a Bayesian learner.[6] The evolution of the distribution of wealth parallels the evolution of posterior beliefs. Market prices are wealth share-weighted averages of the traders' models, and so the pricing function for assets is identical to the rule that assigns a predictive distribution on outcomes to any prior beliefs on states. In other words, the price of asset A in this example is the probability the Bayesian learner would assign to the event that the next state realization will be A.

From these observations we can draw several conclusions. If exactly one trader holds correct beliefs, in the long run his wealth share will converge to 1, and prices will converge to ρ. If several traders have correct beliefs, then the wealth share of this group of traders will converge to 1, and again prices will converge to ρ. The assets will be priced correctly in the long run. Second, if no model is correct, the posterior probability of any model whose Kullback-Leibler distance from the true distribution is not minimal converges a.s. to 0. In this example, selection cannot make the market do better than the best-informed trader. In particular, if there is a unique trader whose beliefs ρ^i are closest to the truth, prices converge in the long run to ρ^i almost surely, and so assets are mispriced.

[6] This idea is developed more fully in Blume and Easley (1993).

7.4.3. Selection in Complete IID Markets

For the remainder of this discussion we assume, as in the example of Section 7.4.2, that the true process and all traders beliefs are IID. A more general analysis appears in Section 7.4.7 and is elaborated in Blume and Easley (2006).

A.7.5. *The true probability p, and all traders beliefs p^i, are distributions on sequences of states consistent with IID draws from probability ρ for the true probability, and ρ^i for trader i's probability.*

Traders are characterized by three objects: a payoff function u_i, a discount factor β_i and a belief ρ^i. However, so long as payoff functions satisfy the Inada condition, they are irrelevant to survival. Only beliefs and discount factors matter. We would expect that discount factors matter in a straightforward way: Higher discount factors reflect a greater willingness to trade present for future consumption, and so they should favor survival. Similarly, traders will be willing to trade consumption on unlikely paths for consumption on those they think more likely. Those traders who allocate the most to the highest-probability paths have a survival advantage. This advantage can be measured by the *Kullback-Leibler* distance of beliefs from the truth, the relative entropy of ρ with respect to ρ^i:

$$I_\rho(\rho_i) = \sum_s \rho_s \log \frac{\rho_s}{\rho_s^i}$$

The Kullback-Leibler distance is not a true metric. But it is non-negative, and 0 if and only if $\rho^i = \rho$.[7] Assumption A.7.3. ensures that $I_\rho(\rho^i) < \infty$ (and this is its only role).

Our results will demonstrate several varieties of asymptotic experience for traders in IID economies. Traders can vanish, they can survive, and the survivors can be divided into those who are negligible and those who are not. Definitions are as follows:

Definition 7.2. Trader i *vanishes* on path σ if $\lim_t c_t^i(\sigma) = 0$. She *survives* on path σ if $\limsup_t c_t^i(\sigma) > 0$. A survivor i is *negligible* on path σ if for all $0 < r < 1$, $\lim_{T\to\infty}(1/T)|\{t \le T : c_t^i(\sigma) > re_t(\sigma)\}| = 0$. Otherwise she is *nonnegligible*.

In the long run, traders can either vanish or not, in which case they survive. There are two distinct modes of survival. A negligible trader is someone who consumes a given positive share of resources infinitely often, but so infrequently that the long-run fraction of time in which this happens is 0. The definitions of vanishing, surviving, and being negligible are reminiscent of transience, recurrence, and null recurrence in the theory of Markov chains.

[7]In fact, it is jointly convex in (ρ, ρ^i), but we will not need to make use of this fact.

7.4.4. The Basic Equations

Our method uses the first-order conditions for Pareto optimality to solve for the optimal consumption of each trader i in terms of the consumption of some particular trader, say Trader 1. We then use the feasibility constraint to solve for Trader 1's consumption. The fact that we can do this only implicitly is not too much of a bother.

Let $\kappa_i = \lambda_1/\lambda_i$. From Eq. 7.3 we get that

$$\frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \kappa_i \left(\frac{\beta_1}{\beta_i}\right)^t \prod_{s\in S}\left(\frac{\rho_s^1}{\rho_s^i}\right)^{n_t^s(\sigma)} \tag{7.7}$$

Sometimes it will be convenient to have this equation in its log form:

$$\log \frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \log \kappa_i + t\log\frac{\beta_1}{\beta_i} - \sum_s n_t^s(\sigma)\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right)$$

We can decompose the evolution of the ratio of marginal utilities into two pieces: the mean direction of motion and a mean-0 stochastic component.

$$\begin{aligned}\log \frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} &= \log \kappa_i + t\log\frac{\beta_1}{\beta_i} - t\sum_s \rho_s\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right)\\ &\quad - \sum_s \big(n_t^s(\sigma) - t\rho_s\big)\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right)\\ &= \log\kappa_i + t\big(\log\beta_1 - I_\rho(\rho^1)\big) - t\big(\log\beta_i - I_\rho(\rho^i)\big)\\ &\quad - \sum_s \big(n_t^s(\sigma) - t\rho_s\big)\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right)\end{aligned}$$

The mean term in the preceding equation gives a first-order characterization of traders' long-run fates.

Definition 7.3. Trader i's *survival index* is $\mathsf{s}_i = \log\beta_i - I_\rho(\rho^i)$.

Then

$$\log \frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \log\kappa_i + t(\mathsf{s}_1 - \mathsf{s}_i) - \sum_s \big(n_t^s(\sigma) - t\rho_s\big)\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right) \tag{7.8}$$

7.4.5. Who Survives? Necessity

Necessary conditions for survival have been studied before, notably by Blume and Easley (2006) and Sandroni (2000). In this economy, a sufficient condition guaranteeing that Trader i vanishes is that Trader i's survival index is not maximal among the survival index of all traders. Consequently, a necessary condition for survival is that the survival index be maximal.

Theorem 7.3. *Assume A.1–5. If* $\mathsf{s}_i < \max_j \mathsf{s}_j$, *then Trader i vanishes.*

The point of writing down ratios of marginal utilities is to compare two traders. If one trader never "loses" a comparison, he must be a survivor. The next lemma describes how ratios of marginal utilities characterize survival.

Lemma 7.1. *On the path* σ, *if* $u_i'\big(c_t^i(\sigma)\big)/u_1'\big(c_t^1(\sigma)\big) \to \infty$, *then Trader i vanishes.*

Proof. If the ratio diverges, either the numerator diverges or the denominator converges to 0. The latter event cannot happen because no trader can consume more than the aggregate endowment, and the aggregate endowments are uniformly bounded from above across periods (A.7.2). The former event implies that $c_t^i(\sigma) \to 0$. ☐

In view of Lemma 7.1, everything to know about selection is captured in the asymptotic behavior of the right side of Eq. 7.7. For the IID case, this behavior can be examined with the strong law of large numbers (SLLN) applied to the log of the left side. Divide both sides of Eq. 7.8 by t to obtain

$$\frac{1}{t}\log\frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \frac{1}{t}\log\kappa_i + (\mathsf{s}_1 - \mathsf{s}_i) - \frac{1}{t}\sum_s \big(n_t^s(\sigma) - t\rho_s\big)\left(\log\frac{\rho_s^i}{\rho_s} - \log\frac{\rho_s^1}{\rho_s}\right) \tag{7.9}$$

The first term on the right side converges to 0 and so, by the SLLN, does the last term. Thus the fate of trader is determined by her survival index.

7.4.6. Selection and Market Equilibrium

The implications for long-run asset pricing are already illustrated in the example that began this section.

Corollary 7.1. If there is a unique Trader i with maximal survival index s_i among the trader population, then market prices converge to ρ^i almost surely.

This corollary is an immediate consequence of Theorems 7.2 and 7.3. If only Trader i has maximal survival index, almost surely all other traders vanish and q_t converges to ρ^i.

The beliefs of the trader with maximal survival index may not be correct, in which case Arrow securities are incorrectly priced in the long run. This may happen because no trader has correct beliefs or because a trader's incorrect beliefs are compensated for by a higher discount factor. In the latter case, allowing for heterogeneous discount factors, the prices need not converge to the most accurate beliefs present in the market.

7.4.7. More General Stochastic Processes

The preceding analysis is actually quite general. Although the particular functional forms that emerge are specific to IID processes, the principles apply in many different situations and lead to the calculation of related survival indices. A *belief* for Trader i is a stochastic process ρ^i on the finite state space S. For general beliefs, an equation like Eq. 7.7 can be derived from Eq. 7.3.

$$\frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \kappa_i \left(\frac{\beta_1}{\beta_i}\right)^t \frac{p_t^1(\sigma^t)}{p_t^i(\sigma^t)}$$

and, in log form,

$$\log \frac{u_i'\big(c_t^i(\sigma)\big)}{u_1'\big(c_t^1(\sigma)\big)} = \log \kappa_i + t \log \frac{\beta_1}{\beta_i} + \log p_t^1(\sigma^t) - \log p_t^i(\sigma^t) \qquad \textbf{(7.10)}$$

Lemma 7.1 is still valid, of course, and so getting selection results in non-IID environments depends only on our ability to unpack and compare the log-likelihood functions of the traders' beliefs.

A simple extension arises when the true process is an irreducible Markov chain, and traders beliefs are also (not necessarily irreducible) Markov chains. Trader 1's belief term in Eq. 7.10 is $\sum_s 1_s(\sigma_t) \log p^1(s|\sigma_{t-1})$, and the term is $\sum_s 1_s(\sigma_t) \log p^i(s|\sigma_{t-1})$ for Trader i. The time average of Trader js term converges to $\sum_s \pi(s')p(s|s') \sum_s \log p(s|s')/p^j(s|s')$, where π is the invariant distribution for the Markov chain. The inner sum is $I_p(p^j|s')$ the entropy of the true distribution given state s' with respect to Trader js beliefs, and the outer sum gives the average of the conditional variances with respect to the ergodic distribution of states. The survival index for Trader i is

$$\mathbf{S}_i = \log \beta_i - E_p I_p(p^i|s')$$

The quality of beliefs is measured by the relative entropy for conditional beliefs given the past, averaged over the past according to the true ergodic distribution of states. Because the true process is irreducible, the initial state and the initial distribution of beliefs do not matter to this calculation, since the limiting time average exists and its value is independent of the initial conditions.

A more subtle example is the case of Bayesian learners. Suppose that traders are all Bayesian with identical discount factors. Each Trader i considers a set of models Θ_i

which is a bounded open subset of a d_i-dimensional Euclidean space. The true model, θ_0, is contained in each Θ_i. Suppose all the processes p^θ, $\theta \in \Theta_i$ are IID, and that the likelihood functions $p(\sigma^t|\theta)$ satisfy some regularity conditions at θ_0. Suppose, finally, that each Trader i has prior beliefs that are represented by a density q^i that is absolutely continuous with respect to Lebesgue measure on Θ_i. Let $p^i(\sigma^t) = \int_{\Theta_i} p^\theta(\sigma^t)q(\theta)d\theta$. A result well known to statisticians and econometricians is

$$\log \frac{p^\theta(\sigma^t)}{p^i(\sigma^t)} - \left(\frac{d_i}{2}\log\frac{t}{2\pi} + \frac{1}{2}\log\det I(\theta) - \log q^i(\theta)\right) \to \chi^2(d_i)$$

where $I(\theta)$ is the Fisher information matrix at θ. The result claims that the difference between the log of the probability ratio and the term in parentheses is finite, converging in probability to a chi-squared random variable.

The implication of this result for survival is that if $d_1 < d_i$, then $\log p_t^1(\sigma^t) - \log p_t^i(\sigma^t)$ diverges to $+\infty$, and so trader i-vanishes. Bayesian learners who satisfy our assumptions learn the true parameter value, but those who have to estimate more parameters, higher d_i, learn slower, and this speed difference is enough for them to be driven out of the market. We can see from the formula that the rate of divergence is $O(\log t)$, and so this speed difference would not be visible had we taken time averages.

7.5. MULTIPLE SURVIVORS

When a single trader (type) has the highest survival index, market prices converge to his view of the world. There is no room for balancing different beliefs because, in the long run, there is only one belief and discount factor present in the market. But if the market process is more complicated than the world view of any single trader so that no trader has correct beliefs, or if traders are asymmetrically informed, it is possible that multiple traders could have maximal survival index. Will all such traders survive, and what are the implications for sufficiency? We now return to the IID world for the discussion in this section.

7.5.1. Who Survives? Sufficiency

Theorem 7.3 shows that traders with survival indices that are less than maximal in the population vanish. This does not imply that all those with maximal survival indices survive. The right side of Eq. 7.8 is a random walk, and the analysis of the previous section is based on an analysis of the mean drift of the right side of Eq. 7.8. Theorem 7.3 shows that a nonzero drift has implications for the survival of some trader. When two traders with maximal survival indices are compared, the drift of the walk is 0, and further analysis of Eq. 7.7 and Eq. 7.8 is required.

We return to the example of Section 7.4.2 to study this question. However, we suppose now that there are $\mathbf{S}$ states, $S = \{1, \ldots, \mathbf{S}\}$, and that each trader has maximal survival index. From Eq. 7.5:

$$\log \frac{w^i_{t+1}}{w^1_{t+1}} = \prod_{s=1}^{\mathbf{S}} \Big(\frac{\rho^i_s}{\rho^1_s}\Big)^{1_s(\sigma_{t+1})} \frac{w^i_t}{w^1_t} = \prod_{t=1}^{t+1} \prod_{s=1}^{\mathbf{S}} \Big(\frac{\rho^i_s}{\rho^1_s}\Big)^{n^s_t(\sigma)} \frac{w^i_0}{w^1_0}$$

Taking logs,

$$\log \frac{w^i_{t+1}}{w^1_{t+1}} = \sum_{s=1}^{\mathbf{S}} \big(n^s_t(\sigma) - t\rho_s\big) \log \frac{\rho^i_s}{\rho^1_s} + t \sum_{s=1}^{\mathbf{S}} \rho_s \log \frac{\rho^i_s}{\rho^1_s} + \log \frac{w^i_0}{w^1_0}$$

The second term on the right is just the difference in survival indices, which is 0 for all survivors, so

$$\log \frac{w^i_{t+1}}{w^1_{t+1}} = \sum_{s=1}^{\mathbf{S}} \big(n^s_t(\sigma) - t\rho_s\big) \log \frac{\rho^i_s}{\rho^1_s} + \log \frac{w^i_0}{w^1_0} \tag{7.11}$$

Eq. 7.11 suggests studying the processes

$$\sum_{s=1}^{\mathbf{S}} \big(n^s_t(\sigma) - t\rho_s\big) \log \rho^i_s$$

Since the terms $\big(n^s_t(\sigma) - t\rho_s\big)$ are deviations from the drift, they sum over all states to 0, and the vector of drifts is a random walk in a space of dimension $\mathbf{S} - 1$. Normalizing with respect to state $\mathbf{S}$, define $z_t = \big(n^s_t(\sigma) - t\rho_s\big)_{s=1}^{\mathbf{S}-1}$ and for each Trader i define $\mathrm{lo}^i = (\log(\rho^i_s/\rho^i_{\mathbf{S}}))_{s=1}^{\mathbf{S}-1}$ to be a vector of log-odds ratios of Trader i's beliefs with respect to state $\mathbf{S}$. Then

$$\log \frac{w^i_{t+1}}{w^1_{t+1}} = z_t \cdot (\mathrm{lo}^i - \mathrm{lo}^1)$$

The beliefs of all surviving traders lie on the same relative entropy level set. In Figure 7.1 the true distribution lies in the center inside the iso-entropy curve.

The fate of consumer A, for instance, requires looking at the inner product of the random walk z_t with the two vectors $\mathrm{lo}^B - \mathrm{lo}^A$ and $\mathrm{lo}^C - \mathrm{lo}^A$, as shown in Figure 7.1. The shaded cone in Figure 7.2 is the polar cone to the cone spanned by $\mathrm{lo}^C - \mathrm{lo}^A$ and $\mathrm{lo}^B - \mathrm{lo}^A$. Whenever the random walk is far out in this polar cone, $z_t \cdot \mathrm{lo}^A \gg z_t \cdot \mathrm{lo}^B$ and $z_t \cdot \mathrm{lo}^A \gg z_t \cdot \mathrm{lo}^C$, so A has a large wealth share relative to B and C.

For any number of traders, those with maximal survival index must lie on the same level set of relative entropy. Thus each is extremal in the polyhedron generated by $\mathrm{lo}^1, \ldots, \mathrm{lo}^I$. From this one can show that for each Trader i there is an open cone C^i such that whenever $z_t \in C^i$, $z_t \cdot \mathrm{lo}^i > z_t \cdot \mathrm{lo}^j$ for all $j \neq i$. Far enough out in this cone, Trader i's wealth share will be arbitrarily large. One can conclude that for any Trader i with maximal survival index, $\limsup r^i_t = 1$.

When we allow for heterogeneous discount factors, the story changes. Discount factors become part of the survival index, and traders with maximal survival indices

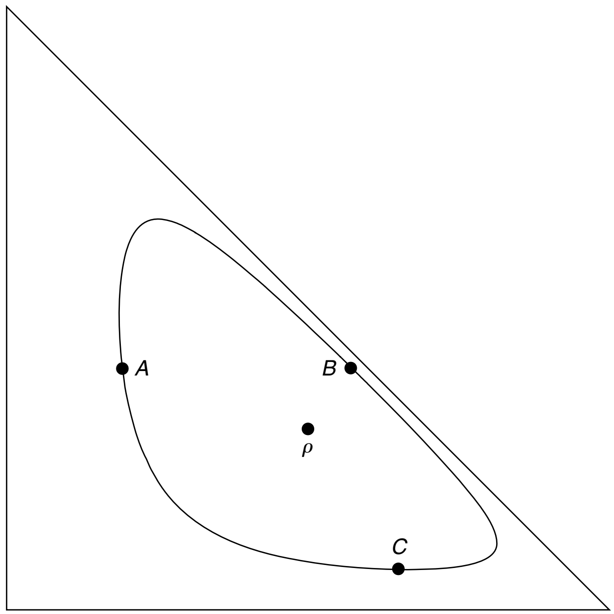

FIGURE 7.1 Three beliefs with identical survivor indices.

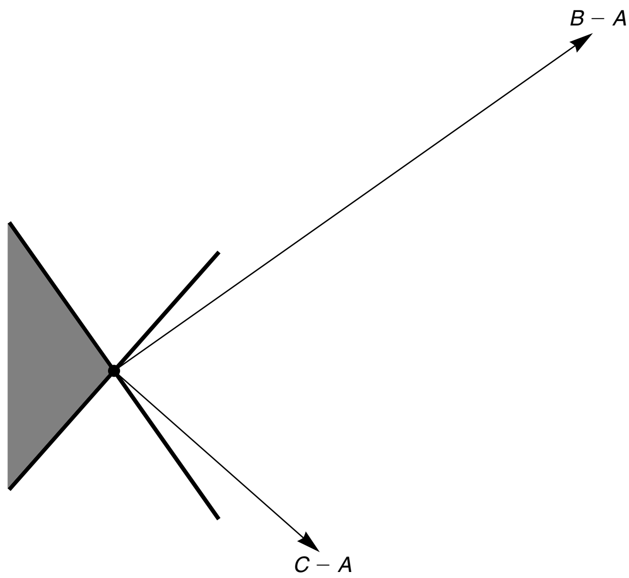

FIGURE 7.2 The cone where *A* survives.

can have beliefs whose relative entropies with respect to the truth differ. The discount factor is nonstochastic, and one can easily check that Eq. 7.11 remains unchanged. But now, since maximal traders' beliefs can be more or less accurate, it is possible that a trader could have a log-odds vector inside the polygon generated by the lo^i. Such a trader will almost surely vanish if the dimension of the walk z_t is large enough. (This requires $\mathbf{S} \geq 4$.) A sufficient condition for survival is that a trader be *extremal*, that his log-odds vector be an extreme point of the polygon. Any extremal trader's wealth share is arbitrarily near 1 infinitely often. The case of traders who are not extremal but also not interior to the polygon is more complicated. They survive, but their wealth share is bounded away from 1.

The implication for equilibrium prices from the existence of multiple survivors in our simple economy is perhaps surprising. If multiple traders have maximal survival index, then for all extremal traders i and all $\epsilon > 0$, $|q_t - \rho^i| < \epsilon$ infinitely often. If $\mathbf{S} > 3$ it is possible that for $\epsilon > 0$ sufficiently small, the event $|q_t - \rho| < \epsilon$ is transient, even if some survivor has rational expectations. Thus, with multiple survivors, asset prices are volatile. Furthermore, asset prices need not be approximately right; that is, approximately right prices may be transient.

7.6. THE LIFE AND DEATH OF NOISE TRADERS

If traders have identical discount factors and markets are complete, the results of Section 7.4 show that noise traders cannot survive. The exponential increase of their marginal utility of consumption growth measures the speed at which they disappear from the market. Yet practitioners of behavioral finance have been fascinated with the idea of noise traders, attributing to them much responsibility for deviations from rational asset pricing. In this section we examine the claims for the survival of noise traders in financial markets.

The current fascination with noise traders seems to have begun with two papers by DeLong et al. (1990, 1991). These papers raise the possibility that noise traders survive, but they do not actually prove it. The early paper observes that, in return for holding more risk than informed traders, noise traders may earn a higher expected return. They go on to show that if the size and influence of the noise-trader population is exogenously pegged, noise traders can influence prices. They point out, however, that larger returns do not imply the ability to survive when their size and influence are determined by the market (DeLong et al., 1991, p. 11). Breiman (1961) taught us that the probability of long-run survival is greatest for those who maximize not expected returns but expected log returns. The extra exposure to risk that comes from maximizing expected returns may, in the long run, be devastating.

7.6.1. The Importance of Market Structure

Many convoluted arguments have been made for the survival of noise traders, some of which (as we will see) are incorrect. But the simplest possible argument seems to have been ignored: Incomplete markets may constrain informed traders from betting against noise traders. Market structure clearly matters for the life and death of noise traders, but we have no market structure characterizations of the survival of noise traders except for the mostly negative conclusions of Sections 7.4 and 7.5. Here we demonstrate the plausibility of the incomplete markets argument by providing an example of an incomplete market economy in which the noise trader survives at the expense of the trader with correct beliefs.

Two traders buy an asset from a third trader. The two traders hold different beliefs about the certain return of the asset. Traders 1 and 3 know the correct return, which is consistently overestimated by Trader 2. We encode this example in the state preference framework by assuming that at each date there are two states: $S = \{s_1, s_2\}$. The true evolution of states has State 1 surely happening every day. There is a single asset available at each date and state that pays off in consumption goods in the next period an amount that depends on next period's state. The asset available at date t pays off, at date $t + 1$,

$$R_t(\sigma) = \begin{cases} \left(1 + \left(\frac{1}{2}\right)^t\right) & \text{if } \sigma_t = s_1 \\ 2\left(1 + \left(\frac{1}{2}\right)^t\right) & \text{if } \sigma_t = s_2 \end{cases}$$

Traders 1 and 2 have CRRA utility with coefficient $\frac{1}{2}$. Trader 3 has logarithmic utility. Traders 1 and 2 have common discount factor $(8)^{-\frac{1}{2}}$ and Trader 3 has discount factor $\frac{1}{2}$. Traders 1 and 3 believe correctly that state s_1 will always occur with Probability 1, and Trader 2 incorrectly believes that state s_2 will always occur with Probability 1. All three traders have endowments that vary with time but not state. Traders 1 and 3 know the correct price sequence. Trader 2 believes that at each date, state s_2 prices will equal the (correct) state s_1 prices. Traders 1 and 2 have endowment stream $(1, 0, 0, \ldots)$. Trader 3's endowment stream is $(0, 2, 2, \ldots)$.

This model has an equilibrium in which the price of the asset is, for every state,

$$q_t = \frac{3}{8}\left(1 + \left(\frac{2}{3}\right)^t\right)$$

In this equilibrium, at each date Trader 3 supplies 1 unit of asset and Traders 1 and 2 collectively demand 1 unit of asset. Of the total wealth belonging to Traders 1 and 2 (not Trader 3), Trader 1's share at date t is

$$\alpha_t = \frac{1}{1 + \left(\frac{3}{2}\right)^{t-1}}$$

which converges to 0 even though he has correct beliefs and Trader 2 has incorrect beliefs. Although the details of the example are complicated, the intuition is simple. At each date, Trader 1 believes that the rate of return on the asset is 2, whereas Trader 2 believes it is 3. Trader 2's excessive optimism causes him to save more at each date, so in the end he drives out Trader 1.

It is more enlightening to understand how this example was constructed than it is to go through the details of verifying the equilibrium claim. In constructing this example, our idea was to fix some facts that would allow us to solve the traders' Euler equations and then to derive parameter values that would generate those facts. Accordingly, we fixed the gross rates of return on the asset at 2 and 3 in states s_1 and s_2, respectively. We also assumed that Traders 1 and 2's total asset demand would be 1. For an arbitrary gross return sequence the Euler equations pinned down prices. We then turned to the supply side and chose an endowment stream for Trader 3 and a gross return sequence that would cause Trader 3 to supply 1 unit of asset to the market at each date.

The contrived nature of the example does not vitiate its point: that the noise trader survives because the informed trader cannot bet against him. Were, say, an Arrow security to exist at each date that pays off in only one of the two states, such bets would be possible. In this example that would pose existence problems because of the point-mass beliefs, but Blume and Easley (2006) contains a more complicated version of this example wherein the belief supports overlap, and equilibrium with the additional assets would exist. Since markets would then be complete, Trader 2 would vanish and Traders 1 and 3 would survive.

7.6.2. Laws of Large Numbers

According to DeLong, Shleifer, Summers, and Waldmann, noise traders survive in financial markets because "idiosyncratic risk reduces the survival probabilities of individual noise traders but not of noise traders as a whole..." (DeLong et al., 1991, p. 12). This argument is perhaps the most compelling of the arguments offered in the literature for the survival of noise traders. Any individual noise trader is likely to do badly, but some will, luck of the draw, do quite well, and so noise traders as a group survive. This argument is an instructive misuse of the strong law of large numbers. In this section we build a simple incomplete market model to demonstrate the problem with this type of argument.

The DeLong et al. (1991) model has traders investing in assets whose return contains a common shock and an asset-specific shock. Traders with correct beliefs can avoid the idiosyncratic risk, but not the common shock. Noise traders underestimate the variance of the asset-specific shock; consequently they take on too much risk, which earns them a higher expected return. Obviously, to make this work one needs some infinities—an infinity of assets, an infinity of traders, and so forth. We will build a much simpler model with no common shock in order to clarify ideas.

In our example there is a finite number I of traders. Trader i has an initial endowment of $w_0^i = 1$ unit of wealth, and 0 endowment subsequently. Each trader bets on IID coin flips of *his* coin, which can be either H (heads) or T (tails). For each coin, $\Pr\{s_t^i = H\} = p$. The return per unit correctly bet is 2. Each trader receives utility $u_i(c_t^i)$ from consumption. We assume that $u_i(c) = (1-\gamma)^{-1}c^{1-\gamma}$ with the coefficient of relative risk aversion $\gamma > 0$. Traders discount the future at rate β. Betting is the only way to move wealth from Date 0 through the date-event tree. At date t on the current path Trader i must choose a fraction δ_t^i to save and eat the rest. Of the wealth to be saved, fraction α_t^i is bet on H, the remainder on T. This can be recast as an investment portfolio of two Arrow securities, one which pays off on H and the other on T. One unit of each asset for each bet is inelastically supplied. The mean return and variance of i's portfolio is

$$\mathrm{E}\{w_{t+1}^i|w_t^i\} = 2\big(p\alpha_t^i + (1-p)(1-\alpha_t^i)\big)\delta_t^i w_t^i$$

and

$$\mathrm{Var}\{w_{t+1}^i|w_t^i\} = 4p(1-p)(1-2\alpha_t^i)^2(\delta_t^i w_t^i)^2$$

Notice that if $\alpha_t^i = 1/2$, the variance of tomorrow's wealth is 0 and the sure return is the amount saved.

It is not hard to compute the optimal policy. Suppose trader i believes the true probability of H is q_i. The optimal policies are independent of time and wealth:

$$\alpha^i = \frac{q_i^{1/\gamma}}{q_i^{1/\gamma} + (1-q_i)^{1/\gamma}}$$
$$\delta^i = 2^{(1-\gamma)/\gamma}\beta^{1/\gamma_i}\big(q_i^{1/\gamma} + (1-q_i)^{1/\gamma}\big)$$

In the case of log utility, corresponding to $\gamma_i = 1$, $\alpha^i = q_i$ and $\delta^i = \beta$. Computing the expected return for the optimal rules given beliefs,

$$\mathrm{E}\{w_{t+1}|w_t\} = (2\beta)^{1/\gamma}(pq^{1/\gamma} + (1-p)(1-q)^{1/\gamma})w_t$$

A computation shows that investors with more extreme beliefs have higher returns. Compare an investor with belief q to one with belief p. If $p > 1/2$ the q-investor has a higher expected return than the p-investor if and only if $q > p$; if $p < 1/2$, the q-investor has a higher expected return if and only if $q < p$.

Let $r_t = w_t^i/w_t^j$, the ratio of Trader i's wealth to Trader j's at date t. Calculations similar to those of Section 7.4 show that

$$\frac{1}{t}\log r_t \to \frac{1}{1-\gamma}\big(I_p(q_j) - I_p(q_i)\big) \tag{7.12}$$

almost surely. Suppose i is a p-investor, and j is a q-investor for some $q \neq p$. Then

$$\frac{1}{t}\log r_t \to \frac{I_p(q)}{1-\gamma}$$

and the long-run wealth share of the q investor converges to 0 almost surely.

Suppose now that there are $2N$ traders; that is, $I = 2N$. The first N traders are informed; they are p-investors. All informed traders have identical initial wealths w_0^i and all noise traders have initial wealths w_0^n. The remaining N investors are q-investors with extreme beliefs. For the sake of argument, suppose that $p > 1/2$ and $q > p$.

We see just by paring off investors of each type that the wealth share of the q-investor pool converges to 0. Now we would like to carry out calculations in the style of DeLong et al. (1991). They compute the log of the ratio of the group wealths, which in our simple model is

$$\lim_{N\to\infty} \log \frac{\sum_{i=1}^{N} w_1^i}{\sum_{i=N+1}^{2N} w_1^i} = \lim_{N\to\infty} \log \frac{\frac{1}{N}\sum_{i=1}^{N} w_1^i}{\frac{1}{N}\sum_{i=N+1}^{2N} w_1^i}$$
$$= \log \frac{pp^{1/\gamma} + (1-p)(1-p)^{1/\gamma}}{pq^{1/\gamma} + (1-p)(1-q)^{1/\gamma}} + \log \frac{w_0^i}{w_0^n}$$

If we assume that noise traders are extreme, the first term in the limit, call it r, is negative. Iterating this relationship, we see that

$$\lim_{t\to\infty} \lim_{N\to\infty} \frac{1}{t} \log \frac{\sum_{i=1}^{N} w_1^i}{\sum_{i=N+1}^{2N} w_1^i} = r < 0$$

so

$$\lim_{t\to\infty} \lim_{N\to\infty} \frac{\sum_{i=1}^{N} w_1^i}{\sum_{i=N+1}^{2N} w_1^i} = 0$$

We showed earlier, however, that

$$\lim_{t\to\infty} \frac{\sum_{i=1}^{N} w_1^i}{\sum_{i=N+1}^{2N} w_1^i} = +\infty$$

for all N, and so

$$\lim_{N\to\infty} \lim_{t\to\infty} \frac{\sum_{i=1}^{N} w_1^i}{\sum_{i=N+1}^{2N} w_1^i} = +\infty \tag{7.13}$$

It should be obvious that the second limit, Eq. 7.13, is the correct one to compute. Infinities do not exist in real economies. We are interested in the long-run behavior of many-agent economies. The second limit calculation shows that in any $2N$-trader economy, after a large enough amount of time, the share of wealth belonging to the noise traders will be nearly 0. The first limit is merely a mathematical artifact. DeLong et al. (1991, p. 12) state, "the wealth share of a randomly selected noise trader type eventually falls with probability one, but the wealth of a small fraction of the noise trader population is increasing fast enough to give them a rising aggregate share of the economy's wealth." Our calculations show that the wealth share of every noise-trader type eventually falls with probability 1, and so the noise-trader wealth share ultimately vanishes. In our example there is no magic whereby the populations of noise and informed traders are fixed; each noise trader's share of wealth vanishes, and yet the group thrives.

7.7. ROBUSTNESS

This section briefly describes generalizations and extensions of the market selection results to other economies.

7.7.1. Unbounded Economies

Several authors (Kogan et al., 2006; Malamud and Trubowitz, 2006; Dumas et al., 2006; and Yan, 2006) have analyzed economies in which the market does not select for traders with maximal survival index. The main thrust of these papers is that payoff functions also matter, so that even in complete markets' economies with a common discount factor, long-run asset prices can be wrong. These papers differ in several ways from our analysis, but the key difference is that they have unbounded aggregate endowments. In this section, we construct a simple example to illustrate the role of unbounded endowments.

Consider an IID economy as in Section 7.4.2 but to make calculation simple (and comparable to the papers cited here), we assume that all traders have CRRA utility, $u_i(c) = \delta_i^{-1} c_i^{\delta}$, with $\delta_i < 1$. Suppose that there are just two traders and let α_t denote Trader 2's consumption share of the aggregate endowment e_t. Suppose $e_t = w^{\lambda t}$. Then, by the SLLN, Eq. 7.9 implies that, almost surely,

$$\frac{1}{t} \log \left(\frac{\alpha_t^{1-\delta_2}}{(1-\alpha_t)^{1-\delta_1}} \right) \to (s_2 - s_1) + (\delta_2 - \delta_1)\lambda \log(w)$$

So the new survival index for Trader i is $\hat{s}_i = s_i + \delta_i \lambda \log(w)$. In the bounded economy $\lambda = 0$ and we again have the old survival index. Suppose that $s_1 > s_2$, so that in the bounded economy Trader 2 would vanish. Now, if $\delta_2 > \delta_1$, and the aggregate endowment grows fast enough, Trader 2 does not vanish and in fact her share of consumption converges almost surely to 1. Alternatively, if $\delta_1 > \delta_2$, and the aggregate endowment falls fast enough, Trader 2 does not vanish and in fact her share of consumption converges almost surely to 1. Now discount rates, relative entropy of beliefs, the relative risk aversion coefficient, and the growth rate of the aggregate endowment all matter. Note, however, that the same analysis would apply even in a deterministic economy. So, interpreting the CRRA coefficient on consumption as "risk aversion" is misleading; instead it matters because of its role in determining intertemporal marginal rates of substitution. This result was foreshadowed by the calculation summarized in Eq. 7.12. The observant reader may well have wondered what the $(1 - \gamma)$ denominator was doing in the survival indices. There too it arises because of the lack of a uniform bound on wealth.

7.7.2. Incomplete Markets

As we saw in Section 7.6.1, the market selection hypothesis can fail to be true in economies with incomplete markets. Blume and Easley (2006) provides examples that show that if markets are incomplete, rational traders may choose either savings rates or portfolio rules that are dominated by those selected by traders with incorrect beliefs. If some traders are irrationally optimistic about the payoff to assets, the price of those assets may be high enough so that rational traders choose to consume more now and less in the future. Their low savings rates are optimal, but as a result of their low savings rates, the rational traders do not survive.

Beker and Chattopadhyah (2006) analyze a market in which the only assets are money and one risky asset, so that (with enough states) the market is incomplete. They give an example in which a trader whose beliefs are correct and whose impatience is low is driven out of the market. They also show that having some agents driven out of the market is a robust feature of their incomplete markets' economies and that beliefs do not determine which trader is driven out.

7.7.3. Differential Information

If traders' differences in beliefs are due solely to information asymmetries, the market selection hypothesis requires that asset markets select for traders with superior information. The research discussed previously asks about selection over traders with different, but exogenously given, beliefs. Alternatively, if traders begin with a common prior and receive differential information they will have differing beliefs, but now they will care about each others' beliefs. In this case, the selection question is difficult because some of the information that traders have will be reflected in prices. If the economy is in a fully revealing rational expectations equilibrium, there is no advantage to having superior information (see Grossman and Stiglitz, 1980). So the question only makes sense in the more natural, but far more complex, case in which information is not fully revealed by market statistics.

Figlewski (1979) shows that traders with information that is not fully reflected in prices have an advantage in terms of expected wealth gain over those whose information is fully impounded in prices. But, as expected wealth gain does not determine fitness, this result does not answer the selection question. Mailath and Sandroni (2003) consider a Lucas trees economy with log utility traders and noise traders. They show that the quality of information affects survival, but so does the level of noise in the economy. Sciubba (2005) considers a Grossman and Stiglitz (1980) economy in which informed traders pay for information and shows that in this case uninformed traders do not vanish.

7.7.4. Selection over Non-EU Traders

To this point we have focused on selection within the class of subjective expected utility maximizers. But in recent years alternative decision theories have been developed and used to interpret various asset market anomalies. The most well known of these alternative decision theories is based on ambiguity aversion that is motivated by the famous experiment of Ellsberg (1961) showing that some decision makers prefer known odds to unknown odds. Ambiguity aversion has been used to explain the equity premium puzzle, portfolio home bias, and lack of participation in asset markets. Typically the research asking about the effect of ambiguity aversion on asset prices considers economies populated only by ambiguity-averse traders. But what if some traders are ambiguity averse and others are expected utility maximizers?

Condie (2008) considers a complete markets economy that is populated by a mixture of ambiguity-averse traders and expected utility traders. He asks if ambiguity-averse traders can survive and have persistent effects on asset prices in these economies. He finds that if there is no aggregate risk in the economy, ambiguity-averse traders can survive even in the presence of expected utility traders with correct beliefs. Condie models ambiguity-averse traders using the Gilboa and Schmeidler (1989) representation of ambiguity aversion in which traders' beliefs are represented by sets of probability distributions. Traders choose a portfolio that maximizes the minimum expected utility over the entire set of distributions. This formalization produces a kink in the trader's indifference curve at a risk-free portfolio. To see how the analysis works in an economy with no aggregate risk suppose, for example, that the economy consists of one expected utility trader with correct beliefs and one ambiguity-averse trader with a set of beliefs that contains the truth, and that the endowments of both traders are risk-free. Then in equilibrium there will be no trade and prices are set by the expected utility trader. These prices are just the (correct) probabilities of the states. So although the ambiguity-averse trader survives, this trader has no effect on trade or on asset pricing.

Alternatively, suppose that there is aggregate risk in the economy. In this case, Condie (2008) shows that ambiguity-averse traders who have a set of beliefs containing the truth in its interior cannot survive in the presence of expected utility traders with correct beliefs. In the earlier two-trader example, with aggregate risk, if the ambiguity-averse trader still holds a risk-free portfolio, the expected utility trader is earning the return to holding the aggregate risk and will drive the ambiguity-averse trader out of the market. Alternatively, if the ambiguity-averse trader holds risk in equilibrium, this trader is behaving as an expected utility trader with incorrect beliefs (those that minimize expected utility for the portfolio), and the analysis of Blume and Easley (2006) shows that they will vanish. So the conclusion is that either ambiguity-averse traders vanish or they survive and have no effect on prices. In either case they have no effect on prices in the long run.

These results have important implications for explanations of market pricing anomalies using non-expected-utility traders. To the extent that these explanations are based on ambiguity-aversion (of the Gilboa and Schmeidler, 1989, type), either they explain only short-run phenomena or there are no expected utility traders with correct beliefs in the economy. It would be interesting to ask whether similar results hold for other types of non-expected-utility motivated behavior.

7.7.5. Selection over Rules

We have focused on selection over traders whose behavior is motivated by preferences, but characteristics of preferences matter only through the behaviors they dictate. Selection analysis is equally relevant to the fate of decision rules (savings rates and portfolios) that do not arise from maximization. This is analogous to biological selection, which works not on genotypes (the full description of biological information) but on phenotypes (the characteristics actually expressed). Behavioral rules, for us, are phenotypes.

Consider an intertemporal general equilibrium economy with a collection of Arrow securities and one physical good available at each date. Suppose that traders are characterized by their stochastic processes of endowments of the good and by portfolio and savings rules. A savings rule describes the fraction of a trader's wealth she saves and invests at each date, given any partial history of states. Similarly, a portfolio rule describes the fraction of her savings the trader allocates to each Arrow security. The savings and portfolio rules that rational traders could choose form one such class of rules. But other, nonrationally motivated rules are also possible.

There are two questions to ask about the dynamics of wealth selection in this economy. First, is there any kind of selection at all? Is it possible to characterize the rules that win? Second, if selection does take place, does every trader using a rational rule survive, and in the presence of such a trader do all nonrational traders vanish?

In repeated betting, with exogenous odds, the betting rule that maximizes the expected growth rate of wealth is known as the *Kelly Rule* (Kelly, 1956). The use of this formula in betting with fixed but favorable odds was further analyzed by Breiman (1961). In asset markets the "odds" are not fixed; instead, they are determined by equilibrium asset prices, which in turn depend on traders' portfolio and savings rules. Nonetheless, the market selects over rules according to the expected growth rate of wealth share they induce. Blume and Easley (1992) show that if there is a unique trader using a rule that is globally maximal with respect to this criterion, this trader eventually controls all the wealth in the economy, and prices are set as if he is the only trader in the economy. A trader whose savings rate is maximal and whose portfolio rule is, in each partial history, the conditional probability of states for tomorrow has a maximal expected growth rate of wealth share. This rule is consistent with the trader having logarithmic utility for consumption, rational expectations, and a discount factor that is as large as any trader's savings rate. Thus, if this trader exists, he is selected for. However, rationality alone does not guarantee a maximal expected growth rate of wealth share. There are rational portfolio rules that do not maximize fitness (even controlling for savings rates), and traders who use these rules can be driven out of the market by traders who use rules that are inconsistent with rationality.

The "preference-based" approach to market selection limits the rules it studies to those generated by particular classes of preferences. Alternatively, one could limit attention to rules exhibiting particular behaviors. This is the approach of Amir, Evstigneev, Hens, and Schenk-Hoppé (2005) and Evstigneev, Hens, and Schenk-Hoppé (2006). They consider general one-period assets and ask if there are *simple* portfolio rules that are selected for, or are evolutionarily stable, when the market is populated by other simple portfolio rules. A *simple* rule is one for which the fraction of wealth invested in a given asset is independent of current asset prices. In this research, either all winnings are invested, or equivalently, traders are assumed to invest an equal fraction of their winnings; consumption rates are the same for all traders. So selection operates only over portfolio rules. Amir et al. (2005) find that a trader who allocates his wealth across assets according to their conditional expected relative payoffs drives out all other traders as long as none of the other traders ends up holding the market. This result is consistent with Blume and Easley (1992) because the log-optimal portfolio rule agrees with the conditional expected relative payoff rule, when only these two rules exist in the market. Hence, both of these rules end up holding the market in the limit. Evstigneev, Hens, and Schenk-Hoppé (2006) show that the expected relative payoffs rule is evolutionarily stable using notions of stability from evolutionary game theory.

7.8. CONCLUSION

The hypothesis that traders are rational in the sense of Savage has few implications for asset prices. The power of rationality lies in the framework that makes analyzing beliefs possible. One approach is to assume that all traders have rational—that is, correct—expectations. This assumption imposes structure on asset prices, but it is surely too strong an assumption. Instead, assuming that some traders may have more accurate beliefs than others is more plausible than the rational expectations assumption. In the short run, traders with incorrect expectations can influence asset prices. But in the long run, they may lose out to those with more accurate expectations who choose better portfolio rules. Expectations also affect savings rates, and traders with incorrect expectations can be induced to over-save, so the conjecture they they are driven out is far from obvious. The literature shows that if markets are dynamically complete, the economy is bounded, and traders have a common discount factor, then in fact the market is dominated in the long run by those with correct expectations. If there is any trader with correct beliefs, in the long run asset prices converge to their rational expectations values.

The assumptions that markets are complete, that the economy is bounded, and that traders have a common discount factor are all important for this selection result. If markets are incomplete, the selection hypothesis can fail and asset prices need not converge to their correct values. Traders with high discount factors and incorrect beliefs can drive out those with correct beliefs and lower discount factors, and this will cause even long-run asset prices to be incorrect. If there is heterogeneity in discount factors and beliefs so that multiple traders have maximal survival index, multiple traders can survive. But we show that having a maximal survival index is not sufficient, and the sufficient condition derived here for the IID economy is new. In this case, long-run price volatility is possible. Finally, in unbounded economies, discount factors, beliefs, and the curvature of the utility function all matter for survival. We provide new results that show, for CARA economies, how all these factors enter into a new survival index.

References

Alchian, A., Uncertainty, evolution and economic theory, *Journal of Political Economy*, 1950, 58, 211–221.

Amir, R., I. Evstigneev, T. Hens, and K. R. Schenk-Hoppé, Market selection and the survival of investment strategies, *Journal of Mathematical Economics*, 2005, 41, 105–122,

Barlow, N. (ed.), *The Autobiography of Charles Darwin, 1809–1882, with Original Omissions Restored*, W. W. Norton, 1969.

Beker, P., and S. Chattopadhyah, Economic survival when markets are incomplete, unpublished, University of Warwick, 2006.

Black, F., Noise, *Journal of Finance*, 1986, 41, 529–543.

Blume, L. E., and D. Easley, Evolution and market behavior, *Journal of Economic Theory*, 1992, 58, 9–40.

Blume, L. E., and D. Easley, Learning to be rational, *Journal of Economic Theory*, 1982, 26, 340–351.

Blume, L. E., and D. Easley, Economic natural selection, *Economics Letters*, 1993, 42, 281–289.

Blume, L. E., and D. Easley, Optimality and natural selection in markets, *Journal of Economic Theory*, 2002, 107, 95–135.

Blume, L. E., and D. Easley, If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets, *Econometrica*, 2006, 74, 929–966.

Breiman, L., Optimal gambling systems for favorable games, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (ed.), Vol. 1, pp. 65–78, University of California Press, 1961.

Condie, S., Living with ambiguity: prices and survival when investors have heterogeneous preferences for ambiguity, *Economic Theory*, 2008, 36, 81–108.

Cootner, P. H., *The Random Character of Stock Market Prices*, MIT Press, 1964.

DeLong, J. B., A. Shleifer, L. Summers, and R. Waldmann, Noise trader risk in financial markets, *Journal of Political Economy*, 1990, 98, 703–738.

DeLong, J. B., A. Shleifer, L. Summers, and R. Waldmann, The survival of noise traders in financial markets, *Journal of Business*, 1991, 64, 1–19.

Dow, J., and G. Gorton, Noise traders, in *The New Palgrave Dictionary of Economics*, 2nd ed., S. Durlauf and L. Blume (eds.), Palgrave Macmillan, 2008.

Dumas, B., A. Kurshev, and R. Uppal, What can rational investors do about excessive volatility and sentiment fluctuations?, unpublished, INSEAD, 2006.

Dutta, P., and R. Radner, Profit maximization and the market selection hypothesis, *Review of Economic Studies*, 1999, 66, 769–798.

Ellsberg, D., Risk, ambiguity, and the savage axioms, *Quarterly Journal of Economics*, 1961, 75, 643–669.

Enke, S., On maximizing profits: A distinction between Chamberlin and Robinson, *American Economic Review*, 1951, 41, 566–578.

Evstigneev, I., T. Hens, and K. R. Schenk-Hoppé, Evolutionary stable stock markets, *Economic Theory*, 2006, 27, 449–468.

Fama, E., The behavior of stock market prices, *Journal of Business*, 1965, 38, 34–105.

Figlewski, S., Subjective information and market efficiency in a betting model, *Journal of Political Economy*, 1979, 87, 75–88.

Friedman, M., *Essays in Positive Economics*, University of Chicago Press, 1953.

Gilboa, I., and D. Schmeidler, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics*, 1989, 18, 141–153.

Grossman, S. J., and J. E. Stiglitz, On the impossibility of informationally efficient markets, *American Economic Review*, 1980, 70, 393–408.

Kelly, J. L., A new interpretation of information rate, *Bell System Technical Journal*, 1956, 35, 917–926.

Knight, F. H., The ethics of competition, *Quarterly Journal of Economics*, 1923, 37, 579–624.

Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield, The price impact and survival of irrational traders, *Journal of Finance*, 2006, 61, 195–229.

Koopmans, T. C., *Three Essays on the State of Economic Science*, McGraw-Hill, 1957.

Kyle, A., Continuous auctions and insider trading, *Econometrica*, 1985, 53, 1315–1335.

Mailath, G., and A. Sandroni, Market selection and asymmetric information, *Review of Economic Studies*, 2003, 70, 343–368.

Malamud, S., and E. Trubowitz, The long horizon behavior of asset prices in heterogeneous economies, unpublished, ETH Zürich, 2006, *http://ssrn.com/paper=953552*.

Marshall, A., *Principles of Economics*, 9th ed., Macmillan, 1961.

Peleg, B., and M. E. Yaari, Markets with countably many commodities, *International Economic Review*, 1970, 11, 369–377.

Samuelson, P., Models of thought in economics and biology, *American Economic Review*, 1985, 75, 166–172.

Sandroni, A., Do markets favor agents able to make accurate predictions?, *Econometrica*, 2000, 68, 1303–1342.

Savage, L. J., The theory of statistical decision, *Journal of the American Statistical Asssociation*, 1951, 46, 55–67.

Schumpeter, J. A., *The Theory of Economic Development; An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*, Harvard University Press, 1934.

Sciubba, E., Asymmetric information and survival in financial markets, *Economic Theory*, 2005, 25, 353–379.

Williams, G. C., *Natural Selection: Domains, Levels, and Challenges*, Oxford University Press, 1992.

Winter, S., Satisficing, selection and the innovating remnant, *Quarterly Journal of Economics*, 1971, 85, 237–261.

Yan, H., On the market selection hypothesis, unpublished, Yale University, 2006.

CHAPTER 8

Rational Diverse Beliefs and Market Volatility

Mordecai Kurz
Stanford University

8.1.	Introduction	440
8.2.	Can Market Dynamics Be Explained by Asymmetric Private Information?	445
	8.2.1. A General Model of Asset Pricing under Asymmetric Information	445
	8.2.2. Dynamic Infinite Horizon Models	450
	8.2.3. Is Asymmetric Information a Satisfactory Theory of Market Dynamics?	453
8.3.	Diverse Beliefs with Common Information: The General Theory	454
	8.3.1. A Basic Principle: Rational Diversity Implies Volatility	455
	8.3.2. Stability and Rationality in a General Nonstationary Economy	457
	8.3.3. Belief Rationality and the Conditional Stability Theorem	461
	8.3.4. Describing Individual and Market Beliefs with Markov State Variables	464
	8.3.5. Asset Pricing with Heterogeneous Beliefs: An Illustrative Model and Implications	474
8.4.	Explaining Market Dynamics with Simulation Models of Diverse Beliefs	485
	8.4.1. Introduction: On Simulation Methods and the Main Results	485
	8.4.2. Anatomy of Market Volatility	486
	8.4.3. Volatility of Foreign Exchange Rates and the Forward Discount Bias	498
	8.4.4. Macroeconomic Applications	499
8.5.	Conclusion and Open Problems	501
	References	502

Note: I thank Andrea Beltratti, Ken Judd, Stephen Morris, Maurizio Motolese, Hiroyuki Nakata, Carsten Nielsen, Eduardo Perez, Ho-Mou Wu, Min Fan, and John O'Leary for helpful discussions of the ideas in this chapter and detailed comments on earlier versions of the manuscript. Ho-Mou Wu helped me write the "Speculation" section, Eduardo Perez made helpful comments on an earlier draft and assisted with the editorial work of the manuscript, and Maurizio Motolese helped me prepare the figures. Finally, special thanks are due to the editors, Thorsten Hens and Klaus Reiner Schenk-Hoppé, who tirelessly read an earlier version of this chapter and made valuable suggestions that helped shape the structure of the final manuscript.

Abstract

This chapter explores theories that explain market dynamics by using rational but diverse beliefs and employing mechanisms of endogenous amplification. Noisy Rational Expectations Asset-Pricing Theory (in which belief diversity arises from diverse private information) does not offer a satisfactory paradigm, because increased idiosyncratic private information *reduces* market volatility. We then focus on models of diverse beliefs *without* private information, whereby economic agents do not know the structure of the complex economy but infer empirical probability from data. A belief is a model of *deviations* from this empirical distribution. It is shown that in a world of diverse beliefs, *to be rational*, *a belief must fluctuate around the empirical frequencies*, generating endogenous amplification. *Market belief*, which is the distribution of individual beliefs, is then observable via sampling. We explore an explicitly solvable asset-pricing model with diverse beliefs to illustrate the central implications of the theory for market dynamics, the nature of uncertainty and risk premia. *Simulations* are employed to illustrate the ability of the theory to explain the stylized empirical facts. The results offer a unified explanation of the key features of market dynamics, such as excess price volatility, the Equity Premium Puzzle, predictability of asset returns, and stochastic volatility.

Keywords: diverse beliefs, private information, rational beliefs, market beliefs, empirical probability, stable probability, excess volatility, endogenous uncertainty, volume of trade, risk premium

8.1. INTRODUCTION

This chapter explores the role of rational diverse beliefs in explaining market dynamics and volatility. To do that we examine alternative reasons for belief diversity, which is rational, and review different models of market dynamics that incorporate such beliefs. The term *market dynamics* refers to the dynamic characteristics of financial markets, but we focus on dynamic phenomena that attracted attention in the literature. Examples include excess volatility of asset returns, high and time-varying equity risk premia, high volume of trade, and so on. Many examples are termed "anomalies" or "puzzles" because they contradict predictions of *rational expectations equilibria* (in short, REE) with full information. Most studies show that fundamental exogenous factors cannot explain the observed dynamics, leading Paul Samuelson to quip that "the stock market predicted nine of the last five recessions." We aim to explore ideas that explain the four recessions the market predicted but that did not occur. The problem of market volatility and the question of whether asset markets exhibit "excess" volatility relative to fundamental factors have been central in financial economics; thus, the approach explored here addresses major questions and offers useful ideas for advancing the scientific study of free markets, for public stabilization policy, and for practitioners in financial markets.

It is important to note that the idea that allocations and prices are affected by agents' perceptions of the future is rather old. Diverse expectations are central to Thornton's

(1802) views of paper money and financial markets. Expectations are crucial to Keynes (1936). Chapter 12 of *The General Theory* examines the "state of confidence" and the importance of investors' expectations to asset pricing. Expectations are key to the "cumulative movements" in Pigou (see Pigou, 1941, Chapter VI) and constitute the mechanism of deviations from a stationary equilibrium in the Swedish school (e.g., see Myrdal's 1939 views in Myrdal, 1962, Chapter III). Also, "subjective values" based on diverse agents' expectations are cornerstones of Lindahl's (1939) theory of money and capital.

Before turning to recent work, we draw attention to an assumption made in the work reviewed later. It holds that *the distribution of beliefs in the market is an observable variable* that can be deduced from forecast panel data. It would thus be useful to briefly review some available raw data.

In the post-World War II era, large databases on heterogeneous forecasts of various variables have been assembled in Holland, Germany, and Sweden. In the United States, the Survey of Professional Forecasts, reporting *quarterly* forecasts of private forecasters, was started in 1968. It was first conducted by the American Statistical Association's National Bureau but has since been taken over by the Federal Reserve Bank of Philadelphia. Since 1980 the Blue Chip Economic Indicators reports *monthly* forecasts of economic variables by over 50 financial institutions. This service was expanded, under the title of Blue Chip Financial Forecasts, to include forecasts of interest rates and other variables. To illustrate, Table 8.1 reports forecasts of GDP growth and change in GDP deflators in May 2000 for the year 2000. Actual GDP growth rate in 2000 was 4.1% and actual inflation rate was 2.3%. Note that in May 2000, *five months into the year*, large heterogeneity persisted. Also, almost all GDP growth forecasts were wrong! To understand this correlated error, place yourself in May 2000 and make a stationary econometric forecast of GDP growth, but make no special judgment about the unique conditions in May 2000. An example of such a model was developed by Stock and Watson (2001, 2002, 2005). They estimate it by using a combination of diffusion indexes and bivariate VAR forecasts and employing a large number of U.S. time series. In May 2000 the nonjudgmental stationary forecast of GDP growth was *lower than most private forecasts*.

Repeating the experiment over time, we find that the distribution of forecasts fluctuates in two ways. First, it exhibits changes in the *cross-sectional variance* of the forecasts, reflecting changes in degree of disagreement. Second, it exhibits large fluctuations over time *in relation to the stationary forecasts* reflecting correlation in forecasters' views about unusual conditions at the time. Sometimes forecast distributions are below the stationary forecast, whereas in May 2000 the distribution was above the stationary forecast. Observe that the noted large data banks of market forecast distributions are publically available for many variables because forecasters are willing to reveal their forecasts.

We thus note that for any variable, *individual forecasts are correlated and the average market forecast fluctuates around the stationary forecast*. This nonjudgmental forecast is a central yardstick in the work reviewed later. Also, observe that the cross-sectional variance of forecasts fluctuates over time.

Despite the impact of the rational expectations paradigm, belief heterogeneity has been used to explain many phenomena such as asset price volatility, risk premia, volume

TABLE 8.1 May 2000 BLUE Forecasts of Growth and Inflation Rates for the Full Year 2000

May 2000 Forecasted Percent Change in Forecast for 2000	Real GDP Growth	GDP Price Deflator
First Union Corp.	5.3H	2.0
Turning Points (Micrometrics)	5.2	2.1
J. P. Morgan	5.2	2.1
Evans, Carroll & Assoc.	5.1	2.2
Mortgage Bankers Association	5.1	2.1
Goldman Sachs Group, Inc.	5.1	2.1
U.S. Trust Company	5.1	2.0
U.S. Chamber of Commerce	5.1	2.0
Bank of America Corp.	5.1	2.0
Morgan Stanley Dean Witter	5.1	1.9
Wayne Hummer Investments LLC	5.0	2.3
Bank One Corp.	5.0	2.1
Nomura Securities Co.	5.0	1.9
Merrill Lynch	5.0	1.9
Perna Associates	4.9	2.3
National Assn. of Home Builders	4.9	2.1
Macroeconomic Advisers, LLC	4.9	2.1
Prudential Securities, Inc.	4.9	2.0
LaSalle National Bank	4.8	2.3
Conference Board	4.8	2.3
Wells Capital Management	4.8	2.2
DuPont	4.8	2.1
Northern Trust Company	4.8	2.1
Chicago Capital, Inc.	4.8	2.0
Deutsche Bank Securities	4.8	1.8
Chase Securities, Inc.	4.8	1.8
Credit Suisse First Boston	4.8	1.8
Comerica	4.7	2.4
Moody’s Investors Service	4.7	2.2
Fannie Mae	4.7	2.0
Federal Express Corp.	4.7	2.0
SOM Economics, Inc.	4.7	1.9
National Assn. of Realtors	4.7	1.9
National City Corporation	4.7	1.9
ClearView Economics	4.7	1.9

(Continued)

TABLE 8.1 (*Continued*)

May 2000 Forecasted Percent Change in Forecast for 2000	Real GDP Growth	GDP Price Deflator
Eggert Economic Enterprises, Inc.	4.6	2.1
WEFA Group	4.6	1.9
Eaton Corporation	4.6	1.9
Bear Stearns & Co., Inc.	4.6	1.2 L
Ford Motor Company	4.5	1.8
Motorola	4.5	1.7
Standard & Poors Corp.	4.5	1.7
UCLA Business Forecasting Project	4.4	2.1
Inforum–University of Maryland	4.4	2.0
Prudential Insurance Co.	4.4	1.9
Weyerhaeuser Company	4.3	2.2
DaimlerChrysler AG	4.3	2.0
Georgia State University	4.2	2.2
Kellner Economic Advisers	4.2	2.0
Econoclast	4.1	2.0
Naroff Economic Advisors	4.0 L	2.5 H

of trade, and money nonneutrality. There are two general theories of rational behavior motivated by the observed diversity. One follows the Harsanyii doctrine, viewing people as Bayesians who hold the same prior probability but with asymmetric private information used in forecasting.

Examples of supporting papers include Grossman and Stiglitz (1980), Phelps (1970), Lucas (1972), Diamond and Verrecchia (1981), Townsend (1978, 1983), Singleton (1987), Brown and Jennings (1989), Grundy and McNichols (1989), Wang (1994), He and Wang (1995), Judd and Bernardo (1996, 2000), Morris and Shin (2002, 2005), Woodford (2003), Hellwig (2002, 2005), Angeletos and Pavan (2006), Angeletos and Werning (2006), Allen, Morris, and Shin (2006), and Hellwig, Mukherji, and Tsyvinski (2006).

An alternative view sees no evidence for the use of private information in forecasts of market prices or economic aggregates. It finds no justification for a common prior and insists that diverse beliefs about state variables are inevitable in a complex world. A sample of papers taking this approach include Harrison and Kreps (1978), Varian (1985, 1989), Harris and Raviv (1993), Kurz (1994, 1996, 1997a, 1997b, 1997c, 2007) Detemple and Murthy (1994), Frankel and Rose (1995), Kandel and Pearson (1995), Cabrales and Hoshi (1996), Kurz and Beltratti (1997), Brock and Hommes (1997, 1998), Kurz and Motolese (2001, 2007), Kurz and Schneider (1996), Kurz and Wu (1996), Kurz, Jin and Motolese (2005a, 2005b), Nielsen (1996, 2003, 2005, 2006), Motolese (2001, 2003), Wu and Guo (2003, 2004), Fan (2006), Acemoglu et al. (2007), Nakata (2007), and Guo and Wu (2007).

We first clarify the differences between these two theories and examine whether they solve the problems outlined. Since the range of issues is wide, we concentrate on rationalized beliefs, and this excludes two types of models. The first type is behavioral finance and noise trading models, in which belief diversity arises from psychological but irrational motives. The second is learning models with common information, typically models of convergence to rational expectations; hence in these models belief diversity is not persistent. Before turning to an evaluation of the difference between these two approaches, it is useful to clarify the standards we set for advocating or rejecting a theory.

Since standard models explain dynamics with exogenous shocks and these are not sufficient to explain the data, to explain excess volatility we search for mechanisms of *endogenous amplification*. In addition, it will become clear that not all belief heterogeneity generates market dynamics. Hence, we ask, What must be the structure of heterogeneity for belief diversity to matter? The data reveal that heterogeneity persists; therefore the two key criteria for effective diversity in any theory are that *it has aggregate effects and that these effects are nonvanishing*. The nonvanishing condition is challenging to models of asymmetric information under rational expectations since information revelation of prices leads back to a common belief, and hence diversity cannot persist. Asymmetric private information in a rational expectations equilibrium is therefore usually supplemented with a "noisy" mechanism to avoid revelation. But then we must ask whether such a mechanism is natural or is it just an artificial construct? Is it testable? Requiring diversity to have persistent aggregate effects implies that heterogeneity by itself is not sufficient and it must be supplemented with dynamic features. To understand the importance of this fact, consider two examples:

1. Beliefs are diverse, randomly and independently distributed over agents with a fixed distribution over time. Here an agent's belief measured by, say, a density over states, changes over time but is randomly determined. The IID distribution causes a cancellation of the effects of beliefs; hence there is a constant, typically small, aggregate effect. Such a distribution implies that diversity is irrelevant.
2. Beliefs are heterogeneous *with a fixed distribution of beliefs so that the belief of each agent is fixed on that distribution*. An agent always has the same superior or inferior information, or else specific agents are always more optimistic or more pessimistic than others. A fixed belief distribution also implies that prices, volume of trade, and risk premia fluctuate only in response to exogenous shocks; hence we are back to a theory, rejected by the data, that publicly observed exogenous shocks are the only cause of fluctuations. Such distributions of beliefs do not generate the desired endogenous amplification to explain excess volatility.

These examples show that the *dynamics* of beliefs over time are essential, and the question is, what is their source? Under asymmetric private information such dynamics could be generated by an exogenous flow of private information, which entails a process of belief updating. How effective or plausible such an assumption is must be carefully weighed, and we discuss it in detail in Section 8.2.3. The situation is different under diverse beliefs with common information, since dynamics and rationality are inherently interrelated. We explain in Section 8.3.1 that the central principle that drives the theory of rational heterogeneous beliefs is that *rational diversity of correlated beliefs without private information implies market volatility*.

The chapter is organized as follows: Section 8.2 reviews the structure of *noisy REE* asset-pricing theory. It shows that although this theory has many useful features, it fails to deliver a consistent theory of financial market dynamics. Indeed, volatility and volume of trade *decline with belief diversity*. Section 8.3 reviews the theory of diverse beliefs with common information. It shows that the theory delivers a consistent and plausible model of endogenous amplification and provides a foundation for understanding market dynamics. Section 8.4 reviews simulation models based on the theory in Section 8.3. It shows that simulations of models with diverse beliefs match the observed data well. Finally, Section 8.5 reviews some open problems.

8.2. CAN MARKET DYNAMICS BE EXPLAINED BY ASYMMETRIC PRIVATE INFORMATION?

The literature on asset pricing in "noisy" REE under asymmetric private information is large, and Brunnermeier (2001) provides a good survey. We discuss it in three stages. In Section 8.2.1 we present a universally used model with exponential utility. In Section 8.2.2 we discuss dynamic versions of the model. In Section 8.2.3 we evaluate the developed ideas.

8.2.1. A General Model of Asset Pricing under Asymmetric Information

The model reviewed here is an adaptation of the short-lived trader model of Brown and Jennings (1989). Similar models were used by Grossman and Stiglitz (1980), Diamond and Verrecchia (1981), Singleton (1987), Grundy and McNichols (1989), Wang (1994), He and Wang (1995), Bacchetta and van Wincoop (2005, 2006), Allen, Morris, and Shin (2006), and others.

There is a unit mass of traders, indexed by the [0, 1] interval and a single aggregate asset with unknown intrinsic unit value V. The economy is static with one period divided into three trading dates (no discounting): At date 1, traders first receive public information y and private signals x^i about the asset value and then they trade. At date 2 they trade again. At date 3 (or end of date 2) uncertainty is resolved, the true liquidation value V is revealed and traders receive it for their holdings. Public information is that V is distributed in accord with $V \sim N(y, \frac{1}{\alpha})$. The private signal x^i about V is $x^i = V + \varepsilon^i$, where ε^i satisfy $\varepsilon^i \sim N(0, \frac{1}{\beta})$ *independently across* i. (α, β) are known. Since these facts are common knowledge, agents know the true unknown value V is "in the market," since by the law of large numbers the mean of all private signals is the true value V. Trader i starts with S^i units of the aggregate asset and can borrow at zero interest rate to finance trading. (D_1^i, D_2^i) are i's demands in the first and second rounds and (p_1, p_2) are market prices in the two rounds. Ending wealth is thus $W^i = S^i p_1 + D_1^i(p_2 - p_1) + D_2^i(V - p_2)$.

All traders are assumed to have the same utility over wealth $u(W^i) = -e^{-(W^i/\tau)}$, and they maximize expected utility. Aggregate supplies (S_1, S_2) of shares, each representing an asset unit, traded in each of the rounds, are *random, unobserved, and independently normally distributed with mean zero*. This noise is crucial to ensure that traders cannot deduce from prices the true value of V. In a noisy REE traders maximize expected utility of final wealth while markets clear after traders deduce from prices all possible information. Indeed, Brown and Jennings (1989) show that equilibrium price at date 1 is of the form

$$p_1 = \kappa_1(\lambda_1 y + \mu_1 V - S_1) \quad \textbf{(8.1a)}$$

and since S_1 is normally distributed, p_1 is also normally distributed. Keep in mind that V and S_1 are unknown, hence Eq. 8.1a shows that prices are not fully revealing.

Since over trading rounds V remain fixed, more rounds of trading generate more price data from which traders deduce added information about V. But with additional supply shocks the inference problem becomes more complicated. That is, at date 2 the price p_2 contains more information about V, but it depends on *two* unobserved noise shocks (S_1, S_2). Hence, the price function is shown to be time dependent and at date 2 takes the form

$$p_2 = \hat{\kappa}_2(\hat{\lambda}_2 y + \hat{\mu}_2 V - S_2 + \psi S_1) \quad \textbf{(8.1b)}$$

Since the realized noise S_1 is not observed, traders condition on the known price p_1 to infer what they can about S_1. They thus use a date 2 price function, which takes an equivalent form

$$p_2 = \kappa_2(\lambda_2 y + \mu_2 V - S_2 + \xi_{21} p_1) \quad \textbf{(8.1c)}$$

By Eq. 8.1a, equivalence means $\kappa_2 = \hat{\kappa}_2, \lambda_2 = (\hat{\lambda}_2 + \lambda_1\psi), \mu_2 = (\hat{\mu}_2 + \mu_1\psi)$ and $\xi_{21} = -\frac{\psi}{\kappa_1}$. Denote by (H_1^i, H_2^i) the information of i in the two rounds. The linearity of the equilibrium price map implies that the payoff is normally distributed. Brown and Jennings (1989) show in their Appendix A that there exist constants (G_1, G_2), determined by the covariance matrix of the model's random variables and assumed by most writers to be the same for all agents, such that the demand functions of i are

$$D_2^i(p_2) = \frac{\tau}{\mathrm{Var}^i(V|H_2^i)}[E^i(V|H_2^i) - p_2] \quad \textbf{(8.2a)}$$

$$D_1^i(p_1) = \frac{\tau}{G_1}[E^i(p_2|H_1^i) - p_1] + \frac{(G_2 - G_1)}{G_1}[E^i(D_2^i|H_1^i)] \quad \textbf{(8.2b)}$$

Most writers assume $\text{Var}^i(V|H_2^i) = \sigma_V^2$ independent of i. The second term in Eq. 8.2b is the "hedging demand" arising from a trader's date 1 perceived risk of price change between date 1 and date 2. The hedging demand in a noisy REE complicates the inference problem and raises equilibrium existence problems. As a result, most writers ignore this demand and study *myopic-investor economies*, where there are only "short-lived" traders who live one period only. They first trade in date 1, gain utility from p_2, and leave the economy. They are replaced by new short-lived traders who receive the information of the first traders but trade in date 2 only and gain utility from the revealed V. None of them have hedging demands. A "long-lived" trader lives through *both* periods, trades in dates 1 and 2, and hence has a hedging demand. For simplicity we follow here the common practice and ignore the second term in Eq. 8.2b. We now average on i, equate to supply and conclude that

$$p_2 = \overline{E}_2(V) - (\frac{1}{\tau})\sigma_V^2(S_1 + S_2), \quad p_1 = \overline{E}_1(p_2) - \frac{G_1}{\tau}S_1 \tag{8.3}$$

$\overline{E}_2(V)$ is date 2 average market forecast of V and $\overline{E}_1(p_2)$ is average market forecast of p_2. In this case $G_1 = \text{Var}_1^i(p_2)$ and it is assumed this variance is the same for all i. Hence, the proof of Eqs. 8.1a and 8.1b amounts to exhibiting a closed form solution of $\overline{E}_2(V)$ and solving the joint system in Eq. 8.3.

The derivation of Eqs. 8.2a and 8.2b used a *general* form of conditional expectations and required only that prices are normally distributed. It is thus a general solution for any informational structure used in the conditioning and *it does not depend on the private character of information*. Hence, Eqs. 8.2a and 8.2b are applicable to models with diverse beliefs and common information as long as their implied prices are normally distributed. Moreover, differences among theories of diverse beliefs are expressed entirely by differences in their implications to the conditional expectations in Eqs. 8.2a and 8.2b. In the case of asymmetric private information discussed here, Eq. 8.2a shows that D_2^i depend on date 2 expectations, which are updated based on the information deduced from p_2 and p_1. This is different from date 1 information, which consists of a public signal, private signals, and inference from p_1 only. This is why equilibrium price maps are time dependent. Allen et al. (2006) present in their Appendix A computations of the closed-form solution. To get an idea of the inference involved, we briefly review the steps they take.

What does a trader learn in Round 1? Given a prior belief $V \sim N(y, \frac{1}{\alpha})$, trader i observes $p_1 = \kappa_1(\lambda_1 y + \mu_1 V - S_1)$. Since $S_1 \sim N(0, 1/\gamma_1)$, all he infers from date 1 price is that

$$[1/(\kappa_1\mu_1)](p_1 - \kappa_1\lambda_1 y) = V - [S_1/\mu_1] \sim N(V, 1/(\mu_1^2\gamma_1))$$

But now his added piece of information is the private signal $x^i = V + \epsilon^i, \epsilon^i \sim N(0, \frac{1}{\beta})$. Using a standard Bayesian inference from these three sources, his posterior belief becomes

$$E_1^i(V|H_1^i) = \frac{\alpha y + \beta x^i + \mu_1^2\gamma_1 \frac{1}{\kappa_1\mu_1}(p_1 - \kappa_1\lambda_1 y)}{\alpha + \beta + \mu_1^2\gamma_1} = \frac{(\alpha - \mu_1\gamma_1\lambda_1)y + \beta x^i + \frac{\mu_1\gamma_1}{\kappa_1}p_1}{\alpha + \beta + \mu_1^2\gamma_1} \quad \textbf{(8.4a)}$$

with precision

$$\alpha + \beta + \mu_1^2\gamma_1 \quad \textbf{(8.4b)}$$

Averaging Eq. 8.4a over the population, we can see that the average market forecast at date 1 is then

$$\overline{E}_1(V|H_1) = \frac{(\alpha - \mu_1\gamma_1)y + \beta V + \frac{\mu_1\gamma_1}{\kappa_1}p_1}{\alpha + \beta + \mu_1^2\gamma_1} \equiv \frac{\alpha y + (\beta + \mu_1^2\gamma_1)V}{\alpha + \beta + \mu_1^2\gamma_1} - \frac{\mu_1\gamma_1 S_1}{\alpha + \beta + \mu_1^2\gamma_1}$$

In Round 2 a trader observes p_2, which, as seen in Eq. 8.1c, is a function of p_1. Given p_1 and the fact that $S_2 \sim N(0, \frac{1}{\gamma_2})$, he infers from $p_2 = \kappa_2(\lambda_2 y + \mu_2 V - S_2 + \xi_{21}p_1)$ that

$$[1/(\kappa_2\mu_2)](p_2 - \kappa_2\lambda_2 y - \kappa_2\xi_{21}p_1) = V - [S_2/\mu_2] \sim N(V, 1/(\mu_2^2\gamma_2))$$

He now updates Eqs. 8.4a and 8.4b. Since supply shocks are IID, the updated posterior is standard

$$E_2^i(V|H_2^i) = \frac{\left[\frac{(\alpha - \mu_1\gamma_1\lambda_1)y + \beta x^i + \frac{\mu_1\gamma_1}{\kappa_1}p_1}{\alpha + \beta + \mu_1^2\gamma_1}\right](\alpha + \beta + \mu_1^2\gamma_1) + \frac{1}{\kappa_2\mu_2}(P_2 - \kappa_2\lambda_2 y - \kappa_2\xi_{21}p_1)(\mu_2^2\gamma_2)}{\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2}$$

Simplification leads to

$$E_2^i(V|H_2^i) = \frac{[\alpha - \mu_1\gamma_1\lambda_1 - \mu_2\gamma_2\lambda_2]y + \beta x^i + [\frac{\mu_1\lambda_1}{\kappa_1}p_1 + \frac{\mu_2\gamma_2}{\kappa_2}p_2 - \mu_2\gamma_2\xi_{21}p_1]}{\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2} \quad \textbf{(8.5a)}$$

$$\text{Var}(V|H_2^i) = \frac{1}{\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2} \quad \textbf{(8.5b)}$$

Finally, to compute Eq. 8.1c we average Eq. 8.5a to conclude that

$$\overline{E}_2(V) = \frac{[\alpha - \mu_1\gamma_1\lambda_1 + \mu_2\gamma_2\lambda_2]y + \beta V + [\frac{\mu_1\lambda_1}{\kappa_1}p_1 + \frac{\mu_2\lambda_2}{\kappa_2}p_2 - \mu_2\gamma_2\xi_{21}p_1]}{\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2} \quad \textbf{(8.6a)}$$

$$\overline{E}_1(p_2) = \kappa_2(\lambda_2 y + \mu_2 \overline{E}_1(V) + \xi_{21} p_1) \quad \textbf{(8.6b)}$$

We now solve for prices by inserting Eqs. 8.6a and 8.6b into Eq.8.3. The final step is to match coefficients of the price functions and identify $(\kappa_1, \lambda_1, \mu_1, \kappa_2, \lambda_2, \mu_2, \xi_{21})$. For more details of these computations see Allen et al. (2006), Appendix A. This verifies that prices are indeed normally distributed as in Eqs. 8.1a and 8.1b.

What is the length of memory in prices? Multiple trading rounds provide opportunities to deduce more information from prices about V. As trading continues, information on *all past prices is used*, since prices depend on all past unobserved supply shocks. In such a case, the price system is not a finite memory Markov process. The model has been extended to multiperiod trading whereby V is revealed N periods later (see Brown and Jennings, 1989; Grundy and McNichols, 1989; He and Wang, 1995; and Allen et al., 2006). In these models the complexity of inference depends on the presence of a hedging demand of long-lived traders. However, for both long- and short-lived traders, *the number of trading rounds is an arbitrary modeling construct*. It would thus be instructive to examine the limit behavior. In a third round of trading by short-lived traders, the price map becomes

$$p_3 = \kappa_3(\lambda_3 y + \mu_3 V - S_3 + \xi_{31} p_1 + \xi_{32} p_2)$$

Hence, the independent supply shock leads to an updating rule, which is again standard:

$$E_3^i(V|H_3^i) = \frac{E^i(V|H_2^i)(\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2) + \frac{1}{\kappa_3\mu_3}(P_3 - \kappa_3\lambda_3 y - \kappa_3\xi_{31}p_1 - \kappa_3\xi_{32}P_2)(\mu_2^3\gamma_3)}{\alpha + \beta + \mu_1^2\gamma_1 + \mu_2^2\gamma_2 + \mu_3^2\gamma_3}$$

By expressing individual and market forecasts in terms of the unobserved variables, one can easily extend the above to N rounds of trade, and it can be shown that they take the general form

$$E_N^i(V|H_N^i) = \frac{\alpha y + \beta x^i + \sum_{j=1}^{N} \mu_j^2\gamma_j V}{\alpha + \beta + \sum_{j=1}^{N} \mu_j^2\gamma_j} - \frac{\sum_{j=1}^{N} \mu_j\gamma_j S_j}{\alpha + \beta + \sum_{j=1}^{N} \mu_j^2\gamma_j} \quad \textbf{(8.7)}$$

A standard argument shows the μ_j converge. For simplicity assume $\gamma_j = \gamma$. The independence of the noise S_j together with Eq. 8.7 and the law of large numbers imply that with probability 1 the first term converges to V and the second converges to 0. Hence, in the limit, with probability 1, all forecasts converge to the true V and the effect of the public signal y disappears. This proves that *repeated trading leads to a full revelation of the true value* and that in the limit, $p = V$ and at that time traders do not forecast prices at all. With repeated trading the effect of y disappears. If the unit of time is short, like a month, trading rounds are not really limited. Hence this result contradicts the claim (e.g., see Allen et al., 2006) that the effect of the public signal y on prices lingers on forever.

Allen et al. (2006) use the model to explain the Keynes (1936) Beauty Contest. To see how, recall that $\overline{E}(S_i) = 0$. Then Eq. 8.3 implies that if there are N rounds of trade, then

$$p_1 = \overline{E}_1\overline{E}_2 \ldots \overline{E}_N(V) - [\mathrm{Var}_1(p_2)/(\tau)]S_1 \tag{8.8}$$

The authors then propose that Eq. 8.8 represents the Beauty Contest metaphor, since the price is not equal to the market expectations of V but rather to the average expectation of what the market expects the average expected value of V will be in the future. We comment on this interpretation in Section 8.3.5.

What have we learned so far? The key conclusions of the private information paradigm are that equilibrium prices vary with each date's true intrinsic value of the asset and with the random supply shock of that date, both of which are fundamental factors. Private information, as such, has no separate direct effect on price volatility since the effect of private information is averaged out by the law of large numbers. Hence, the model does not possess the endogenous amplification we seek. In addition, since supply shocks are never observed, the repeated inference causes prices to have infinite memory. In applications such as Brown and Jennings (1989) and Grundy and McNichols (1989), this property was used to explain the phenomenon of "Technical Analysis" defined as the traders' use of past prices, in addition to today's price, to form their demand. More broadly, the static model of asymmetric private information was used in widely diverse applications. One of the more celebrated application in macroeconomics are the Phelps (1970) and Lucas (1972) island models.

8.2.2. Dynamic Infinite Horizon Models

In the model of Section 8.2.1, trades can occur, but it is not a truly dynamic model. Extensions to infinite horizon were developed for many applications, and to get a sense of the issues involved, we review two very different applications. We start with Wang's (1994) study of trade volume.

Wang's (1994) aim is to overcome the no-trade theorems of REE and explain the volume of trade in asset markets. With REE perspective, his hypothesis is that trade is the result of asymmetric private information. Using his notation, he assumes that agent i maximizes expected utility over consumption flows $-E_t^i[\sum_{s=0}^{\infty} \beta^s e^{-\gamma c_{t+s}^i}|H_t^i]$ where expectations are conditioned on information of i. Wang (1994) assumes that there are two assets with payoff in consumption units. A riskless asset with infinitely elastic supply pays a constant rate r and where $R = 1 + r$. The second asset is a risky stock with a fixed supply set at 1, which pays a dividend D_t at date t. The law of motion of dividend is

$$D_t = F_t + \varepsilon_{D,t} \quad \text{where} \quad F_t = a_F F_{t-1} + \varepsilon_{F,t}$$

$(\varepsilon_{D,t}, \varepsilon_{F,t})$ are IID normally distributed, zero mean shocks. Here F_t is the persistent component of the dividend process and $\varepsilon_{D,t}$ is the transitory component. The structure of information is intended to ensure that a closed-form solution is possible. To that end Wang (1994) assumes that there are two types of investors. I-investors have perfect private information and observe F_t. The U-investors receive only a noisy signal about F_t in the form $S_t = F_t + \varepsilon_{S,t}$ where $\varepsilon_{S,t}$ are IID normal, zero mean shocks. Since all investors observe the dividends, the I-investors observe the persistent as well as the transitory components of dividends, whereas the U-investors observe neither.

In addition to the public asset, Wang (1994) assumes the I-investors have a private production technology that is risky and constant returns to scale. If they invest at t the amount I_t, they receive at $t+1$ the amount $I_t(1+r+q_{t+1})$, where excess return on the private technology q_{t+1} follows:

$$q_{t+1} = \Xi_t + \varepsilon_{q,t+1} \quad \text{and} \quad \Xi_{t+1} = a_{\Xi}\Xi_t + \varepsilon_{\Xi,t+1}$$

$(\varepsilon_{q,t+1}, \varepsilon_{\Xi,t+1})$ are IID normal, zero mean shocks. Expected excess return Ξ_t is observed only by the I-investors. This sharp information structure is called *hierarchical* since it requires one class of investors to permanently have inferior information. The economy's structure is common knowledge, and all agents are Bayesians with normal priors about parameters they do not know.

Two forces are used to explain the volume of trade. First, asymmetric information between the U- and the I-investors is measured by $\sigma_S^2 = \text{var}(\varepsilon_{S,t})$. If $S_t = F_t$, information about the stock is *symmetric* and $\text{var}(\varepsilon_{S,t}) = 0$. When $\sigma_S^2 > 0$ we have $S_t \neq F_t$ and information is asymmetric. Second, the private technology of the I-investors is unavailable to the U-investors. The effect of this factor on asset demands operates via the correlation between private excess returns q_{t+1} and dividends, measured by $\sigma_{D,q}$—the covariance of $\varepsilon_{D,t+1}$ with $\varepsilon_{q,t+1}$. To understand how this correlation impacts asset demands and trade, suppose that $\sigma_{D,q} \neq 0$ and Ξ_t increases leading I-investors to increase investments in private technology because expected return on such investments increased. But due to $\text{Cov}[\varepsilon_{D,t+1}, \varepsilon_{q,t+1}] \neq 0$, such increased investment changes the risk posture of their portfolio, calling for control of the risk by changing their investments in the publicly traded stock. If $\sigma_{D,q} > 0$, control of risk leads to lower investments in the stock; if $\sigma_{D,q} < 0$, it leads to increased demand for the stock.

The effect of asymmetric information about the private technology is thus due to the need of the I-investors to control their risk, whereas the U-investors are unable to distinguish between changes in F_t and Ξ_t. It shows that the setup of public and private technologies is crucial for Wang's (1994) results. Without private technology, just the asymmetry $S_t \neq F_t$ does not lead to trade, since in this case the uninformed investors deduce F_t from prices, so Wang's (1994) REE becomes fully revealing and we are back to no trade. With $\sigma_{D,q} \neq 0$ the price is linear in F_t and Ξ_t, and uninformed investors are "confused" and cannot deduce either one from the price. This confusion of the U-investors makes it impossible for them to determine the cause of price changes. The U-investors now use the history of the process to conduct a Kalman Filtering to form expectation of their unobserved F_t. In sum, Wang (1994) shows that the model generates trade due to the exogenous shocks F_t and Ξ_t, which cause time variability in the investment composition of the I-investors.

We pause briefly to examine the *causes* of price and volume volatility in models of noisy rational expectations equilibria. In the earlier models, equilibrium prices, such as

Eqs. 8.1a and 8.1b, vary only with exogenous shocks to supply. In Wang (1994) prices vary only in response to exogenous shocks F_t and Ξ_t. This result continues to hold in all other dynamic models of private information, such as He and Wang (1995). What is the effect of increased *diversity* of private information? In the earlier models, private information was so diverse that the law of large numbers was invoked and, as a result, private information had no effect on prices. If diverse private information is to be the cause of trade, we would expect that *increased* diversity of private information should *increase* the volume of trade. The problem is that Wang (1994, page 145) shows the opposite: *Increased* diversity of private information *decreases* the volume of trade. This results from the fact that rising diversity of private information causes uninformed traders to have rising difficulties in deducing from prices useful information needed for trade. We thus conclude that in noisy REE price volatility and volume of trade are caused by exogenous shocks, whereas *diverse private information does not cause or explain them*. We are thus back to the standard model without any endogenous amplification effect.

A second example is Woodford (2003), who revisits the Lucas (1972) model. It is motivated by the fact that Lucas (1972) explains transitory effects of monetary policy but fails to account for the observed fact that monetary disturbances have persistent real effects. Woodford assumes that agents are Dixit-Stiglitz monopolistic competitive price setters who select the nominal price of their product but cannot set the real price since they do not observe the aggregate price level and aggregate output. Although producers cannot observe the real price, their own output is determined by the real price. Aggregate *nominal* GNP is the exogenous state variable (e.g., determined by monetary policy). In equilibrium date t aggregate price level and aggregate output are functions of date t nominal GNP and of all higher-order market expectations (i.e., market expectations of market expectations of . . .) about it. As in Lucas (1972), agents cannot observe nominal GNP and receive private signals about it. Being rational, they learn from the available information and, as in Wang (1994), use a Kalman filtering procedure to learn about the unobserved state variable. With incomplete learning of the persistent exogenous nominal GNP, Woodford (2003) demonstrates persistent money nonneutrality.

Limited space prevents our discussing other applications. Examples include Townsend (1978, 1983); Amato and Shin (2006); Hellwig (2002, 2005) and Angeletos and Werning (2006), who study business cycles; Morris and Shin (2002, 2005), who study the transparency of monetary policy; Singleton (1987), who studies bond markets; Bacchetta and van Wincoop (2006); Hellwig, Mukherji, and Tsyvinski (2006), who study the volatility of foreign exchange rates; and many other policy-oriented papers using global coordination games. These applications use persistent belief heterogeneity to explain the behavior of market *aggregates*. However, why is it asymmetric private information that should provide a basis for heterogeneity? With asymmetric private information agents clearly make different forecasts. Therefore there is the temptation to assume private information to model diversity, and many authors have done just that. This is so common that for some, agents with different opinions are *synonymous* to agents with different private information. Such identification should be rejected. Private information is a very sharp sword that must be used with care. As we have seen, deducing information from prices is complicated and should be employed only when well justified. The virtual equivalence between belief diversity and private information is particularly wrong in macroeconomic applications when agents are assumed to have asymmetric private information with respect to aggregate variables such as interest rates or growth rate of GNP. We have serious doubts about the applicability of the private information paradigm to study asset market dynamics and will now pause to evaluate the results derived so far.

8.2.3. Is Asymmetric Information a Satisfactory Theory of Market Dynamics?

In questioning the use of asymmetric private information assumption, we recall that phenomena studied with private information include the volatility of asset price indices, interest rates, risk premia, foreign exchange rates, business cycles, and the like. In such models individual agents forecast aggregate variables. We thus break our query into two questions. First, does the model of asymmetric private information deliver a satisfactory mechanism of market volatility? Second, as a modeling device, is it reasonable to assume that economic agents have private information about such aggregate variables?

Starting with the first question, our answer is no. In all noisy REE with diverse and independent private signals, asymmetric private information has no impact either on price volatility or on volume of trade. In general, *increased* diversity of private information *decreases* volatility and volume of trade. In any noisy REE, all dynamic characteristics are fully determined by either the standard exogenous shocks such as dividends or exogenous “noise,” which is often questionable if it is unobserved. Since we aim to explain excess volatility of markets with mechanisms of *endogenous amplification*, it follows that the asymmetric private information paradigm does not offer such amplification; rather, it leads us simply back to the traditional causes of market dynamics. We examine the second question by outlining five problems raised by models of noisy rational expectations equilibria:

1. *What is the data that constitute “private” information*? For the case of individual firms, the nature of private information is clear, and we discuss it under 2. Now, if forecasters of GNP growth or future interest rates use private information, one must specify the data to which a forecaster has an exclusive access. Without it one cannot interpret a model’s implications, since all empirical implications, are deduced from restrictions imposed by private information. In reality it is difficult to imagine the data that constitute private information.
2. *Without correlation, private information explains little*. Even if some agents have some private information about some firms, an aggregate model may have no implications for market dynamics. To deduce any implications, private information has to be repetitive over time, correlated, and widespread. There is no *empirical* evidence for that. Indeed, all models of noisy REE assume private information to be IID distributed, and in that case private information has no effect on volatility, asset pricing, or any other dynamic characteristics.

3. *Asymmetric information implies a secretive economy*. Forecasters take pride in their models and are eager to make their forecasts public. Consequently, there are vast data files on market forecasts of many variables. In discussing public information, forecasters explain their interpretation of information, which is often framed as "their thesis." In contrast, an equilibrium with private information is secretive. Agents do not divulge their private information, since it would deprive them of the advantage they have. In such an equilibrium, private forecast data are treated as sources of *new information* used by all to update their own forecasts. The fact that forecasters are willing to reveal their forecasts is not compatible with private information being the cause of persistent divergence of forecasts.
4. *If private signals and noise are unobserved, how could common knowledge of the structure be attained, and how can we falsify the theory*? For us to deduce private information from public data, the *structure* of private signals must be common knowledge. For example, in Section 8.2.1, $x^i = V + \varepsilon^i$, where ε^i are pure noise, independent across i. A simple question arises: If these are not publicly observed signals, how does common knowledge come about? How does agent i know her own signal takes the form $x^i = V + \varepsilon^i$ and that the signal of k is $x^k = V + \varepsilon^k$? Also, if the crucial data of a theory are not observable, how can one falsify the theory? What, then, are the true restrictions of the theory?
5. *Why are private signals more informative than audited public statements*? Most results of models with private information are based on the assumption that private signals are more informative than public information. For example, in the model of Section 8.2.1, the public signal is $y = E(V)$, where V is unknown. Knowing y is inferior to knowing V. The private signals are $x^i = V + \varepsilon^i$ with ε^i IID and there is a continuum of traders. It is then assumed that "the market" aggregates the private x^i and learns V; hence equilibrium price is a function of the unknown V. This procedure raises two questions:
   - Why are private signals more precise than the professionally audited public statements?
   - Who does the aggregation and knows the IID structure needed for aggregation? If that agent is a neutral agent, why does he not announce V? Or if he is not neutral, he should be part of the model.

In summary, models with an asymmetric private information paradigm fail to explain the observed volatility and the assumptions made have questionable empirical basis. Hence, we must conclude that asymmetric private information is not a persuasive assumption for modeling market dynamics.

8.3. DIVERSE BELIEFS WITH COMMON INFORMATION: THE GENERAL THEORY

Rational expectations and behavioral economics have staked out two extreme positions in contemporary thinking. Under *rational expectations* people know all structural details needed for perfect forecasting, whereas under *behavioral economics* they are driven by psychological impulses that lead to irrational behavior. The theory of *rational belief* offers an intermediate concept of rationality that begins with imperfection of human knowledge, assumes people optimize given the limited knowledge they have, and concludes with a recognition that with imperfect knowledge *rational people make mistakes*. Rational mistakes may be magnified to a point where changing perceptions dominate public life and asset markets. This is the road to rational diverse beliefs and endogenous amplification that we explore. Before proceeding, we mention the papers of Harrison and Kreps (1978), Varian (1985, 1989), Harris and Raviv (1993), Detemple and Murthy (1994), Kandel and Pearson (1995), and Cabrales and Hoshi (1996). These, together with the early writers mentioned in the Introduction, recognized the importance of diverse expectations. We do not review them since they did not anchor the theory with a concept of rationality. In this connection we also note the *adaptive equilibrium* model of Brock and Hommes (1997, 1998) in which agents are boundedly rational.

8.3.1. A Basic Principle: Rational Diversity Implies Volatility

In contrast with asymmetric private information models, we now explain that theories of rational diverse beliefs provide a mechanism for endogenous amplification of volatility. Start by noting that any model discussed here assumes that agents do not know a true probability and hold diverse beliefs about it. This induces two basic questions. First, why do agents not know what they do not know? Second, what is their common knowledge basis? Before proceeding to these questions we clarify our notation. The symbols Π and m are reserved for special probability measures over infinite sequences to be defined shortly. Letters such as Q or P describe probabilities over *infinite sequences*. However, it is useful to think of a "belief" as a collection of conditional probabilities. Hence, instead of i's belief Q^i, we also use terms like *belief* or *date t belief* in reference to date t conditional probability $Q^i(\cdot|H_t)$, where H_t is date t information. Here *belief*, or *date t belief*, refers to a density, a joint distribution or transition function at t that is a component of Q^i. This abuse of notation avoids multiple definitions of belief and should not be confusing when the context is clear. We now return to the two questions.

Starting with the second question, although assumptions about what is common knowledge vary, one answer is general: It is *past data on observed variables*. There is a vector of observable economic variables $x_t \in \mathbb{R}^N$ over time with a data-generating process under a true unknown probability Π on infinite sequences. At t, agents have a long history of past observations $(x_0, x_1, \ldots, x_t)$, allowing rich statistical analysis. Given this data, all agents compute the same finite dimensional distributions of the data; hence all know the same empirical moments, *if they exist*. They then deduce from the data an empirical probability on infinite sequences denoted by m, which is then the empirical common knowledge of all. It will be seen later that m is stationary. Turning now to the first question, the basic cause of diverse beliefs is the fact that m and Π *are not the same*. We briefly explain why.

Our economy has undergone changes in technology and institutions, and these have had deep economic effects, rendering the data process $\{x_t, t = 1, 2, \ldots\}$ nonstationary. Although this means that the distributions of the x_ts are time dependent[1], a simple way to express it is to say that the data process constitutes a sequence of "regimes." But each "regime" is relatively short, with insufficient data to enable agents to learn each of these regimes with any degree of precision. Just to recall a sample of environments we have witnessed in recent years, note that before 1973 to 1979 we had never seen oil shocks and before the 1980s we had never encountered a S&L crisis of the size we had. The dot-com technology cycle of 1996 to 2001 resulted from the novelty of the Internet and the market's failure to predict the timing of its effect: Google was not even a factor then. Finally, the current subprime mortgage crisis results from the fact that the securitization it generated has never been seen before. One source of the crisis is the fact that there is no prior data with which we can predict with accuracy the effect of lower home prices on the rate of default of these securities. In short, it is impossible to learn the unknown probability Π. The stationary probability m (if it exists) is then just an average over an infinite sequence of changing regimes. It reflects long-term frequencies, but it is not the true probability under which the data are generated. Belief diversity arises when agents believe that m is not the truth and the past is not adequate to forecast the future. All surveys of forecasters show that subjective judgment contributes more than 50% to final forecasts (e.g., Batchelor and Dua, 1991). Individual subjective models are thus the way agents express their interpretation of the data. Being common knowledge, the stationary probability m is a reference point for any rationality concepts.

Is it rational to believe that m is the truth? Those who believe that the economy is stationary hold this belief. The theory of *rational beliefs* (see Kurz, 1994, 1997a) defines an agent to be *rational* if her model cannot be falsified by the data m. The theory then has a simple implication that addresses the crucial question of dynamics. It says that an agent's date t belief *cannot be constant or time invariant* unless she believes the stationary probability m is the truth. To see why, consider an agent who holds a constant belief at date t (e.g., time-invariant transition function), which is different from the one implied by m. Since it is constant, the time average of his belief is not m. Since m is the time average in the data, this *proves* that the agent is irrational. In simple terms, it is irrational to be permanently optimistic or pessimistic relative to m. By implication, if a rational agent's belief persists in disagreeing with m, then such a belief *must fluctuate over time around* m. Hence rationality induces dynamic fluctuations on the level of individuals! Now assume that a population exhibits a persistent diversity of beliefs across agents. It implies that most hold beliefs that disagree with m. But then, we have seen that this requires individual beliefs to fluctuate over time. Finally, for an aggregate effect of beliefs, we only add the empirically established fact (see Section 8.3.4, "Individual States of Belief") that individual beliefs are correlated, and this leads to the conclusion that *rational diversity implies aggregate dynamics*.

[1]The technical definition of "nonstationary" that we use requires the process to be time dependent, and this is the customary terminology in ergodic theory and stochastic processes. It is different from the use of this term in the time series literature, which requires the process to have infinite variance.

Diversity of beliefs without private information is often questioned by asking how agents could be wrong and rational at the same time. The idea that rational agents may be wrong relative to an unknown truth is a central component of the theory. Indeed, when rational agents hold diverse beliefs while there is only one true probability law of motion, *then most agents are wrong most of the time*. Since agents' beliefs are correlated, the average market belief is also often wrong, and this is the source of endogenous propagation of market risk and volatility, called *endogenous uncertainty* by Kurz (1974) and Kurz and Wu (1996). Sections 8.3.2 and 8.3.3 provide a precise outline of these ideas.

8.3.2. Stability and Rationality in a General Nonstationary Economy

In a stationary economy, joint probabilities are time invariant and the *Ergodic Theorem* holds: Time averages equal expected values under the true probability, and this probability is deduced from relative frequencies of events. In such environments, the empirical distribution reveals the truth, and since human history is long, agents learn the structure from the data. The fact is that the real data-generating mechanism is not stationary, and history, matters. The question is then how can we discuss rationality and empirical distributions in a complex environment? What is the regularity we can use for analytical evaluation? The answer leads to a definition of *rational beliefs*, which we outline now. The development in this section is based on the material in Kurz (1994).

Let $x_t \varepsilon X \subseteq \mathbb{R}^N$ be a vector of the N observables and let $x = (x_0, x_1, x_2, \ldots)$. Let a future sequence from t on be $x^t = (x_t, x_{t+1}, x_{t+2}, \ldots)$, hence $x^0 \equiv x$. The history to date t is defined by $(x_0, x_1, x_2, \ldots, x_t)$. Let X^∞ be the space of infinite sequences x, and let $\mathfrak{B}(X^\infty)$ be the Borel σ-field of X^∞. Events in $\mathfrak{B}(X^\infty)$ are denoted by letters U, S, T, and so on. For an event $S \varepsilon \mathfrak{B}(X^\infty)$ define the sets $S^{(k)} = \{x | x^k \varepsilon S, k \geq 0\} \equiv$ the event S occurring k periods later. Clearly, $S = S^{(0)}$.

Definition 8.1. A set $S \varepsilon \mathfrak{B}(X^\infty)$ is said to be invariant if $S^{(1)} = S$.

Definition 8.2. A probability Π on $(X^\infty, \mathfrak{B}(X^\infty))$ is said to be *ergodic* if for any invariant set S we have $\Pi(S) = 1$ or $\Pi(S) = 0$.

Throughout the discussion we assume ergodicity so as to simplify the exposition. Under this assumption we develop the basic equivalence theorem, which is the basis of the theory of rational belief. We start with the concept of *statistical stability*. For any finite dimensional set $U \varepsilon \mathfrak{B}(X^\infty)$ define the following time average:

$$m_n(U)(x) = \frac{1}{n} \sum_{k=0}^{n-1} 1_U(x^k) = \left\{ \begin{array}{l} \text{The relative frequency that } U \text{ occurred} \\ \text{among } n \text{ observations since date 0} \end{array} \right\}$$

where

$$1_U(y) = \begin{cases} 1 & \text{if } y \in U \\ 0 & \text{if } y \in U. \end{cases}$$

Although the set U is finite dimensional, it can be a complicated set. For example:

$$U = \left\{ \begin{array}{l} \text{State 1 today} \leq \$1, \text{State 6 next year} \geq 16, \\ 2 \leq \text{State 14 five years later} \leq 5 \end{array} \right\}$$

Definition 8.3. (Property 1) A probability Q on $(X^\infty, \mathfrak{B}(X^\infty))$ is statistically stable if for each cylinder (i.e., finite dimensional) set $U \in \mathfrak{B}(X^\infty)$:

$$(1) \ \lim \ m_n(U)(x) \ = \ m^Q(U)(x) \text{ exists} \qquad Q \text{ a.e.}$$

Since by ergodicity $m^Q(U)(x)$ is Q a.e. independent of x, we add the notation:

$$(2) \ m^Q(U)(x) = m^Q(U) \qquad Q \text{ a.e.}$$

The restriction to finite-dimensional sets results from the fact that we have only finite data; hence we cannot ascertain whether an infinite dimensional event actually occurred.[2] The first property of *statistical stability* means that the process satisfies the conclusions of the ergodic theorem, although it does not satisfy the standard condition of stationarity used to prove it. Hence, data generated by a stable process have the property that relative frequencies of events converge and all finite moments exist.

The limits in Definition 8.3 might not exist for infinite dimensional sets. Hence the set function defined by the preceding limits is not a probability. However, standard extension theorems permit extension of m^Q to a probability measure on $(X^\infty, \mathfrak{B}(X^\infty))$. To avoid multiple notation, we do not distinguish between these two set functions and denote the extension m^Q as well. We also have:

Theorem 8.1. *m^Q is unique and stationary. It is thus called the stationary measure of Q.*

We now introduce the concept of *weak asymptotic mean stationary* probability measure.

Definition 8.4. (Property 2) A probability Q on $(X^\infty, \mathfrak{B}(X^\infty))$ is *weak asymptotic mean stationary* if for each cylinder set $U \in \mathfrak{B}(X^\infty)$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q(U^{(k)}) = m_Q(U)$$

exists.

[2]This in contrast to the theory of checking rules (e.g., Dawid, 1985), which can be effectively implemented only "at infinity" after one has infinite number of observations.

Averaging probabilities of future events $U^{(k)}$ is like averaging over date t beliefs over time. Such average is required to converge. The set function $m_Q(S)$ is not a probability, but by extension one deduces a probability on $(X^\infty, \mathfrak{B}(X^\infty))$ that is unique and stationary. Again, use the same notation for the extension. Thus, if *Q is weak and asymptotic mean stationary*, then m_Q is the probability on $(X^\infty, \mathfrak{B}(X^\infty))$ induced by Property 2. A key result of the theory can now be stated:

Theorem 8.2. *Properties 1 and 2 are equivalent and $m^Q(S) = m_Q(S)$ for all events $S \in \mathfrak{B}(X^\infty)$.*

Theorem 8.2 says a data-generating process that is statistically stable with probability Q has an associated stationary probability $m^Q = m_Q$, *which can be computed in two different ways*. First, it can be deduced empirically from relative frequencies computed from the data generated by the process. Second, it can be deduced *analytically* by averaging probabilities of each event over future dates. How do we use this equivalence to define the rationality of belief?

The data process under Π is nonstationary and we now assume it is *statistically stable and ergodic*. This is a reasonable assumption because in our economy additional data increase accuracy of statistical analysis and moments exist. Agents do not know Π and compute empirical frequencies from the data. Using extension, they discover from the data[3] the probability m induced by the dynamics under Π. We reserve the notation $m(U)$ for the limit of the empirical frequencies under Π; hence, under our convention $m \equiv m^\Pi$. This is the stationary measure of Π and we refer to it as *the* stationary measure, or the empirical distribution. Although agents have only finite data, *we assume they actually know the limits $m(U)$ in Definition* 8.3, an assumption made for simplicity.[4] All have the same data; hence there is no disagreement among agents about the probability m.

If the economy was actually stationary, $m = \Pi$ *but agents could not know this fact*. The fact is that $m \neq \Pi$ and we seek a concept of rationality of belief that requires an agent to hold a belief that is not contradicted by the empirical evidence represented by the probability m.

[3]We always have only finite data that enable agents to compute at date t only $m_t(U)(x)$. This depends on the data used and with time $m_t(U)(x)$ converges to the limit probability. The assumption made in the text is that the data sequence is very long and the probability m is simple enough (i.e., Markov with short memory) that with finite data agents can obtain a good approximation for the limit measure m. The assumption that agents know the limits is very strong and should not be interpreted to mean we assume that agents have an infinite sequence of observations, since in that case agents will consider not only limits on sequences but also limits on all infinitely many possible subsequences. With finite data we can observe only a finite number of subsequences at all dates; for this reason we do not incorporate restrictions that would be implied by limits on subsequences. We also note that in the nonergodic case the data requirement is greater, since then we need data for many alternative sequences x with different starting points, but the basic theory remains unchanged. For details see Kurz (1994).

[4]The assumption that the limit in Definition 8.3 is known to the agents is made to avoid the complexity of an approximation theory. Without this assumption the diversity of beliefs would be increased due to the diverse opinions about the finite approximations that would be made by different agents. In this context we also mention that the assumption of *ergodicity* is also not needed and is not made in Kurz [1994].

Definition 8.5. A probability belief Q is said to be *a rational belief relative to m* if

1. Q is a *weak asymptotic mean stationary* probability on $(X^{\infty}, \mathfrak{B}(X^{\infty}))$,
2. $m_Q(S) = m(S)$ for all events $S \in \mathfrak{B}(X^{\infty})$.

A *rational* agent holds a belief Q that is compatible with the empirical evidence m if:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q(S^{(k)}) = m_Q(S) = m(S)$$

for all cylinder sets $S \in \mathcal{B}(X^{\infty})$.

The first of these conditions is checked analytically by averaging date t beliefs and these should average to a probability m_Q. The second condition requires that $m_Q = m$ where m is common knowledge in the market.

Three implication of Definition 8.5 are notable:

1. Π is a *rational belief* and hence rational expectations are also rational beliefs.
2. m is a *rational belief*, although it is not the truth, since $m \neq \Pi$.
3. A rational belief is generally not equal to Π, showing that agents are wrong and rational at the same time. Agents holding *rational beliefs* disagree more about the short-term forecasts of variables than about forecasts of long-run averages of these same variables. For example, we expect greater disagreement about the forecast of one-year inflation or output growth than about averages of these variables over the next ten years.

Applications of these concepts requires a simplification of the general rationality conditions in Definition 8.5. We thus start with two examples that use a method due to Nielsen (1996).

EXAMPLE 8.1

Agents observe a black box generating numbers x_t in $\{0, 1\}$ without long-term serial correlation between x_t and x_{t+k} all $k > 0$. Using a long dataset, they find the mean is 0.5. The probability m is then the probability measure induced by a sequence of IID random variables on $\{0, 1\}$ with probability of 1 being 0.5. If the box contains a single coin, the x sequence has an empirical distribution of an IID fair coin, which is the truth. What other processes generate the same empirical measure? As an example, consider a belief in a two-coin family that uses *the realization* of an IID sequence of random variables $g_t, t = 0, 1, \ldots$ in $\{1, 2\}$ with probability of 1 being, say, 1/3. A sequence is produced in advance, and is therefore known to the agent. Pick an infinite sequence $g = (g_0, g_1, g_2, \ldots)$. These realizations of $g_t = 1$ or $g_t = 2$ are treated as *fixed parameters* of a new process. It is defined by a process $\{v_t \in \{0, 1\}, t = 0, 1, \ldots\}$ with two IID coins in the box that appear at different times, depending on the $g = (g_0, g_1, \ldots)$. v_t is then a sequence of independent random variables of the form

$$P\{v_t = 1\} = \begin{cases} 0.60 & \text{if} \quad g_t = 1 \text{ (coin type 1)} \\ 0.45 & \text{if} \quad g_t = 2 \text{ (coin type 2)} \end{cases} \qquad \textbf{(8.9a)}$$

Since $(1/3)(0.60) + (2/3)(0.45) = 0.50$, the empirical distribution is the same as m for almost all g and v. It is easy to see that instead of two possible "regimes" we could have an infinite number of regimes. Note that we have not even specified the true probability Π.

EXAMPLE 8.2

The data x reveal that the empirical probability m is represented by, say, a 10-dimensional matrix M. Again, select an IID sequence of random variables $g_t, t = 0, 1, \ldots$ in $G = \{1, 2\}$ with a probability of 1 being, say, α. Next, construct a joint probability on infinite sequences (g, x) on the space $((X \times G)^\infty, \mathcal{B}((X \times G)^\infty))$, assuming the joint (g, x) process is a stationary Markov process on a 2×10 transition matrix. Suppose that over these 20 (g, x) states the matrix takes the form

$$F = \begin{bmatrix} \alpha F_1 & (1-\alpha)F_1 \\ \alpha F_2 & (1-\alpha)F_2 \end{bmatrix} \tag{8.9b}$$

States in the upper part of F are of $g = 1$ and in the lower part of $g = 2$; hence, the marginal of F on g is the IID distribution $(\alpha, (1-\alpha))$. Now, when $g_t = 1$ the probabilities assigned to x_{t+1} are given by F_1, and when $g_t = 2$ the probabilities of x_{t+1} are given by F_2. The nonstationary probability we seek is represented by the *conditional probability* of F given g. What is the empirical distribution under this conditional probability? Assuming Theorem 8.2 applies we compute the mean probability. With probability α the matrix F_1 is used, and with probability $(1-\alpha)$ the matrix F_2 is used. Hence the stationary distribution implied by the conditional probability of F on g is the expected value $\alpha F_1 + (1-\alpha)F_2$. It follows from Definition 8.5 that F is a *rational belief* if $M = \alpha F_1 + (1-\alpha)F_2$.

Note that in Example 8.1 we define a *rational belief* by knowing in advance the infinite sequence $g = (g_0, g_1, \ldots)$. In Example 8.2 we select only the matrix F so that at date t our forecaster knows his type g_t but is uncertain about g_{t+1}. This approach is the one we follow in the rest of this chapter.

8.3.3. Belief Rationality and the Conditional Stability Theorem

By its own nature, nonstationarity is difficult to describe since it entails a potential infinite variability. Examples 8.1 and 8.2 reveal a simple method to describe nonstationary probability of a real system or as an agent's belief. The question is, how general are such systems, and is there a general principle that generalizes Examples 8.1 and 8.2 to stable but nonstationary systems? The *conditional stability theorem* (Kurz and Schneider, 1996) gives the answer. We explain it now.

Although the theorem holds for general dynamical systems we avoid excessive formalism and discuss here only stochastic processes with probability measures over infinite sequences under which the data are generated.[5] We characterize a family

[5]In the language of Ergodic Theory, the theorem applies to general dynamical systems, but we confine our attention only to dynamical systems under a shift transformation. Since we approach the problem from the point of view of stochastic processes, we avoid the notation of Ergodic Theory altogether.

of probabilities described by sequences of parameters in $G \subseteq \mathbb{R}^L$. These parameters represent structural change and could be thought of as a sequence of "regimes" over time. To do that, consider *joint* sequences of $(x_t, g_t) \in X \times G$ generated under a probability P over the space $((X \times G)^\infty, \mathcal{B}((X \times G)^\infty))$. The following theorem is stated under the assumption that P is stationary, although it is sufficient that it be stable. Now, let P_g be a regular *conditional probability* of x given g. That is:

$$P_g(S) : G^\infty \times \mathcal{B}(X^\infty) \to [0, 1] \quad g \in G^\infty \quad S \in \mathcal{B}(X^\infty) \tag{8.9}$$

such that for each $S \in \mathcal{B}(X^\infty)$, $P_g(S)$ is a measurable function of g and for each g, $P_g(\bullet)$ is a probability on $(X^\infty, \mathcal{B}(X^\infty))$. We now consider the data $(x_t, t = 0, 1, 2, \ldots)$ as being generated under the conditional probability P_g parametrized by g, where we consider g_t as the parameters of the regime in place at date t. The question we ask is, under what conditions is the data-generating system under the probability $P_g(\bullet)$ stable for almost all parameter sequences g?

Before proceeding, we pause and ask how we should think of the joint system. The joint process on data and parameters could be considered in two ways. One is as a description of the way our world evolves, inducing statistical regularity of the data-generating process. The joint system is then a true unobserved law of motion of our economy, the $g_t \in G$ are *unobserved parameters*, and the statistical properties of the parameters are interrelated with the statistical properties of the data. Both arise from the stability of the joint system. Or else, which is the way we use it here, the joint process is a *model* that a rational agent uses to formulate his belief. The parameter g_t then pins down the state of belief or the agent "type" at date t.

To proceed, we need two technical definitions. Let P_X be the *marginal measure* of P on $(X^\infty, \mathcal{B}(X^\infty))$ and P_G be the marginal of P on $(G^\infty, \mathcal{B}(G^\infty))$ defined by

$$P_X(S) = P(S \times G^\infty) \quad \text{for all} \quad S \in \mathcal{B}(X^\infty)$$
$$P_G(Y) = P(X^\infty \times Y) \quad \text{for all} \quad Y \in \mathcal{B}(G^\infty)$$

Our perspective is then simple. The joint is a process on data and parameters under P, but the data $(x_t, t = 0, 1, 2, \ldots)$ is generated under the conditional probability P_g parametrized by g.

Theorem 8.3. *Conditional Stability Theorem (Kurz and Schneider 1996)—Suppose* G *is countable and the probability* P *on* $((X \times G)^\infty, \mathcal{B}((X \times G)^\infty))$ *is stationary and ergodic. Then:*

- *The conditional probability* P_g *is stable and ergodic for* P *almost all* g. *The stationary measure of* P_g *is denoted by* m^{P_g}.
- m^{P_g} *is independent of* g P *almost all* g.
- $m^{P_g} = P_X$.

A sufficient condition for stability and ergodicity of P_g is then the stability and ergodicity of P. In addition to the stability of the conditional probability P_g, we also have the result that the stationary measure of P_g is the marginal measure P_X. It is well known that for all $S \in \mathcal{B}(X^\infty)$ we have

$$P(S \times G^\infty) = \int_{G^\infty} P_g(S) P_G(dg)$$

Hence, P_X is computed by averaging the conditional probability P_g over frequencies at which it is used, as is the case in Examples 8.1 and 8.2.

The theorem defines a general family of probabilities that are *rational beliefs relative to a known* m. That is, let the data be generated under a stable, ergodic, but unknown probability Π with a stationary measure m. Now an agent formulates a joint process of (x, g) under probability P, which induces, with parameters g, a belief P_g on data sequences x. The question is then under what condition is P_g a stable and ergodic rational belief? Theorem 8.3 tells us that if the joint P is stationary and ergodic, then P_g is stable and ergodic with a stationary measure satisfying $m^{P_g} = P_X$. It is a *rational belief* if the joint satisfies $P_X = m$. The joint is then a model an agent uses to formulate his belief. Our development that follows is based on this way of constructing beliefs relative to a known empirical probability m. But since we also assume Gaussian processes with a continuum of states, we comment on Theorem 8.3's condition that G is countable. In general, Theorem 8.3 is false for continuum state space without more restrictions. For Gaussian processes the theorem holds and we can give a direct proof. A more general theorem is given by Nielsen (2007) for Harris processes.

Theorem 8.3 offers a tractable way to describe beliefs about general asset structures. To simplify exposition, we concentrate in the rest of this chapter on a simple asset structure. To that end we postulate an exogenous environment in which there is a single risky asset or a single risky portfolio of assets paying an exogenous risky payoff $\{D_t, t = 1, 2, \ldots\}$ with a nonstationary and unknown true probability. We assume that the available long history of the data reveals that the empirical distribution of the D_t's constitute a Markov process with transition

$$D_{t+1} = \mu + \lambda_d(D_t - \mu) + \rho^d_{t+1} \quad \rho^d_{t+1} \sim N(0, \sigma^2_d)^6$$

and unconditional mean μ. Let $d_t = D_t - \mu$; then the process $\{d_t, t = 1, 2, \ldots\}$ is a zero mean, nonstationary with unknown true probability Π and empirical probability m. Hence, $\{d_t, t = 1, 2, \ldots\}$ has an empirical distribution that implies a transition function of the first-order Markov process

$$d_{t+1} = \lambda_d d_t + \rho^d_{t+1} \quad \rho^d_{t+1} \sim N(0, \sigma^2_d) \qquad \textbf{(8.10a)}$$

Since the implied stationary probability is denoted by m, we write $E^m[d_{t+1}|d_t] = \lambda_d d_t$.

[6]As will shortly be explained, in many applications the dividend or payoff D_t grows without bound, does not have a finite mean, and has growth rates that have an empirical distribution characterized by a stationary transition of a Markov process. The same applies to other statistically stable processes with trends, in which case the concept of stability is applied to growth rate data, not to the absolute quantities.

We also review papers that assume finite state spaces. In simulation work which study volatility it is assumed $D_{t+1} = v_{t+1} D_t$, where v_t is the random *growth rate* of dividends, which is a Markov process over a finite state space. In studying the equity risk premium Mehra and Prescott (1985) assume v_t takes two values. They find that the long-term empirical distribution is represented by a stationary and ergodic Markov process over a state space that reflects extreme business cycle states of "recession" and "expansion." They estimate the transition matrix of the two states to be

$$\begin{bmatrix} \varphi & 1-\varphi \\ 1-\varphi & \varphi \end{bmatrix} \quad \varphi = 0.43 \tag{8.10b}$$

Is the stationary model of Eq. 8.10a or 8.10b the true process? Those who believe the economy is stationary would accept Eq. 8.10a or 8.10b as the truth. Most do not believe past empirical record is adequate to forecast the future, and this leads to nonstationary and diverse beliefs. The problem is, then, how do we describe an equilibrium in such an economy? The belief structure is our next topic.

8.3.4. Describing Individual and Market Beliefs with Markov State Variables

The approach taken by Theorem 8.3 raises a methodological question. In formulating an asset-pricing theory, do we need to describe in detail each agent's model? Are such details needed for a study of price dynamics? Although an intriguing question, we suggest that such details are not needed. To describe an equilibrium, all we need is to specify how beliefs affect agents' perceived stochastic transition of state variables. Once specified, Euler equations are well defined and market clearing leads to equilibrium pricing. Theorem 8.3 leads to this approach by proposing to treat individual beliefs as state variables, *generated within the economy*.[7] This is the approach we now explain.

Start with the fact that agents who hold heterogeneous beliefs are willing to reveal their forecasts when surveyed. We thus assume that distributions of individual forecasts are publicly observable. An individual's belief is described with a *personal* state of belief that uniquely pins down his perception of the transition to next period's state variables. It follows that personal state variables and economy-wide state variables are not the same. A personal state of belief is the same as any other state variables in an agent's decision problem but can also be interpreted as defining the "type" of agent who is uncertain of her future belief type but knows the dynamics of her belief state. The distribution of belief states is then an economy-wide state variable. Endogenous variables depend on the economy's state variables. Hence, moments of the market distribution

[7] In using Theorem 8.3 there are two possible approaches that can be taken. The first is based on Nielsen (1996), who treats the infinite sequence g_t of parameters as fixed and known to each agent in advance, as in Example 8.1. Hence, in Example 8.1 a belief is a Dirichlet distribution in $(G^\infty, \mathcal{B}(G^\infty))$. In this chapter we follow the developments in Kurz and Motolese (2001), Kurz, Jin, and Motolese (2005a, 2005b), and Kurz and Motolese (2007), who treat the sequence g_t as state variables that define the belief of an agent or identify his type. These papers assume that at date t an agent does not know his own type at date $t+1$.

of beliefs may have an effect on endogenous variables such as prices. Also, in a large economy, an agent's *anonymity* means that a personal state of belief is perceived to have a negligible effect on prices and is assumed not to be public. Some papers assume an exponential utility that results in equilibrium endogenous variables depending only on the *mean market belief*. Finally, in the following discussion the set of agents is implicit and not specified: It might be finite or infinite. It is specified when needed.

Individual States of Belief

We introduce agent i's *state of belief* at t, g_t^i, which pins down his transition functions. Apart from "anonymity," we assume that agent ℓ knows his own g_t^ℓ and the *distribution* of the g_t^i over the agents for all past dates $\tau \leq t$. This last assumption is justified by the fact that an infinite horizon economy consists of a sequence of decision makers. An agent knows his states of beliefs but does not know the states of belief of *all* his own specific predecessors. Past belief distributions are public information, since samples of g_t^i are made public. We specify the dynamics of g_t^i by

$$g_{t+1}^i = \lambda_Z g_t^i + \rho_{t+1}^{ig} \quad \rho_{t+1}^{ig} \sim N(0, \sigma_g^2) \tag{8.11}$$

where ρ_t^{ig} are correlated across agents, reflecting correlation of beliefs across individuals. The *state of belief* is a central concept and Eq. 8.11 is taken as a primitive description of type heterogeneity. One can, however, deduce Eq. 8.11 from more elementary principles (see the next subsection).

How does g_t^i pin down the stochastic transition? In various models agent i's *perception* of date t distribution of d_{t+1} (denoted by d_{t+1}^i) is described by using the belief state as follows:

$$d_{t+1}^i = \lambda_d d_t + \lambda_d^g g_t^i + \rho_{t+1}^{id} \quad \rho_{t+1}^{id} \sim N(0, \hat{\sigma}_d^2) \tag{8.12}$$

The assumption that $\hat{\sigma}_d^2$ is the same for all i is made only for simplicity. An agent who believes the empirical distribution is the truth expresses it by $g_t^i = 0$. It follows that given information H_t, the state of belief g_t^i measures the deviation of her forecast from the empirical stationary forecast

$$E^i[d_{t+1}^i | H_t, g_t^i] - E^m[d_{t+1} | H_t] = \lambda_d^g g_t^i \tag{8.13}$$

Eq. 8.13 shows how g_t^i is measured in practice. For any x_t, publicly available data on i's forecasts of x_{t+h}, measure $E^i[x_{t+h} | H_t, g_t^i]$, where h is the forecast horizon. To estimate the difference in Eq. 8.13 one then uses standard techniques such as Stock and Watson (2001, 2002, 2005) to compute the stationary forecast $E^m[x_{t+h} | H_t]$. Average market belief is then computed by averaging the left side of Eq. 8.13 over agents. Fan (2006) and Kurz and Motolese (2007) offer examples of such construction. Figures 8.1, 8.2, and 8.3 illustrate the time series of *average market belief* with horizon of six-month for the six-month Treasury Bill rate, for percent change in the GDP

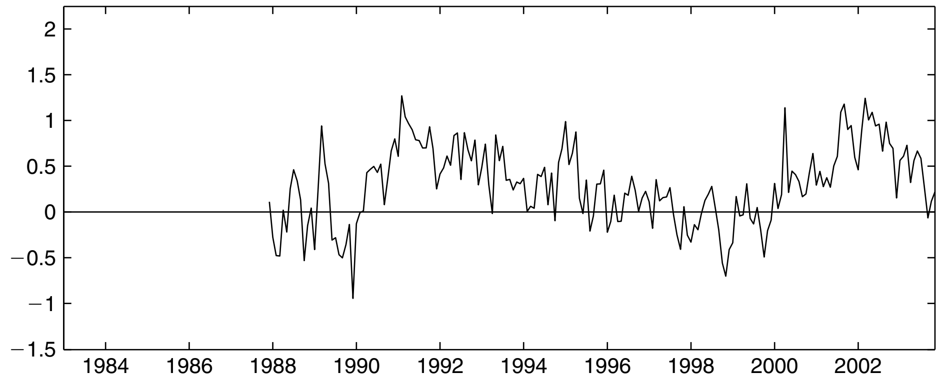

FIGURE 8.1 Six-month Treasury Bill rate: six-months-ahead market belief.

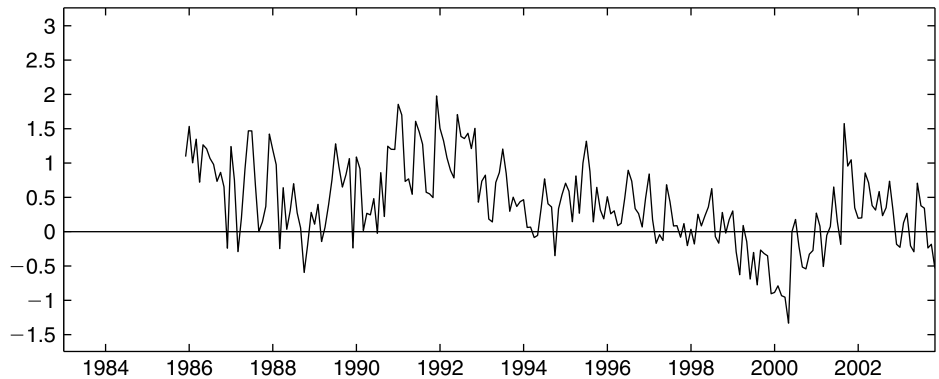

FIGURE 8.2 GDP deflator for inflation rate: two-quarters-ahead market belief.

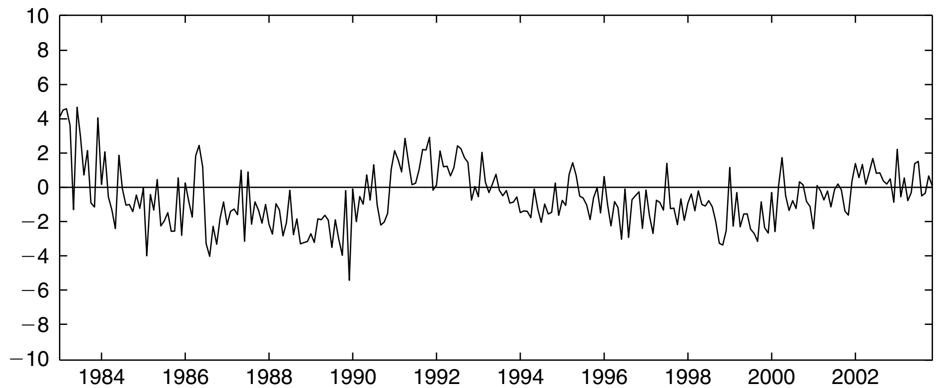

FIGURE 8.3 Month-over-month, annualized growth rate of industrial production: six-months-ahead market belief.

deflator (measuring inflation) and growth rate of industrial production. For each variable, the average forecasts are given by the blue-chip financial forecasts, whereas $E^m[x_{t+h}|H_t]$ is computed by Kurz and Motolese (2007) using methods of Stock and Watson (2001, 2002, 2005).

In Figures 8.1 and, 8.3, average market belief fluctuates around zero as predicted by the theory, *even during the short period at hand*. In Figure 8.2 this pattern is not exactly maintained by market belief about inflation due to persistent deviations of inflation forecasts from the normal pattern during the 1980s and early 1990s. Over a longer horizon, the pattern of fluctuations around zero is restored. All three figures are compatible with the Markov property assumed in Eq. 8.11 (and later in Eq. 8.15).

Note that since belief variables arise from structural change, g_t^i in 1900 has nothing to do with the one in 2000: They reflect different social and technological environments. Also observe that a belief g_t^i is not "information" about unknown structural parameter; rather, it describes the *opinion* of agent i. Hence, agents do not treat individual beliefs of others as *information*, and even if they observed them they do not deduce from them anything about unknown parameters.

Deducing the Dynamics of Individual Belief from Bayesian Inference

Although Eq. 8.11 is a primitive, we can deduce it from elementary principles. There are many ways to do this; we review the approach of Kurz (2007). He shows how subjective interpretation of data arises from public *qualitative* information, which always accompanies the release of quantitative data. To highlight this idea, note first that a Markov property in Eq. 8.11 is not surprising since Bayesian posteriors have a Markov form. This is not sufficient, since if agents knew that the dividend process has an unknown parameter b, which takes the form

$$d_{t+1} - \lambda_d d_t = b + \rho_{t+1}^d \qquad \rho_{t+1}^d \sim N\left(0, \frac{1}{\beta}\right)$$

then a Bayesian posterior would be a *convergent* belief sequence. Hence, the key object is to explain where the *random term* in Eq. 8.11 comes from.

From Eqs. 8.16a and 8.16b, agents know λ_d, and Kurz (2007) assumes they also know that under the true probability Π the transition of d_t is

$$d_{t+1} - \lambda_d d_t = b_t + \rho_{t+1}^d \qquad \rho_{t+1}^d \sim N\left(0, \frac{1}{\beta}\right) \tag{8.14}$$

b_t are the unknown, exogenous time-varying mean values of $d_{t+1} - \lambda_d d_t$ and hence g_t^i are beliefs about b_t. Since there is no universal method to learn a sequence of parameters, Kurz (2007) outlines a Bayesian updating procedure that is supplemented by subjective estimates of $d_{t+1} - \lambda_d d_t$, *which are based on qualitative public information*. We start with the qualitative data.

Qualitative data about all aspects of our economy are provided at all times, and financial markets pay a great deal of attention to them. Profit is just one number in a financial report that covers many additional issues. Firms may announce new research projects, new organizational structures, or new products. Qualitative data are rarely comparable over time. For example, when a firm starts research into a new topic, no past data exist on it. Qualitative information is modeled by Kurz (2007) in the form of qualitative *statements that can potentially impact future profits*. The list of statements may change with time and the impact on profits may be positive or negative. For each statement, a realization at $t + 1$ may be a "success" or "failure" in its effect on profits. Agent i has subjective maps from the list of potential successes or failures to potential future value of $(d_{t+1} - \lambda_d d_t)$. Finally, conditional on the statements, agent i attaches subjective probabilities to *vectors* of successes or failures, and by taking expected value she makes a subjective estimate Ψ_t^i of $(d_{t+1} - \lambda_d d_t)$. Ψ_t^i varies with time since new statements are made each date. Because the long-term average of $(d_{t+1} - \lambda_d d_t)$ is zero, rationality requires the Ψ_t^i to be zero mean random variables. We now examine an alternative Bayesian updating procedure for estimating the same quantity.

Kurz (2007) starts the updating process by assuming that β in Eq. 8.14 is known. At first decision date t (say, $t = 1$), an agent has two pieces of information. He observes d_t and receives public qualitative information with which he assesses Ψ_t^i. Without Ψ_t^i his prior belief at $t = 1$ is normal with mean b. However, to start the process, he uses *both sources* to form a prior belief $E_t^i(b_t|d_t, \Psi_t^i)$ about b_t (used to forecast d_{t+1}). However, the changing parameter b_t leads to a problem. When $d_{t+1} - \lambda_d d_t$ is observed, Agent i updates his belief to $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$[8] in a standard Bayesian procedure. But he needs an *estimate of* b_{t+1}, not of b_t. Hence, his problem is how to go from $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$ to a prior of b_{t+1}. Without new information, his belief about b_{t+1} is unchanged and $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$ would be the new prior of b_{t+1}. This is his first estimate of b_{t+1}. Next the agent observes the qualitative information released publicly before trading at $t + 1$, which provides an alternate subjective estimate Ψ_{t+1}^i of b_{t+1}. Now the agent has two independent sources for belief about b_{t+1}: $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$ and Ψ_{t+1}^i. Kurz (2007) now assumes:

Assumption 8.1. Agent i uses a subjective probability τ to form date $t + 1$ prior belief defined by

$$E_{t+1}^i(b_{t+1}|d_{t+1}, \Psi_{t+1}^i) = (1 - \tau)E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i) + \tau\Psi_{t+1}^i \quad 0 < \tau \leq 1$$

At $t = 1$ it was assumed that the initial prior mean is b, hence for consistency, if Ψ_t^i is Normal, then

$$b_1 \sim N\left((1 - \tau)b + \tau\Psi_1^i, \frac{1}{\vartheta}\right)$$

for some ϑ.

This assumption is the element that permits $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$ to be upgraded into a prior belief at date $t + 1$, $E_{t+1}^i(b_{t+1}|d_{t+1}, \Psi_{t+1}^i)$, before d_{t+2} is observed. The following result is then shown:

[8]We use the notation $E_t^i(b_t|d_t, \Psi_t^i)$ for the *prior* belief at date t about the unknown parameter b_t used to forecast d_{t+1}. We then use the notation $E_{t+1}^i(b_t|d_{t+1}, \Psi_t^i)$ for the posterior belief about b_t given the observation of d_{t+1}. Assumption 8.1 will then use this posterior belief as a building block in the formation of the new *prior* $E_{t+1}^i(b_{t+1}|d_{t+1}, \Psi_{t+1}^i)$.

Theorem 8.4. (Kurz, 2007): *Suppose* $\Psi_t^i \sim N(0, \frac{1}{v})$, *IID and Assumption (8.1) hold. Then there exists a constant* $0 < \kappa < 1$ *such that for large t the prior belief* $E_t^i(b_t|d_t, \Psi_t^i)$ *is a Markov random variable, and by identifying* $g_t^i = E_t^i(b_t|d_t, \Psi_t^i)$ *and* $(1-\tau)\kappa = \lambda_Z$ *we have that the dynamics of Eq. 8.11 hold: Assumption 8.1 implies Eq. 8.11.*

Theorem 8.4 shows that as the length of the data increases with time, nothing new is learned. The posterior fluctuates forever, but its dynamic *law of motion* converges. That is, in the limit Eq. 8.11 holds and any new data alter the conditional probability of agents but do not change the law of motion of g_t^i. If $\tau = 0$, the agent ignores all qualitative information and the posterior converges. But the results hold no matter how small τ is, since even the slightest perturbation of the Bayesian updating process cause it to fluctuate forever. This result is compatible with the work of Freedman (1963, 1965), who first demonstrated the general nonconvergence of Bayesian posteriors in an IID context but when the parameter space is countable. The work of Acemoglu et al. (2007) also relates to the issue of diversity in a setting in which the data do not permit an identification of the state.

Individual Perceptions, Market Belief, and Endogenous Amplification

We assume that the market is large and anonymous and that the distribution of beliefs is observable; hence its moments are known. Let $Z_t = \int g_t^i di$ be the first moment and refer to it as *average market belief*. Due to correlation across agents, averaging over the agents does not result in a constant, and the average $\epsilon_t = \int \rho_t^{ig} di$ is a random variable, not a constant 0. Hence we have

$$Z_{t+1} = \lambda_Z Z_t + \epsilon_{t+1} \tag{8.15}$$

Correlation of ρ_t^{ig} *across agents* may exhibit nonstationarity, and that would be inherited by the ϵ process. The empirical distribution of the ϵ process is denoted by a process ρ^Z. If the ϵ process is stationary, $\epsilon_t = \rho_t^Z$. Since the Z_t are observable, market participants have data on the *joint* process (d, Z); hence they know their *joint empirical distribution*. We assume that, this distribution is described by the system of equations

$$\begin{array}{ll} \textbf{(8.16a)} & d_{t+1} = \lambda_d d_t + \rho_{t+1}^d \\ \textbf{(8.16b)} & Z_{t+1} = \lambda_Z Z_t + \rho_{t+1}^Z \end{array} \quad \begin{pmatrix} \rho_{t+1}^d \\ \rho_{t+1}^Z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_d^2 & 0 \\ 0 & \sigma_Z^2 \end{bmatrix} = \Sigma\right) \quad \text{IID}$$

Eqs. 8.16a–8.16b are the first formal structure to explain the mechanisms of endogenous amplification of volatility. We started with one exogenous shock, and we find that correlation of beliefs expanded the economy's state space to include an aggregate market belief variable. If this variable affects prices, it causes endogenous amplification of market dynamics and volatility. Note that Z_t does not arise from individual choice; rather, it is a market externality arising from the correlation of individual beliefs. Indeed, in the theory reviewed here, the emergence of the distribution of market belief, as an observable variables that has economic impact, is the single most important development. But, to demonstrate that amplification is actually present, we need to show that equilibrium prices depend on market beliefs. This suggests a natural definition that is useful in assessing equilibria:

Definition 8.6. An economy exhibits *endogenous uncertainty* if an equilibrium price map is a function of the market belief.[9]

We now explain agent i's perception model. In Eq. 8.12, ${g_t}^i$ pins down agent i's forecast of d^i_{t+1}. We now broaden this idea to a perception model of the two state variables (d^i_{t+1}, Z^i_{t+1}) given d_t and Z_t. Following Theorem 8.3, her belief takes the joint form:

$$\begin{array}{ll}\textbf{(8.17a)} & d^i_{t+1} = \lambda_d d_t + \lambda^g_d g^i_t + \rho^{id}_{t+1} \\ \textbf{(8.17b)} & Z^i_{t+1} = \lambda_Z Z_t + \lambda^g_Z g^i_t + \rho^{iZ}_{t+1} \\ \textbf{(8.17c)} & g^i_{t+1} = \lambda_Z g^i_t + \rho^{ig}_{t+1}\end{array} \qquad \begin{pmatrix} \rho^{id}_{t+1} \\ \rho^{iZ}_{t+1} \\ \rho^{ig}_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \hat{\sigma}^2_d & \hat{\sigma}^2_{Zd} & 0 \\ \hat{\sigma}^2_{Zd} & \hat{\sigma}^2_Z & 0 \\ 0 & 0 & \sigma^2_g \end{bmatrix} = \Sigma^i \right)$$

${g_t}^i$ defines belief (d_{t+1}, Z_{t+1}), and Eqs. 8.17a and 8.17b show it pins down i's perceived transition of (d^i_{t+1}, Z^i_{t+1}). This simplicity ensures that one state variable pins down agent i's subjective belief; hence

$$ {E_t}^i \begin{pmatrix} d_{t+1} \\ Z_{t+1} \end{pmatrix} - {E_t}^m \begin{pmatrix} d_{t+1} \\ Z_{t+1} \end{pmatrix} = \begin{pmatrix} \lambda^g_d {g_t}^i \\ \lambda^g_Z {g_t}^i \end{pmatrix} $$

We stress two facts. First, market belief is shaped by correlation across individuals, but such correlation is a market externality with implications to efficiency considerations. Second, from the perspective of agents, Z_t is an economy-wide state variable like any other. But market belief is often wrong: It has forecast more recessions than actually occurred. In contrast with asymmetric information models, agents do not use Z_t to update beliefs about future exogenous variables: Eq. 8.17a does not depend on Z_t. Agents do not view Z_t as information about d_{t+1}, since it is not a "signal" about unobserved private information. They do consider Z_t as crucial "news" about what the *market thinks* about d_{t+1}. Since $t + 1$ prices depend on $t + 1$ market belief, to forecast future endogenous variables an agent must forecast Z_{t+1}, which express future beliefs of other agents.

Rationality Conditions for the Gaussian Model

Theorem 8.3 gives general rationality conditions, and we now explore the specific conditions that must be satisfied by the perception models (Eqs. 8.17a–8.17c). We note first that some rationality conditions have already been imposed. First, we argued that rational agents exhibit fluctuating beliefs, since a *constant* belief that is not the empirical

[9]Earlier we stressed the notion of *endogenous uncertainty* as entailing excess price volatility due to the effect of beliefs. The precise definition as given here was introduced in Kurz and Wu (1996) in the context of a General Equilibrium model. Kurz and Wu (1996) define the term as a property of the price map that has multiple prices for the same exogenous state.

probability is irrational. Second, $g_t{}^i$ are required to have an unconditional zero mean since beliefs are all about deviations from empirical frequencies. Third, any belief is a conditional probability of a stationary joint system. We now turn to Eqs. 8.17a–8.17c.

For Eqs. 8.17a–8.17c to be a *rational belief* it needs to induce the same empirical distribution of the observables (d_t, Z_t) as Eqs. 8.16a and 8.16b. In accord with Theorem 8.3, we then treat $g_t{}^i$ symmetrically with other random variables and require that for Eqs. 8.17a–8.17c to be a rational belief, we must have:

$$\text{Empirical distribution of the process} \begin{Bmatrix} \lambda_d^g g_t{}^i + \rho_{t+1}^{id} \\ \lambda_Z^g g_t{}^i + \rho_{t+1}^{iZ} \end{Bmatrix} = \qquad \textbf{(8.18)}$$

$$\text{the distribution of} \begin{pmatrix} \rho_{t+1}^d \\ \rho_{t+1}^Z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_d^2, & 0, \\ 0, & \sigma_Z^2 \end{bmatrix} \right), \text{IID}$$

To compute the implied statistics of the model, we first compute the moments of $g_t{}^i$. From Eq. 8.17c, the unconditional variance of $g_t{}^i$ is $\mathbb{V}\text{ar}(g^i) = \sigma_g^2/(1 - \lambda_Z^2)$. Hence, we have two sets of rationality conditions that follow from Eq. 8.18. The first arises from equating the covariance matrix

$$\text{(i)} \quad \frac{(\lambda_d^g)^2 \sigma_g^2}{1 - \lambda_Z^2} + \hat{\sigma}_d^2 = \sigma_d^2 \qquad \text{(ii)} \quad \frac{(\lambda_Z^g)^2 \sigma_g^2}{1 - \lambda_Z^2} + \hat{\sigma}_Z^2 = \sigma_Z^2 \qquad \text{(iii)} \quad \frac{\lambda_d^g \lambda_Z^g \sigma_g^2}{1 - \lambda_Z^2} + \hat{\sigma}_{Zd} = 0$$

The second set arises from equating the serial correlations of the two systems

$$\text{(iv)} \quad \frac{(\lambda_d^g)^2 \lambda_Z \sigma_g^2}{1 - \lambda_Z^2} + \mathbb{C}\text{ov}(\hat{\rho}_t^{id}, \hat{\rho}_{t+1}^{id}) = 0 \qquad \text{(v)} \quad \frac{(\lambda_Z^g)^2 \lambda_Z \sigma_g^2}{1 - \lambda_Z^2} + \mathbb{C}\text{ov}(\hat{\rho}_t^{iZ}, \hat{\rho}_{t+1}^{iZ}) = 0$$

(i) to (iii) fix the covariance matrix in Eqs. 8.17a–8.17c, and (vi)–(v) fix the serial correlation of $(\hat{\rho}_t^{id}, \hat{\sigma}_t^{iZ})$. An inspection of Eqs. 8.17a–8.17c reveals the choice left for an agent are the two parameters $(\lambda_d^g, \lambda_Z^g)$. But under the rational belief theory, these are not free either, since there are natural conditions they must satisfy. First, $\hat{\sigma}_d^2 > 0, \hat{\sigma}_Z^2 > 0$ place two strict conditions on $(\lambda_d^g, \lambda_Z^g)$:

$$|\lambda_d^g| < \frac{\sigma_d}{\sigma_g} \sqrt{1 - \lambda_Z^2} \quad |\lambda_Z^g| < \frac{\sigma_Z}{\sigma_g} \sqrt{1 - \lambda_Z^2}$$

Finally, we need to ensure the covariance matrix in Eqs. 8.17a–8.17c is positive definite. The following is a sufficient condition

$$\frac{1 - \lambda_Z^2}{\sigma_g^2} > \frac{(\lambda_Z^g)^2}{\sigma_Z^2} + \frac{(\lambda_d^g)^2}{\sigma_d^2}$$

The "free" parameters $(\lambda_d^g, \lambda_Z^g)$ are thus restricted to a narrow range, which is empirically testable.[10]

Comments on the Finite State Space Case

Much of the simulation work reported in Section 8.4 uses finite state space economies.[11] For example, Kurz (1997c), Kurz and Beltratti (1997), Kurz and Schneider (1996), Kurz and Motolese (2001, 2007), Motolese (2003), Nielsen (1996, 2003), Nakata (2007), and Wu and Guo (2003) all use OLG models with two "dynasties" of finite lived agents in which each agent has, at each date, two belief states. They assume that the sequence of parameters ${g_t}^i$ are IID with $Q^i\{{g_t}^i = 1\} = \alpha_i$, but ${g_t}^1$ and ${g_t}^2$ *are correlated*. This marginal distribution is fixed in the following discussion. The empirical distribution of dividends is typically assumed as Markov on two values (d^H, d^L) with a transition matrix as in Eq. 8.10b. We use this matrix together with Example 8.2 to review the main ideas.

Starting with the *endogenous amplification effect*, note that although the exogenous Markov dividend growth rate process takes two values, with two dynasties of agents, each with two belief states, the economy's state space is of Dimension 8. That is, the economy has four market belief states defined by possible values of the pair (g^1, g^2) and eight economy-wide states defined by the identification:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} \Leftrightarrow \begin{bmatrix} d = d^H, g^1 = 1, g^2 = 1 \\ d = d^H, g^1 = 1, g^2 = 2 \\ d = d^H, g^1 = 2, g^2 = 1 \\ d = d^H, g^1 = 2, g^2 = 2 \\ d = d^L, g^1 = 1, g^2 = 1 \\ d = d^L, g^1 = 1, g^2 = 2 \\ d = d^L, g^1 = 2, g^2 = 1 \\ d = d^L, g^1 = 2, g^2 = 2 \end{bmatrix}. \qquad \textbf{(8.19)}$$

[10] It may appear that the empirical evidence consists of more than the moments of the data series as stipulated in Section 8.3.1 and Definition 8.1; that is, one should look not only at the full data series but also at subsequences. Kurz (1994) argues that economic time series have deterministic patterns in seasonal and cyclical frequencies; hence if these are cleaned out so that we look at seasonally and cyclically and adjusted data, then under ergodicity, with probability 1 the empirical distribution along any subsequence over dates whose selection does not depend on the observed data is the same as the distribution along the entire sequence of data. Also, with finite data there are always an infinite number of unobservable sequences. Hence, there are no new restrictions that can be deduced from looking at subsequences. See Dawid (1985) for the Calibration literature view on the question of rationality conditions along subsequences.

[11] States of belief are described either with finite or continuous state models. Continuous state models tend to be more complex than discrete state models, which are more tractable, but the simulation results of the two models are essentially the same. To avoid repetition we report later detailed results deduced only for the continuous state models. Since a reader may find either one of these two more suitable for his or her application, we describe in the text the basic structure of *both* models. Hence, on first reading, *one may skip the sections on finite state modeling* and study these only after covering the full development of the continuous state models together with the numerical results of the simulations described later in Section 8.4.

The endogenous amplification in Eq. 8.19 induces an expansion of the state space and explains how beliefs increase price volatility above the “fundamental” volatility of dividends.

Under the assumptions of marginal Markov and IID distributions, the empirical distribution of the eight states is characterized by an 8×8 stationary transition matrix M. Hence, this matrix has two important properties: The dividend transition matrix (Eq. 8.10b) must be one marginal probability and the unconditional belief probabilities $Q^i\{g_t{}^i = 1\} = \alpha_i$ must be the second marginal probabilities. Denoting the multiplication $\alpha[M_{v,u}] = [\alpha M_{v,u}]$, it turns out that M must take the form

$$M = \begin{bmatrix} \varphi M(a) & (1-\varphi)M(a) \\ (1-\varphi)M(b) & \varphi M(b) \end{bmatrix} \quad \textbf{(8.20)}$$

where $M(a)$ and $M(b)$ are 4×4 matrices that take the form

$$M(a) = \begin{bmatrix} a_1 & \alpha_1 - a_1 & \alpha_2 - a_1 & 1 + a_1 - \alpha_1 - \alpha_2 \\ a_2 & \alpha_1 - a_2 & \alpha_2 - a_2 & 1 + a_2 - \alpha_1 - \alpha_2 \\ a_3 & \alpha_1 - a_3 & \alpha_2 - a_3 & 1 + a_3 - \alpha_1 - \alpha_2 \\ a_4 & \alpha_1 - a_4 & \alpha_2 - a_4 & 1 + a_4 - \alpha_1 - \alpha_2 \end{bmatrix}$$
$$M(b) = \begin{bmatrix} b_1 & \alpha_1 - b_1 & \alpha_2 - b_1 & 1 + b_1 - \alpha_1 - \alpha_2 \\ b_2 & \alpha_1 - b_2 & \alpha_2 - b_2 & 1 + b_2 - \alpha_1 - \alpha_2 \\ b_3 & \alpha_1 - b_3 & \alpha_2 - b_3 & 1 + b_3 - \alpha_1 - \alpha_2 \\ b_4 & \alpha_1 - b_4 & \alpha_2 - b_4 & 1 + b_4 - \alpha_1 - \alpha_2 \end{bmatrix} \quad \textbf{(8.21)}$$

φ is due to dividends and (α_1, α_2) is due to individual beliefs. This leaves open a and b, which reflect correlation *among* beliefs and *between* dividend growth and beliefs. In an uncorrelated world $a_i = b_i = 0.25$ for all i. If beliefs and dividend growth are uncorrelated $a = b$. Correlation does not arise by individual choices, *hence* (a, b) *reflect the externality of beliefs determined by social interaction and communication*. It is the crucial component of endogenous amplification of beliefs.

Now use Theorem 8.3 and Example 8.2 to define a Markov belief as a conditional probability $Q^i(s|g_t{}^i, H_t), s = 1, 2, \ldots, 8$, on the eight states in Eq. 8.19; implied by a 16×16 matrix F which is a joint probability on the eight economy-wide states and the two individual belief states. By Theorem 8.3 (see also Example 8.2) agent i is using two transition matrices $(F_1{}^i, F_2{}^i)$, with a conditioning as follows:

$$F_v{}^i = Q^i(\bullet|g_t{}^i = v, H_t) \quad \text{if} \quad g_t{}^i = v \quad \text{for} \quad i = 1, 2,\ v = 1, 2$$

and α_i is then the unconditional frequencies at which agent i uses matrix $F_1{}^i$. *Since* M *is deduced from the data, rationality of belief* requires the two pairs of matrices to satisfy the conditions

$$M = \alpha_i F_1{}^i + (1 - \alpha_i)F_2{}^i \quad \text{for } i = 1, 2 \quad \textbf{(8.22)}$$

Keeping in mind that the empirical distribution is given to an agent, rationality implies that relative to M agent i can select only ${F_1}^i$, since Eq. 8.22 implies that ${F_2}^i = \frac{1}{1-\alpha_i}(M - \alpha_i {F_1}^i)$. This imposes additional restrictions on ${F_1}^i$ due to the nonnegativity inequality conditions $M \geq \alpha_i {F_1}^i$. In sum, we have:

Theorem 8.5. *In the case of two types, two states of belief and two state Markov process of the dividend process the set of all rational beliefs relative to M is characterized by the set of* $(0 \leq \alpha_1 \leq 1, 0 \leq \alpha_2 \leq 1)$ *and pairs of nonnegative transition matrices* $({F_1}^i, {F_2}^i)$*, which satisfy Eq. 8.22.*

But what are the matrices of the two agents? Since any two matrices $({F_1}^i, {F_2}^i)$ of agent i are perturbations of M, there is a limited choice of matrices satisfying the inequalities $M \geq \alpha_i {F_1}^i$. In the simulation models of asset volatility cited previously, researchers chose a simple formulation that permits an agent to be either optimistic or pessimistic about the probability of high dividend states tomorrow. To do that, note that the first four rows and columns of M correspond to the high dividend state and the second four rows and columns to the low dividend state. Hence, the 4×4 matrix $\varphi \chi^i M(a)$ would express optimism or pessimism by a factor χ_i, relative to $\varphi M(a)$, in transition probabilities from a high dividend state today to a high dividend state tomorrow. For ${F_1}^i$ to be a transition matrix one must also adjust the matrix $(1 - \varphi)M(a)$.

For a specified parameter $0 \leq \alpha_i \leq 1$ and subject to the nonnegativity restrictions specified earlier, the following matrix is a rational belief that expresses optimism (if $\chi^i > 1$) or pessimism (if $\chi^i < 1$) in transition to high dividend state from all states today:

$$F_1^i = \begin{bmatrix} \varphi \chi^i M(a) & (1 - \varphi \chi^i) M(a) \\ (1 - \varphi)\chi^i M(b) & (1 - (1 - \varphi)\chi^i) M(b) \end{bmatrix} \tag{8.23}$$

In short, for a given $(0 \leq \alpha_1 \leq 1, 0 \leq \alpha_2 \leq 1)$ the set of rational beliefs of this form is summed up by one parameter χ^i, which varies over an interval defined by the inequality $M \geq \alpha_i {F_1}^i$. Now, if $\chi^i > 1$, an agent using ${F_1}^i$ is optimistic about d_{t+1}: his conditional probability of $d_{t+1} = d^H$ is higher than the stationary probability implied by M. When this agent uses ${F_2}^i$, his conditional probabilities of $d_{t+1} = d^H$ are lower than the probability implied by M. The results of these studies are reported later.

8.3.5. Asset Pricing with Heterogeneous Beliefs: An Illustrative Model and Implications

Having outlined the structure of beliefs, we now return to the infinite horizon model of Section 8.2.2 and adapt it to an economy with diverse beliefs but common information. Assume a continuum of agents on $[0, 1]$ and an exogenous risky payoff process $\{d_t, t = 1, 2, \ldots\}$ with an unknown stable and ergodic probability $\prod$ and empirical distribution described by a transition $d_{t+1} = \lambda_d d_t + \rho_{t+1}^d$ and where $\rho_{t+1}^d \sim N(0, \sigma_d^2)$, IID. The asset structure of the economy consists of an aggregate stock index (think of it as the S&P 500) and a risk-free bond. We assume the *riskless rate* r *is constant over time* and positive; hence $R = 1 + r > 1$ and $0 < \frac{1}{R} < 1$. Agent i borrows the amount $B_t^{\,i}$ at t and receives with certainty $B_t^{\,i}R$ at $t+1$. At date t, agent i buys θ_t^i shares of stock and receives dividend $D_t = d_t + \mu$ for each θ_{t-1}^i held. Consumption is then standard: $c_t^{\,i} = \theta_{t-1}^i[p_t + d_t + \mu] + B_{t-1}^i R - \theta_t^i p_t - B_t^{\,i}$. Equivalently, we define wealth $W_t^{\,i} = c_t^{\,i} + \theta_t^i p_t + B_t^{\,i}$ and derive the familiar transition for wealth

$$W_{t+1}^i = (W_t^{\,i} - c_t^{\,i})R + \theta_t^i \pi_{t+1} \quad \pi_{t+1} = p_{t+1} + (d_{t+1} + \mu) - Rp_t \qquad \textbf{(8.24)}$$

π_t is excess returns per share. For initial values (θ_0^i, W_0^i) the agent maximizes the expected utility

$$\underset{(\theta^i, c^i)}{E_t^i}\left[\sum_{s=0}^{\infty} -\beta^{t+s} e^{-(\gamma c_{t+s}^i)} | H_t\right] \qquad \textbf{(8.25)}$$

subject to transitions Eqs. 8.17a–8.17c of the state variables $\psi_t^i = (1, d_t, Z_t, g_t^{\,i})$. H_t is date t information.

Our assumptions are restrictive. Constant R is not realistic and exponential utility exhibits no income effects. Nevertheless, these assumptions have the great advantage of leading to closed form solutions that are helpful vehicles to explain the main ideas. Hence the term *illustrative* in this section's title. To seek a closed-form solution we conjecture that prices are linear in the economy's state variables; hence equilibrium price p_t is conditionally normally distributed. In Theorem 8.4 we confirm this conjecture. For an optimum (for details, see the Appendix of Kurz and Motolese, 2007) there exists a constant vector u, so the demand functions for the stock is

$$\theta_t^i(p_t) = \frac{R}{\gamma r \hat{\sigma}_\pi^2}[E_t^{\,i}(\pi_{t+1}) + u\psi_t^i],\ u = (u_0, u_1, u_2, u_3),\ \psi_t^i = (1, d_t, Z_t, g_t^i) \qquad \textbf{(8.26)}$$

$\hat{\sigma}_\pi^2$ is an *adjusted* conditional variance (the "adjustment" is explained in Section 8.4.3) of excess stock return π_{t+1}, which is assumed to be constant and the same for all agents. The term $u\psi_t^i$ is the intertemporal hedging demand that is linear in agent i's state variables.

For an equilibrium to exist, we impose stability conditions on the dynamics of the economy:

Stability Conditions We require that (i) $0 < \lambda_d < 1$, (ii) $0 < \lambda_Z + \lambda_Z^g < 1$.

(i) requires that $\{d_t, t = 1, 2, \ldots\}$ is dynamically stable, and (ii) requires dynamic stability of *belief*. It requires the market, on average, to believe that (d_t, Z_t) is stable. To see why, look at this definition.

Definition 8.7. The average market belief operator is $\overline{E}_t(\bullet) = \int E_t^{\,i}(\bullet)di$.

Now take expectations of Eq. 8.17b, average over the population and recall that Z_t are averages of $g_t{}^i$. This implies that

$$\overline{E}_t[Z_{t+1}] = (\lambda_Z + \lambda_Z^g)Z_t$$

Kurz (2007) and Kurz and Motolese (2007) then demonstrate the following results.

Theorem 8.6. *Consider the model with heterogeneous beliefs under the stability conditions specified with supply of shares equal to 1. Then there is a unique equilibrium price function, which takes the form*

$$p_t = a_d d_t + a_z Z_t + P_0 \tag{8.27a}$$

with coefficients

$$a_d = \frac{\lambda_d + u_1}{R - \lambda_d} \tag{8.27b}$$

$$a_z = \frac{(a_d + 1)\lambda_d^g + (u_2 + u_3)}{R - (\lambda_Z + \lambda_Z^g)} \tag{8.27c}$$

$$P_0 = \frac{(\mu + u_0)}{r} - \frac{\gamma\hat{\sigma}_\pi^2}{R} \tag{8.27d}$$

The linearity of the price thus confirms the earlier conjecture that the price is conditionally normal.

Now, closed-form solutions for the hedging demand parameters $u = (u_0, u_1, u_2, u_3)$ are not available; hence Kurz and Motolese (2007) compute numerical Monte Carlo solutions. For all values of the model parameters they find (i) $a_d > 0$, (ii) $(a_d + 1)\lambda_d^g + (u_2 + u_3) > 0$, and (iii) $a_z > 0$. These conclusions are reasonable: *Today's asset price increases if d_t or Z_t rise*.

Our main objective now is to assess the implications of theory to *the effects of diverse rational beliefs on asset market dynamics*. We do it in two ways. First, we use the closed-form solution of the illustrative model as a simple reference. Second, we use the developments up to now, together with a citation of other papers, to develop results on the questions at hand. We devote the rest of this section to a discussion of such implications of the theory.

Endogenous Uncertainty

The most direct implication of the theory is that asset markets are subject to *endogenous uncertainty*. To explore Definition 8.6, we examine the price map $p_t = a_d d_t + a_z Z_t + P_0$ and find that endogenous uncertainty is expressed in two ways. First the term Z_t says that the risks of asset returns are, in part, due to the risk of future market belief. By Eqs. 8.3.4a and 8.3.4b, in the long run $\sigma_p^2 = a_d^2\sigma_d^2 + a_z^2\sigma_Z^2$; hence price volatility is caused by exogenous as well as endogenous forces, and this has far-reaching implications to market efficiency, risk premia, and public policy. Second, Kurz (2007) shows that the variance in P_0 can be approximated by $\hat{\sigma}_{\pi}^2 \simeq (a_d + 1)^2 \hat{\sigma}_d^2 + 2(a_d + 1) a_z \hat{\sigma}_{Zd} + a_Z^2 \hat{\sigma}_Z^2$. These are terms from the covariance matrix in the agent belief Eqs. 8.17a–8.17c; hence they depend on perception rather than on the actual empirical moments in Eqs. 8.16a–8.16b. But as the perceived volatility of dividend and average market belief increases, the price declines. We show in the next section that this fact implies an increased risk premium.

The presence of endogenous uncertainty in asset markets has far-reaching implications to asset-pricing theory and market dynamics and we note a few of these general conclusions:

- An asset's price is not equal to a unique fundamental value determined by the flow of future payoffs. Moreover, market belief about exogenous states matters since it is often wrong; hence market belief is an independent and dominant component of asset price volatility.
- Moral hazard and the large dimension of market belief make it impossible for markets to trade contracts contingent on market belief; hence markets are fundamentally incomplete.
- In scales of days or weeks, changes in productivity, growth, and profits are slow. Hence, it is absolutely clear without much formal analysis that most volume of trading results from changes in the market distribution of beliefs. Indeed, over the short run the key function of asset markets is to permit agents to *trade their belief differences*.
- Expected individual excess returns and "efficient frontiers" are both subjective concepts. Hence, in markets with diverse beliefs most predictions of CAPM theory do not hold.
- By anonymity of individuals, the market belief is a public externality and hence subject to the effect of coordination and public policy. Stabilization policy can thus have a strong effect on market volatility, and this carries over to monetary economies as well.

The Endogenous Uncertainty Risk Premium

We now turn to an exploration of the risk premium under heterogeneous beliefs in the illustrative model of Section 8.3.5, and we review results in Kurz and Motolese (2007). Recall that the premium on a long position, *as a random variable*, is defined by

$$\frac{\pi_{t+1}}{p_t} = \frac{p_{t+1} + d_{t+1} + \mu - Rp_t}{p_t}. \qquad \textbf{(8.28)}$$

We seek a measure of the premium as a *known expected quantity* recognized by market participants, but we have a problem, since with diverse beliefs the premium is subjective. From Eqs. 8.27a–8.27d we compute three equilibrium measures to consider. One is the subjective expected excess returns by i,

$$E_t^i(\pi_{t+1}) = (a_d + 1)(\lambda_d d_t + \lambda_d^g g_t^i) + a_z(\lambda_Z Z_t + \lambda_Z^g g_t^i) + \mu + P_0 - Rp_t \qquad \textbf{(8.28a)}$$

Aggregating over i we define the *market premium* as the average market expected excess returns. It reflects what the market expects, not necessarily what the market gets:

$$\overline{E}_t(\pi_{t+1}) = (a_d + 1)(\lambda_d d_t + \lambda_d^g Z_t) + a_z(\lambda_Z Z_t + \lambda_Z^g Z_t) + \mu + P_0 - Rp_t \qquad \textbf{(8.28b)}$$

Eqs. 8.28a and 8.28b are not necessarily "correct," and we focus on a third, objective, measure, which is common to all. Econometricians who study the long-term time variability of the premium measure it by the empirical distribution of Eq. 8.28, which, by Eqs. 8.27a–8.27d and Eqs. 8.16a and 8.16b, is

$$E_t^m[\pi_{t+1}] == (a_d + 1)(\lambda_d d_t) + a_z \lambda_Z Z_t + \mu + P_0 - Rp_t \qquad \textbf{(8.28c)}$$

Eq. 8.28c is the common way all researchers cited previously have measured the risk premium; therefore we refer to it as *the* risk premium.

We thus arrive at two conclusions. First, the difference between the individual perceived premium and the market perceived premium is

$$E_t^i[\pi_{t+1}] - \overline{E}_t[\pi_{t+1}] = [(a_d + 1)\lambda_d^g + a_z \lambda_Z^g](g_t^i - Z_t) \qquad \textbf{(8.29a)}$$

From the perspective of trading, all that matters is the difference $g_t^i - Z_t$ between individual and market belief. In addition, the following difference is important:

$$E_t^m[\pi_{t+1}] - \overline{E}_t[\pi_{t+1}] = -[(a_d + 1)\lambda_d^g + a_z \lambda_Z^g]Z_t \qquad \textbf{(8.29b)}$$

The risk premium is different from the market perceived premium when $Z_t \neq 0$. But the important conclusion is the analytical expression of the risk premium:

$$E_t^m[\pi_{t+1}] = \left(\frac{\gamma r \hat{\sigma}_\pi^2}{R} - u_0 - u_1 d_t\right) - a_z(R - \lambda_Z)Z_t \qquad \textbf{(8.30)}$$

Since $a_z > 0$, $R > 1$, and $\lambda_Z < 1$, it follows that the premium per share *declines* with Z_t. We then have Theorem 8.7.

Theorem 8.7. *The risk premium $E_t^m[\pi_{t+1}]$ is increasing in the variance $\hat{\sigma}_\pi^2$ and decreasing in the mean market belief Z_t.*

This theorem exhibits what Kurz and Motolese (2007) call the "Market Belief Risk Premium." It shows that the risk premium depends on market belief in two ways:

1. A direct effect on the permanent mean premium $\frac{\gamma r \hat{\sigma}_\pi^2}{R}$. We have seen that the variance is approximately $\hat{\sigma}_\pi^2 \simeq (a_d + 1)^2 \hat{\sigma}_d^2 + 2(a_d + 1)a_z \hat{\sigma}_{Zd} + a_Z^2 \hat{\sigma}_Z^2$; hence it increases with the perceived volatility of dividend and the volatility of average market belief.

2. An effect on the time variability of the risk premium, expressed by $-a_z(R - \lambda_Z)Z_t$ with a negative sign when $Z_t > 0$.

To understand the second result, note that it says if we run a regression of excess returns on the observable variables, the effect of the market belief on excess return *is negative*. From an REE perspective this sign is somewhat surprising, since when $Z_t > 0$ the market expects *above-normal* future dividends but instead, the risk premium on the stock *declines*. When the market holds bearish belief about dividends ($Z_t < 0$), the risk premium *rises*. This requires some further explanation.

Why is the effect of Z_t on the risk premium *negative*? The result shows that when the market holds abnormally favorable belief about future payoffs of an asset, the market views the long position as less risky and the risk premium on the long position of that asset falls. Fluctuating market belief implies time variability of risk premia, but fluctuations in risk premia are inversely related to the degree of market optimism about future prospects of asset payoffs.

To further explore the result, it is important to explain *what it does not say*. One might interpret it as confirming a common claim that to maximize excess returns it is an optimal strategy to be a "contrarian" to the market consensus by betting against it. To understand why this is a false interpretation, note that when an agent holds a belief about future payments, the market belief does not offer any new information to alter the individual's belief about the exogenous variable. If the agent believes that future dividends will be abnormally high but $Z_t < 0$, the agent does not change her forecast of future dividends. She uses the market belief information only to forecast future *prices* of an asset. Thus, Z_t is a useful input to forecasting returns without changing the forecast of d_{t+1}. Given the available information, an optimizing agent is already placed on her demand function defined relative to her own belief; hence it is not optimal for her to just abandon her demand and adopt a contrarian strategy.

This argument is the same as the one showing why it is not optimal to adopt the log utility as your own utility, even though it maximizes the growth rate of your wealth. Yes, it does, but you dislike the sharp expected declines in the values of your assets. By analogy, following a "contrarian" policy may imply a high long-run average return in accord with the empirical probability m. However, if you disagree with this probability, you will dislike being short when your true optimal position is to be long. Indeed, this argument explains why *most people hold positions that are in agreement with the market belief most of the time* instead of betting against it. The crucial observation to make is that a maximizing agent has his own belief about future events, and he does not select a new belief when he learns the market belief. From his point of view, the market belief is an important state variable used to forecast future prices. When it is wrong, the market may forecast a recession that never arrives.

Theorem 8.7 was derived for an exponential utility function. Kurz and Motolese (2007) show that this result is more general and depends only on the positive coefficient a_z of Z_t in the price map. For more general utility functions, they use a linear approximation to show that the result depends only on the condition that the slope of the stock price is positive with respect to Z_t. This condition requires the current stock price to increase if the market is more optimistic about the asset's future payoffs.

Finally, Kurz and Motolese (2007) use data compiled by the Blue Chip survey of forecasts to test the theory proposed in Eq. 8.35b. They report that the data support the theoretical results.

Rational Overconfidence

Evidence from the psychological and behavioral literature (e.g., Svenson, 1981; Camerer and Lovallo, 1999; and Russo and Schoemaker, 1992) shows a majority of individuals assess their own probability of success in performing a task (investment, economic decisions, driving, etc.) above the empirical frequency of success in a population. Hence a majority of people often expect to outperform the empirical frequency measured by the median or mean. In a *rational expectations* paradigm, individuals know the true probability of success; hence the observed inconsistency is taken to be a demonstration of irrational behavior. Indeed, inconsistency between individual assessments and empirical frequencies has been cited extensively as a "proof" of irrational behavior and in support of behavioral/psychological impulses for belief and forecasting. This phenomenon has thus been called *overconfidence*. We reject this conclusion and show that it reveals a fundamental flaw.

The work cited previously (and other empirical and experimental work) provides evidence against *rational expectations*. But rational expectations is an extreme theory in demanding agents to know the full structure of the economy and make exact probability assessments. Behavioral economics takes the other extreme view and assumes that people are irrational and motivated by psychological impulses. Hence, a rejection of rational expectations does not imply acceptance of irrationality of agents. Indeed, we may reject these two extreme perspectives by observing the fact that most people do the best they can, given the limited knowledge they have. Rational people do not know everything and make "mistakes" relative to a true model they do not know. The theory of *rational beliefs* rejects both extremes in favor of an intermediate concept of rationality. We then show that *overconfidence* is compatible with *rational beliefs* and, indeed, *agents who hold rational belief will universally exhibit "rational overconfidence."* Hence, the cited empirical evidence is no proof that people are irrational and motivated by pure psychological factors.

We explain the preceding by using Example 8.1 (also, see Nielsen, 2006). A group of gamblers look at the black box in Example 8.1 and form beliefs using different sequences $g = (g_0, g_1, \ldots)$ as in that example. Each belief is then defined by a sequence of *independent* random variables satisfying Eq. 8.9a. Gamblers vary with the sequences g they use. A survey is taken and the distribution of beliefs is publically announced. Hence, belief distributions of past gamblers are known but not their individual beliefs. Since $(1/3)(0.60) + (2/3)(0.45) = 0.50$, all are rational beliefs for almost all g. In the *rational beliefs* literature the ratio 1/3 is referred to as the "frequency" of optimism. When the frequency of bull and bear states is not the same, we have a *market asymmetry* between them. The probability 0.60 is the "intensity" of optimism when optimistic. In defining a *rational belief* these characteristics are selected separately: For each frequency there is a range of feasible intensities that are rational.

The gamblers decide at date $t-1$ how they want to bet. They can gamble \$1 on $v_t = 1$ or on $v_t = 0$: They win \$1 if they are right and they lose \$1 if they are wrong. Since it's a small bet, they will all bet. Those who put money on 1 expect to win with a probability 0.60; those who put their money on 0 expect to win with probability 0.55. *They are all overconfident and rational!* The constancy of the high (0.60) and low (0.45) probabilities is not essential since we can, instead, put in any sequences of parameters that converge to 0.50 from above and from below and the result will be the same.

Observe that in this example all deviations from empirical frequencies lead to optimal behavior that exhibits *universal* overconfidence. When a subjective probability is above the empirical frequency, a long position is optimally taken with overconfidence. When a probability is below the empirical frequency, a short position is optimally held with overconfidence. *Hence, all agents are then optimally overconfident at all times*.

Generalizing the example is natural. Beliefs are all about deviations from empirical frequencies on the basis of which economic decisions are made. Optimistic agents engage in taking the risk of success in an activity, and pessimistic agents engage in gambles against success. If they cannot gamble against it (e.g., short positions are not allowed), they refrain from participation. This type of behavior is then natural to the *rational belief* paradigm. Moreover, this behavior is natural to any complex environment in which aggregation of subjective probability beliefs of agents may not be equal to the empirical frequencies. But then all creative work and all innovative decisions are based on beliefs that exhibit "overconfidence." Indeed, one can hardly think of entrepreneurship, inventive activity, and any speculative behavior without beliefs that exhibit rational overconfidence.

Properties of Average Market Belief and Higher-Order Beliefs

By Eqs. 8.17a and 8.17b and Definition 8.7 it follows that:

$$\overline{E}_t \begin{pmatrix} d_{t+1} \\ Z_{t+1} \end{pmatrix} - E_t^m \begin{pmatrix} d_{t+1} \\ Z_{t+1} \end{pmatrix} = \begin{pmatrix} \lambda_d^g Z_t \\ \lambda_Z^g Z_t \end{pmatrix} \quad \textbf{(8.31)}$$

and Eq. 8.31 exhibits the dynamics of the average market belief operator. However, Eqs. 8.17a–8.17c also show that properties of conditional probabilities do not apply to the market belief operator $\overline{E}_t(\bullet)$ since it is *not a proper conditional expectation*. To see why, let $X = D \times Z$ be a space where (d_t, Z_t) take values, and let G^i be the space of g_t^i. Since i conditions on g_t^i, his unconditional probability is a measure on the space $((D \times Z \times G^i)^\infty, \mathcal{F}^i)$, where $\mathcal{F}^i$ is a sigma field. The market conditional belief operator is just an average over conditional probabilities, each conditioned on a *different* state variable. Hence, this averaging does not permit one to write a probability space for the market belief. *The market belief is neither a probability nor rational!* This is then formulated (see Kurz, 2007) as Theorem 8.8.

Theorem 8.8. *The market belief operator violates iterated expectations:* $\overline{E}_t(d_{t+2}) \neq \overline{E}_t\overline{E}_{t+1}(d_{t+2})$.

Our earlier comment about the importance of the treatment of market belief as a state variable with independent dynamics is now complemented by Theorem 8.8. Market belief is an externality that does not arise from rational social choice of a collective agent. It cannot arise in a model of intertemporal choice of a single representative agent.

Turning to higher-order beliefs, we must distinguish between higher-order beliefs that are *temporal* and those that are *contemporaneous*. Eqs. 8.17a–8.17c define agent i's belief over future sequences of (d_t, Z_t, g_t^i) and as is the case for any probability, it implies i's temporal higher-order belief with regard to future events. For example, we deduce from Eqs. 8.17a–8.17c statements such as:

$$E_t^i(d_{t+N}) = E_t^i E_{t+1}^i \ldots E_{t+N-1}^i(d_{t+N}), \quad E_t^i(Z_{t+N}^i) = E_t^i E_{t+1}^i \ldots E_{t+N-1}^i(Z_{t+N}^i) \quad \textbf{(8.32)}$$

Properties of temporal higher-order beliefs are thus familiar properties of conditional expectations.

As to market belief, since Eq. 8.15 is implied by Eq. 8.17c, the average market belief operator satisfies $\overline{E}_t(d_{t+N+1}) = \lambda_d\overline{E}_t(d_{t+N}) + \lambda_d^g\overline{E}_t(Z_{t+N})$. We deduce perceived higher-order temporal market beliefs by averaging over i. For example,

$$\lambda_d^g\lambda_Z\overline{E}_t(Z_{t+N}) = \overline{E}_t\overline{E}_{t+N+1}(d_{t+N+2}) - \overline{E}_t E_{t+N+1}^m(d_{t+N+2}) \quad \textbf{(8.33)}$$

Contemporaneous higher-order beliefs have attracted attention (e.g., Allen, Morris, and Shin, 2006; Bacchetta and van Wincoop, 2005; and Woodford, 2003) despite being unobservable. They occur naturally in strategic situations. In a market context they can *formally* arise in Eqs. 8.17a–8.17c as follows: Let Z_t in Eq. 8.15 be defined as Z_t^1. We can argue that agents may form beliefs about the future of this variable by using a *second* belief index g_t^{i2} about Z_{t+1}^1 whose transition would be deduced from the transition of g_t^{i2}. Now $Z_t^2 = \int g_t^{i2}di$ would be a second-order aggregate belief for which a *third* belief index g_t^{i3} could be introduced whose average would be Z_t^3, and so on. Such infinite regress is problematic and leads us to reject contemporaneous higher-order beliefs in markets for two reasons. First, higher-order beliefs are degenerate in Eqs. 8.17a–8.17c *because* the single-belief index g_t^i fully pins down agent i's belief. Moreover, since agents know the beliefs of others and all variables in the price map (which embody the beliefs of others), there is nothing else about which to form beliefs. There is a second and more general reason why, in markets, *all higher-order beliefs* Z_t^j, *for* $j > 1$ *are degenerate*. This is so since they are averages of g_t^{ij} and since for $j > 1$ the Z_t^j are not observable; they can only exist in the minds of the agents, and hence there is no possible mechanism for individual g_t^{ij} to be correlated as in Eq. 8.15. Hence, higher-order beliefs cannot have an aggregate effect, since with independent g_t^{ij} the averages Z_t^j for $j > 1$ are zero at all t.

On Beauty Contests

The Keynes Beauty Contest metaphor has been extensively discussed. Some have associated it with asset-pricing equilibrium, where the price is expressed as iterated expectations of average market belief of the future *fundamental value* of the asset. In Eq. 8.8 we presented the Allen Morris and Shin (2006) example of such pricing with private information. But this interpretation should be questioned. An examination of Keynes' view (see Keynes, 1936, page 156) shows that the crux of Keynes's conception is that there is little merit in using fundamental values as a yardstick for market valuation. Hence what matters for the asset demand of an agent is the perception of what the market believes the future price of that asset will be rather than what the intrinsic value is. Keynes insists that future price depends on future market belief and that *may be right or wrong* without a necessary relation to an intrinsic fundamental value. The Beauty Contest parable is thus simple: A price does not depend on an intrinsic value but on what the market believes future payoffs and valuations will be. Keynes's Beauty Contest is thus a statement that to forecast the price in the future, an individual must forecast the future market state of belief, *when such forecasts may be "right" or "wrong."* We now observe that a *rational belief equilibrium* captures the essence of the Keynes Beauty Contest.

To explain, we make two observations. From Eq. 8.27a, equilibrium price is $p_t = a_d d_t + a_z Z_t + P_0$, and this is clearly in accord with the preceding: *In any model of the Beauty Contest, equilibrium price should not depend on a true intrinsic value; rather, it should depend on market belief.* It follows from the rationality conditions that price/earning ratios exhibit fluctuations with reversion to the long-run stationary mean, but such long-term value is not an intrinsic fundamental value. Indeed, in a model with diverse belief there is no such thing as fundamental intrinsic value, since all prices depend on market belief. Second, to forecast future prices, an agent forecasts Z_{t+1}, which is the market belief tomorrow. From Eq. 8.17b, we have $Z^i_{t+1} = \lambda_Z Z_t + \lambda^g_Z g^i_t + \rho^{iZ}_{t+1}$, which means that an agent forecasts the future market belief with his own subjective model. In sum, this equilibrium concept reflects the Beauty Contest parable because the price map depends on market beliefs, not on some agreed-on intrinsic value, and to forecast future prices agents must forecast the belief of others.

Speculation

Although market practitioners have an intuitive idea of what "speculation" is, there is no scientific consensus on how to define this concept. Keynes (1936) viewed asset markets as a "beauty contest," and many writers have interpreted this to be a form of speculation. A different perspective was proposed by Kaldor (1939), who defined speculation as "the purchase (or sale) of goods with a view to resale (repurchase) at a later date." It is clear that for such asset trades to make sense, prices of assets must regularly deviate from their fundamental values, and agents must believe that prices, are or will not be equal to their fundamental values. It is also clear that in a perfect REE world with homogenous beliefs and complete information, a Kaldor speculation is not possible (e.g., see Tirole, 1982; Milgrom and Stokey, 1982). Here we explore the perspective of a diverse belief equilibrium with respect to Kaldor (1939) speculation.

Following the definition of Kaldor (1939), Harrison and Kreps (1978) study the consequences of risk-neutral investors having different beliefs about the dividend process of a risky asset. At date t, investor i can expect a payment $E_t^i(\beta^k p_{t+k} + \sum_{s=0}^{k} \beta^s(d_{t+s} + \mu))$ if he chooses to resell k periods later, where $\{p_t\}$ and β denote the stock price process and the discount rate. The equilibrium market price, called a *consistent price scheme*, is the supremum over all stopping times k and across all investors. That is, this price is

$$p_t = \max_i \sup_k E_t^i \left(\beta^{t+k} p_{t+k} + \sum_{s=0}^{k-1} \beta^{t+s}(d_{t+s} + \mu) \right)$$

Agents hold diverse beliefs and are assumed to have infinite wealth for each class of investor type. A speculative premium is then defined to be the difference between the consistent price scheme and the value, $\max_i E_t^i(\sum_{s=0}^{\infty} \beta^{t+s}(d_{t+s} + \mu))$, expected when all investors are obliged to hold the asset forever. Harrison and Kreps (1978) show that under the assumptions made, there exists a positive speculative premium, or a price bubble, when short sales are not allowed.

Morris (1996) further examines asset pricing during initial public offerings when investors have different prior distributions, but the difference of belief disappears as investors learn from observations. A major weakness of both the works of Harrison and Kreps (1978) and Morris (1996) is their assumption of the unrestricted heterogeneity of beliefs.

Wu and Guo (2003) use the theory of rational belief to explain the persistence of diverse beliefs in the Harrison and Kreps (1978) model and to narrow the equilibrium results. They adopt the finite state Markov assumption for dividend and the two state of belief model. Wu and Guo (2003) then show that in contrast with the complex solution of Harrison and Kreps (1978) a rational belief equilibrium price vector (over states) is a simple expression that is computed via a finite algorithm. As to dynamics, they show that speculative bubbles and endogenous uncertainty emerge. They further characterize how the speculative premium increases with the degree of heterogeneity.

To explore the phenomenon of *simultaneous* increase in asset prices and trading volume, Wu and Guo (2004) study a model of heterogeneous rational beliefs held by a continuum of agents on the unit interval, as in Miller (1977). In contrast with Harrison and Kreps (1978) and Morris (1996), Wu and Guo (2004) permit limited short sales and impose a wealth constraint of a finite investment fund. They assume an IID dividends process over two states and arrange investors in the order of their optimism along the unit interval. They then derive a steady-state rational belief equilibrium price and show that in equilibrium optimistic investors hold the entire supply. In this framework, Wu and Guo (2004) demonstrate the emergence of endogenous uncertainty with a positive speculative premium that increases with the size of the investment fund and degree of optimism and decreases with the size of short-sale constraint. Furthermore, the model generates a positive relationship between trading volume and the directions of price changes: Volume is high when prices rise and it is low when prices decline. There is also a positive relationship between trading volume and price level. These results are consistent with the empirical evidence (e.g., Karpoff, 1987, and Basci et al., 1996).

8.4. EXPLAINING MARKET DYNAMICS WITH SIMULATION MODELS OF DIVERSE BELIEFS

Although much of our discussion is analytical, significant results about excess volatility are deduced from simulation models. Simulations require specification of functions, parameters, and beliefs and aim to show that a model replicates the statistics of the economy.

8.4.1. Introduction: On Simulation Methods and the Main Results

Since most models reviewed have only two agent types, the beliefs selected are representative of only two classes of agent. Since we might question the validity of such an approach as too narrow, it is useful to explain the *common features* of all simulation models reviewed that replicate the dynamics of real markets we observe. Our view on this issue is simple: a simulation model is a very good tool to explore the impact of the *qualitative features* of feasible belief structure on market volatility. The specific parameter configurations used to attain these qualitative features are less important.

The best way to explain this view is to highlight the central conclusions of the work we review in this section, and the summation of Kurz et al. (2005a) is useful. This paper starts from the view that in any non-REE-based asset market theory there are basically two *natural individual states*: optimistic (i.e., bull states) and pessimistic (i.e., bear states). The authors then explain that given these two basic states, there are three central characteristics of individual beliefs that fully account for all characteristics of market volatility and risk premia observed in real markets. These are:

1. Large (i.e., high-intensity) fat tails in the belief densities of agents
2. Asymmetry in the proportion of bull and bear states in the market over time
3. Belief states are correlated, resulting in regular joint dynamics of belief distributions

Large fat tails means that the densities of the agents' beliefs have very fat tails. "Intensity" measures size of deviations from stationary probabilities, as in the review of rational overconfidence. The *asymmetry in the time frequency of belief states* needed to reproduce the results is a subtle feature that says that on average, agents are in bear states at more than 50% of the dates. Equivalently, on average, at more than half the time, agents do not expect long positions to make *above-normal* returns on their investments. Therefore, it follows from the rational belief principle that when agents are in bull states and expect *above-normal* returns, their expected excess returns must be very high. We shall see later that this asymmetry is empirically supported by the fact that major abnormal rises in stock prices occur over a relatively small fraction of time. Hence, when agents believe a bull market is ahead, they expect to make excess return in relatively short periods.

Correlation of beliefs is a market externality, not determined by individual choices, which regulates the probability of agreement or disagreement of beliefs in the market and the transitions among such states. This is central because the distribution of beliefs determines prices and returns, and the dynamic of the belief distribution is crucially affected by the correlation. It is then natural that asymmetry in the transitions is important, since it regulates the dynamics of bear vs. bull markets.

Kurz et al. (2005a) then make two observations. First, exactly the same simulation model used to study market volatility is also used to study all other aspects of market dynamics. In that model, stock prices and returns exhibit a structure of forecastability observed in the real data. Also, the same model implies that market returns exhibit stochastic volatility generated by the dynamics of the market beliefs. Second, examination of alternative configurations of belief shows that no other configuration of qualitative features than the three previously specified yields predictions that *simultaneously replicate the empirical record*. Many feasible model parameters generate volatility of prices and returns, but as we move away from the three features, the model fails to generate some essential components of the empirical record, frequently the riskless rate and the risk premium. Thus, the main reason the models are able to explain the empirical record is that they have the needed configuration of qualitative factors. In each case they imply a unique parameter structure of the computational model needed to explain the empirical record, but the specific implied belief is not of central significance.

8.4.2. Anatomy of Market Volatility

Papers on excess volatility simulate computed equilibria with finite or infinite belief states. Those with finite belief states are OLG models, whereas those with infinite belief states are infinite horizon models. This division guides our review.

Understanding the Parametrized Structure of Beliefs

The papers that fall into the first category (that is, OLG) include Nielsen (1996, 2003, 2005, 2006), Kurz (1997b), Black (1997, 2005), Kurz and Beltratti (1997), Kurz and Motolese (2001), Kurz and Schneider (1996), Motolese (2003), Nakata (2007), and Wu and Guo (2003). Papers using infinite horizon models with infinite belief states include Kurz et al. (2005a, 2005b), Kurz (2007), Kurz and Motolese (2007), and Guo and Wu (2007).

We start by discussing OLG models *with finite belief states* and use the parametrization of Kurz and Motolese (2001) as a prototype.[12] All models have two assets: a stock and a riskless bond.

[12]Here again we present the parametrization of finite state and continuous state models. The empirical results reported later are all deduced from the continuous state model; hence it may be useful for the reader to skip, on first reading, material that pertains to finite state models. This material would be especially relevant if the reader wants to replicate any of these results by visiting the Web pages provided in footnotes 13 and 15 to download the programs with which to compute the solutions. It will be found that the finite state models are much easier to handle.

The stock pays dividends with a two-state growth rate. There are two types of agent, each living two periods with a power utility function of agent i over consumption:

$$u(c_t^1, c_{t+1}^2) = [1/(1-\gamma)](c_t^1)^{1-\gamma} + [\beta/(1-\gamma)](c_{t+1}^2)^{1-\gamma}, \gamma > 0, 0 < \beta < 1$$

The belief structure is as in Eqs. 8.20–8.23 with two states; thus there are eight economy-wide states.

Rational beliefs are represented by the two pairs of matrices (F_1^i, F_2^i) with frequencies (α_1, α_2). Since deviations from the long-term mean growth rate of dividends could be either *above* it or *below* it, at each date an agent must be either a bull or a bear about future growth of dividends. This is expressed by a pair of parameters (χ^1, χ^2) that measure *the intensity of optimism when in an optimistic state*, while (α_1, α_2) measure *the frequency of each of the two agents being optimistic*.

To see the implications, note that χ^i are revisions of the probabilities of states (1, 2, 3, 4) and (5, 6, 7, 8) relative to M. $\chi^i > 1$ imply increased probabilities of (1, 2, 3, 4) in matrices F_1^i when the first four prices occur at $d_t = d^H$ states. But these are actually states of high prices as well; hence $\chi^i > 1$ implies that agent i is optimistic about high prices at $t+1$. In all simulations, $\chi^i \geq 1$; hence one interprets g_t^i so that $g_t^i > 0$ means agent i is optimistic (*relative to* M) at t about *high prices* at $t+1$.

In the transition M, the matrices $M(a)$ and $M(b)$ regulate the correlation across beliefs and the effect of dividends on that correlation. This is a *correlation externality* given to agents, which is the same as the correlation among the ρ_t^{ig} across i in Eq. 8.11, a correlation that gives rise to the dynamics of the aggregate Z_t in Eq. 8.15. The correlation is crucial, but it turns out that it does not need to be complex. The case $\chi^1 = \chi^2 = 1$, $\alpha_1 = 0.50$, $\alpha_2 = 0.50$, and $a_i = b_i = 0.25$ is the case of REE. Kurz and Motolese (2001) postulate a simple model with $M(a) = M(b)$; hence beliefs are not correlated with dividends. However, beliefs are correlated with a simple description of $a = b = (.50, .14, .14, .14)$. This simple parametrization implies that the dynamics of prices have the feature that bull and bear markets are *asymmetric*. For the market to transit from the "crash" state of the lowest price to the states of the highest prices, it must take several steps: It cannot go *directly* from the low to the high prices. The opposite is possible, since at the bull market states there is a positive probability of reaching the crash states in one step. *This implies that a bull market that reaches the high prices must evolve in several steps, but a crash can occur in one step.*

To sum up, there are three classes of parameters, and simulation work explores *only regions of the parameter space that are compatible with rationality*. Kurz and Motolese (2001) report a set of parameters under which the model replicates the empirical record with great accuracy. These are: utility function parameters: $2.00 \leq \gamma \leq 3.00$, the common risk aversion coefficient; $0,90 \leq \beta \leq 0.95$, the common discount rate; correlation parameters: $a_1 = 0.50$, $a_2 = a_3 = a_4 = 0.14$; belief parameters: $\chi^1 = \chi^2 = 1.7542$, the maximal intensity permitted by rationality; and $\alpha_1 = \alpha_2 = 0.57$, the frequency with which an agent is in an optimistic state. The model replicates well the empirical record, and the results are similar to those of infinite horizon models. Hence there is no point repeating the same results twice, and we report the precise numerical results only for the infinite horizon models[13]. We thus turn to models with infinite belief states.

Models with infinite horizon and infinite belief states typically have two assets: A stock paying dividends and a zero net supply bond. The model has a large number of identical agents of two types with the same utility and endowment. Across types they differ in their beliefs. For consistency we use a model developed by Kurz et al. (2005a) to illustrate the belief structure and report results for this simulation model. This has the merit that all simulation results reported *are derived from a single model*. These authors assume $D_{t+1} = D_t e^{v_{t+1}}$ with an empirical distribution of the growth rate

$$v_{t+1} = (1 - \lambda_v)v^* + \lambda_v v_t + \rho^v_{t+1} \quad \rho^d_{t+1} \sim N(0, \sigma^2_d)\text{IID}$$

where v^* is the unconditional mean. Given his belief, agent i maximizes an infinite horizon expected utility with date t utility of $\beta^t[1/1-\gamma](c^i_t)^{1-\gamma}$.

To explain the perception models of the agents, we could have postulated $v^i_{t+1} = (1 - \lambda_v)v^* + \lambda_v v_t + \lambda^g_v g^i_t + \tilde{\rho}^{v^i}_{t+1}$ as in Eq. 8.17a. Such a model is sufficient for the conceptual needs of the illustrative model, but it would contradict the finite state model reviewed earlier. The reason is that the belief state g^i_t is a *symmetric variable* that does not meet two of the three principles advocated earlier, namely the condition of *asymmetry* and *fat tails* in the belief densities. Limited space permits us to present only a sketch of the complex structure in Kurz et al. (2005a)[14]. To introduce asymmetry and fat tails, the procedure we follow is to transform the g^i_t into a new random variable $\eta^i(g^i_t)$ and define the perception model for the growth rate of dividend to be (8.30)

$$v^i_{t+1} = (1 - \lambda_v)v^* + \lambda_v v_t + \lambda^g_v \eta^i_{t+1}(g^i_t) + \tilde{\rho}^{v^i}_{t+1} \quad \tilde{\rho}^{v^i}_{t+1} \sim N(0, \tilde{\sigma}^2_v)\text{IID}$$

We then look for asymmetry and fat tails in the density of $\Delta^i_{t+1}(g^i_t) = \lambda^g_v \eta^i_{t+1}(g^i_t) + \tilde{\rho}^{v^i}_{t+1}$, *conditioned* on g^i_t. Keep in mind that for a computational model, we must choose a specific functional form, and this may appear to be too strong a set of assumptions about beliefs.

[13]For details of the finite belief state results, see Kurz and Motolese (2001, pp. 530–533). For computational procedures to reproduce these results, go to www.stanford.edu/∼mordecai/ and click "computable models with heterogeneous beliefs." Keep in mind that an OLG model has a unique market-oriented feature not shared by an infinite horizon model that requires an agent to sell his position when old, regardless of his beliefs. This feature is important for a model of market volatility, since agents who aim to preserve capital by holding a portfolio of a riskless asset must sell the asset into the market, regardless of their beliefs. This fact tends to generate additional volatility that would not be present in an infinite horizon model. This feature has two results that are not shared by the infinite horizon model. First, the riskless rate has a much larger standard deviation in simulated OLG equilibria than in the infinite horizon models. Second, to generate a low average riskless rate, a result needed to replicating the 6–7% equity premium, it is necessary to assume an asymmetry where the majority of agents are optimists about earning abnormal excess returns and the frequency of optimism is greater than 50%. In the infinite horizon model, it is necessary to have the pessimists in the majority, with a frequency of pessimism being more than 50%. We shall comment on this issue again later, when we discuss the Equity Premium Puzzle.

[14]Indeed, the main deficiency of the Kurz et al. (2005a) model is its complexity, which, in our view today, could have been avoided. Both the model itself as well as the computational procedures could have been drastically simplified, since the basic ideas are rather simple, as explained in Section 8.4.1.

How do we know that belief densities take these specific forms? We do not. But here we return to the three *qualitative properties* discussed in Section 8.4.1. What drives the results are not the functional forms selected for a computational model but their qualitative properties. Any other functional form with the same qualitative properties would generate the same results but, naturally, with different parametrization. Since $\eta^i(g_t^i)$ cannot be a simple symmetric variable, Kurz et al. (2005a) specify the conditional distribution of $\eta_{t+1}^i(g_t^i)$. They define it as a step function:

$$P(\eta_{t+1}^i | g_t^i) = \begin{cases} \varphi_1(g_t^i)[1/\sqrt{2\pi}]e^{-\frac{\eta_{t+1}^{i\,2}}{2}} & \text{if } \eta_{t+1}^i \geq 0 \\ \varphi_2(g_t^i)[1/\sqrt{2\pi}]e^{-\frac{\eta_{t+1}^{i\,2}}{2}} & \text{if } \eta_{t+1}^i < 0 \end{cases}$$

where η_{t+1}^i and $\tilde{\rho}_{t+1}^{v^i}$ (in Eq. 8.30) are independent. The functions $(\varphi_1(g), \varphi_2(g))$ are defined by a logistic function with two parameters κ and Λ:

$$\varphi(g^i) = \frac{1}{1 + e^{\Lambda(g^i - \kappa)}}, \kappa < 0, \Lambda < 0 \quad \text{and} \quad \varphi_1(g^i) = \frac{\varphi(g^i)}{E_g \varphi(g^i)}, \varphi_2(g^i) = 2 - \varphi_1(g^i)$$

The parameter κ measures *asymmetry*, and the parameter Λ measures *intensity of fat tails* in beliefs. When $g_t^i > \kappa$, then $E_i[\eta_{t+1}^i | g_t^i] > 0$. Choosing $\lambda_v^g > 0$ implies that when $g_t^i > \kappa$ agent i is in a *bull state* and is optimistic about $t + 1$ dividend growth being above-normal. Since $\kappa < 0$, it also implies that bull states occur with frequency higher than 50%. "Normal" is defined relative to the empirical forecast. In sum, for the basic case $\lambda_v^g > 0$:

- $g_t^i > \kappa$ means that agent j is optimistic about profit growth and excess stock returns at $t + 1$.
- $g_t^i < \kappa$ means that agent j is pessimistic about profit growth and excess stock returns at $t + 1$.

The parameter κ measures asymmetry and determines the frequency at which agents are bears, and when $\kappa < 0$ the probability of $g^i > \kappa$ is more than 50%. The density of η_{t+1}^i is exhibited in Figure 8.4 and shows that asymmetry arises from a redistribution of the probability mass. However, the empirical distribution of $\eta_{t+1}^j(g_t^j)$ averaged over time and over the g_t^i is Normal.

Each component of $\Delta_{t+1}(g_t^i)$ is a sum of two random variables: one as in Figure 8.4 and the second is normal. In Figure 8.5 we draw two densities of $\Delta_{t+1}(g_t^j)$, each being a convolution of the two constituent distributions, one density for $g^i > \kappa$ and a second for $g^i < \kappa$, showing both have "fat tails." Since Λ measures intensity by which the positive portion of the distribution in Figure 8.4 is shifted, it measures the degree of fat tails in the distributions of $\Delta_{t+1}(g_t^j)$.

The assumption of a power utility $\beta^t[1/1 - \gamma](c_t^i)^{1-\gamma}$ implies that income effects matter and beliefs do not aggregate. Hence, the state variable in the simulation model is the actual distribution of beliefs. Since there is a large number of identical agents of *two types*, this distribution is a vector (z_t^1, z_t^2). The fact that we denote it by (z_t^1, z_t^2) and

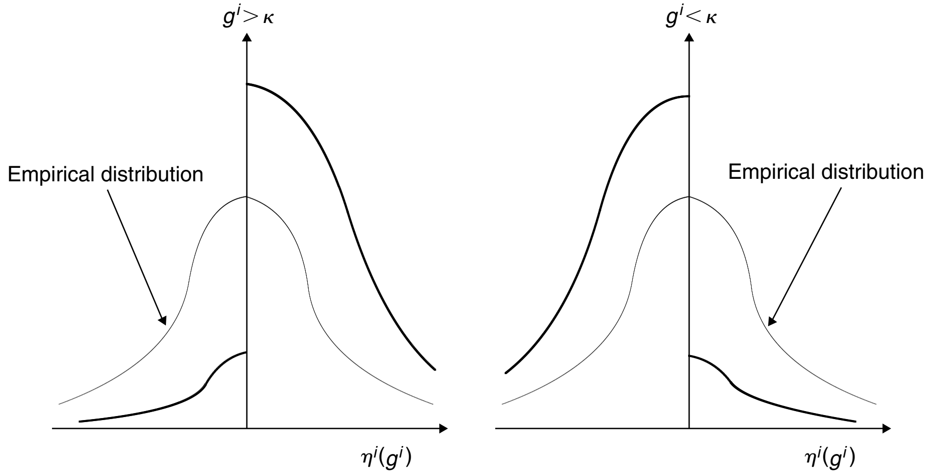

FIGURE 8.4 Nonnormal belief densities.

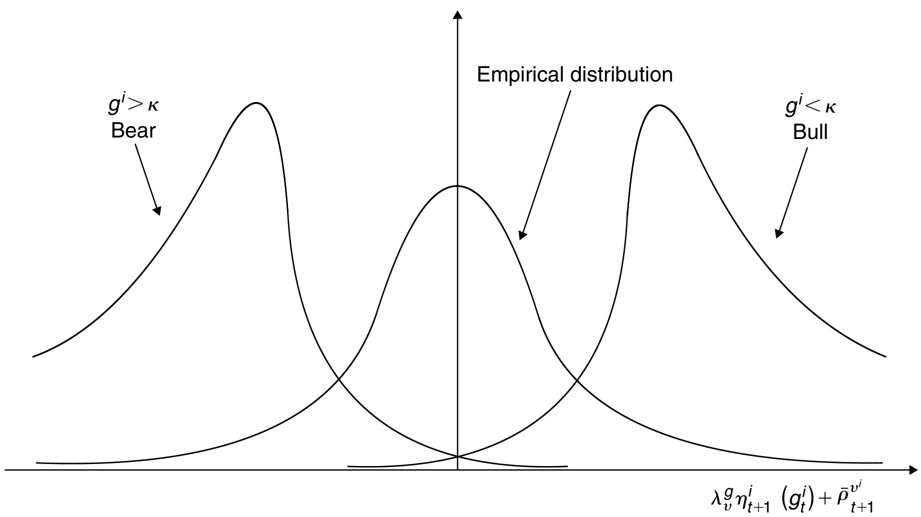

FIGURE 8.5 Density $\Delta^i_{t+1}(g^i_t)$ with fat tails.

not (g_t^1, g_t^2) is an important technical issue arising due to the assumption of anonymity. Agent i knows his own belief as g_t^i, which he uses to forecast all state variables in the economy, whereas (z_t^1, z_t^2) are observed state variables and the agent uses g_t^i to forecast v_{t+1} as in Eq. 8.30 and (z_{t+1}^1, z_{t+1}^2) with a fully developed perception model that is analogous to Eqs. 8.17a–8.17c in the illustrative model we have developed here. For technical details of the perception model and the implied rational belief restrictions, see

Kurz et al. (2005a, pp. 12–19). We now turn to a detailed examination of the simulation results of the finite and infinite belief state model.

Explaining the Volatility Moments

We report simulation results of models with infinite belief states. *All results reported in this chapter are derived from a single model by Kurz et al.* (2005a). In these simulations Kurz et al. compute various measures of volatility using 20,000 observations. Raw moment calculations were carried out by Kurz et al. (2005a) for the following list of long-term volatility measures, which are moments in accord with the stationary measure:

- q: Average price/dividend ratio
- σ_q: Standard deviation of the price/dividend ratio q_t
- R^q: Average risky return on equity
- σ_R: Standard deviation of R^q
- r: Average riskless rate of interest
- σ_r: The standard deviation of r
- e_p: Average the equity premium
- $\rho(d, R^q)$: Correlation between the risky rate and dividends' growth; it is also the correlation between consumption growth and the risky rate as consumption and dividends grow at the same rate
- $(Shrp)$: The Sharpe ratio

Table 8.2 reports the results.[15] The model matches *simultaneously* the moments and, as we see later, it also matches most other features of market dynamics. Kurz et al. (2005a) further observe that the results in Table 8.2 are not due to the particular beliefs used or their parameter values. They are due to fat tails in asset returns, to asymmetry, and to correlation of beliefs as explained in the previous section. Are these three key characteristics supported by the data?

The fact that the distribution of asset returns exhibits fat tails is well documented (e.g., see Fama, 1965, and Shiller, 1981). It is natural to ask where these tails come from. The theory at hand says they come from fat tails in the probability models of agents' beliefs. Correlation of beliefs across agents is documented in sources such as the Blue Chip Financial Forecasts. The evidence in support of the hypothesis that the frequency of bear states is higher than 50% is more complicated.

The hypothesis is supported by the empirical fact that, on average, most above-normal stock returns are realized over *relatively small proportion of time* when asset prices rally (see Shilling, 1992). It is thus reasonable that, on average, the proportion of time when agents expect to make above-normal returns is less than 50%[16]. Additional

[15]For computational procedures to reproduce the simulation results of the Kurz et al. (2005a) model click on "computable models with heterogeneous beliefs" at *www.stanford.edu/∼mordecai/.*

[16]Shilling (1992) shows that during the 552 months from January 1946 through December 1991, the mean real annual total return on the Dow Jones Industrials was 6.7%. However, if an investor missed the 50 strongest months, the real mean annual return over the other 502 months was −0.8%. Hence the financial motivation to time the market is very strong, as is the case for the agents in the model.

TABLE 8.2 Simulation Results of Kurz et al. (2005a) (*all moments are annualized*)

Moment	Simulation Results	Empirical Record* (1889–1998)
q	25.54	25.00
σ_q	5.46	7.10
R^q	7.57%	7.00%
σ_R	18.81%	18.00%
r	1.08%	1.00%
σ_r	5.44%	5.70%
e_p	6.49%	6.00%
$\rho(d, R)$	0.21	0.10
Shrp	0.34	0.33

*The main data source for the empirical record is Shiller at *www.econ.vale.edu/~shiller/data.htm.* It was updated by Kurz et al. (2005a, page 23) to 1998. Since the discussion here does not aim to evaluate the precision of the estimates, the numbers in the table were rounded off to indicate orders of magnitude.

indirect support comes from the psychological literature, which suggests that agents place heavier weight on losses than on gains. In the treatment here, agents fear losses at majority of dates since on those dates they place higher probabilities of abnormally lower returns. By the rational belief principle, a higher frequency of bear states implies that in bull states, an agent's intensity of optimism is higher than the intensity of pessimism. This means that the average positive tail in the belief densities is bigger than the average negative tail. That is, the asymmetry hypothesis implies that optimistic agents tend to be intense.

Together with correlation of beliefs across agents, this hypothesis also implies that we should observe periods of high optimism for a majority of agents. Optimism leads to agents' desire to borrow and finance present and future consumption. At such dates the only way for markets to clear is by exhibiting sharp rises in stock prices together with high borrowing rates. Hence, this theory predicts that we should expect to observe rapidly rising stock prices induced by bursts of optimism correlated with high realized growth rates of dividends. The structure of correlation also implies that we should also expect to see crashes induced by correlated pessimistic agents together with low realized growth rates of dividends.

We make one comment with respect to the low riskless rate. Matching many volatility moments *except the riskless rate* depend mostly on intensity and correlation. These moments exhibit relatively low sensitivity to asymmetry. Hence, apart from the riskless rate, many long-term volatility measures are explained by a broad configuration of the intensity parameters and correlation across agents' beliefs. The low riskless rate and a few others require asymmetry.

Why Does the Model of Diverse Beliefs Resolve the Equity Premium Puzzle?

Risk premia are compensations for risk perceived by risk-averse agents. In single-agent models, the market portfolio is identified with a security for which the payoff is aggregate consumption, mostly taken to be exogenous. The Equity Premium Puzzle is thus an observation that the small volatility of aggregate consumption growth cannot justify a 6% equity premium, given the degree of risk aversion. The theory of diverse beliefs offers a resolution of the puzzle by studying optimal behavior and consumption growth rate volatility on the individual, not the aggregate, level.

Any theory of diverse beliefs implies that at each date the risk premium perceived by an agent is subjective. The risk premium required by an investor with a bullish outlook is smaller than the risk premium required by a bear. Hence to resolve the Equity Premium Puzzle a theory must explain why some agents are willing to hold a riskless asset paying a real return of only 1% when the average return on the risky stock is 7%. The 7% return on the stock is entirely explainable by fundamental factors of growth and productivity, together with the added high volatility of returns induced by factors of intensity and correlation of beliefs that generate endogenous uncertainty. The problem is the low riskless rate. Pessimistic agents who aim to preserve capital are willing to earn low return on their investment, and with enough of them around, the riskless rate would indeed fall. But can a desire to preserve capital by those avoiding the risky stock be compatible with fat tails in returns? This is the role of *asymmetry*. Symmetry between bulls and bears generates fat tails only due to intensity and correlation. Agents are intense when they are bulls and correlation causes the majority sentiment to fluctuate. Fat tails then reflect fluctuations of the majority between bull or bear averages. After all, when a majority of agents try to sell or buy the stock, the price fluctuates. But to push the riskless rate down we need the asymmetric persistence of the bear view by those who expect the stock to deliver low excess returns. Expecting low excess returns, they would rather avoid the risky stock and hold the bond at lower return. The fact that bears are in the majority of investors at the majority of dates constitutes the extra factor, which lowers the riskless rate as well.

We turn to the low volatility of aggregate consumption growth. Diverse beliefs cause diverse individual consumption growth rates, even if aggregate consumption is exogenous, which is the case in the models here. This is true not only because of idiosyncratic factors but also because under diverse beliefs markets are inherently incomplete and the representative agent model does not capture the conditions of individual consumers. Hence, volatility of individual consumption growth rates is higher than the volatility of the aggregate rate, an empirical fact supported by household survey data. Since agents' perceived volatility of their own consumption growth is different from the aggregate rate, they do not seek to own a portfolio whose payoff is aggregate consumption. Consequently, we must not focus on the relation between asset returns and aggregate consumption growth but instead on the relation between perceived asset returns and *perceived volatility of individual consumption growth*. The key question is, then, how volatile do individual consumption growth rates need to be to generate an equity premium of 6% and a riskless rate of 1%? The answer is: *not very much*. Relative to their equilibrium, Kurz et al. (2005a) report that although the standard deviation of the aggregate consumption growth rate is 0.03256, the standard deviation of individual consumption growth rates supporting the premium in the simulations is only 0.039, and the required correlation between individual consumption growth rate and the growth rate of dividends is only 0.83 (compared to 1.00 in a representative household model). Both figures are compatible with survey data.

Predictability of Stock Returns

The problem of predictability of risky returns generated a large literature in empirical finance (e.g., Fama and French, 1988a, 1998b; Poterba and Summers, 1988; Campbell and Shiller, 1988; and Paye and Timmermann, 2003). This debate originates in the theoretical observation that under risk aversion asset prices and returns are not martingales and contain a predictable component. In this context Kurz et al. (2005a) use the model associated with Table 8.2 to generate simulated data with which they examine the following: (1) variance ratio statistic; (2) autocorrelation of returns and of price/dividend ratios; and (3) predictive power of the dividend yield. They then apply to simulated data the standard tests used for market data, and we report their results in the following subsections. Recalling that v_t is the growth rate of dividends and q_t is the price dividend ratio, the standard notation used in this literature is to let $\varrho_t = \log[\frac{(q_t+1)e^{v_t}}{q_{t-1}}]$ be the log of gross one year stock return, $\varrho_t^k = \sum_{i=0}^{k-1} \varrho_{t-i}$ be the cumulative log-return of k-year length from $t-k+1$ to t, and $\varrho_{t+k}^k = \sum_{j=1}^{k} \varrho_{t+j}$ be the cumulative log-return over a k-year horizon from $t+1$ to $t+k$.

Variance Ratio Test The variance-ratio is $VR(k) = \frac{var(\varrho_t^k)}{(k\, var(\varrho_t))}$. If returns are uncorrelated, this ratio converges to 1 as k rises. If returns are negatively autocorrelated at some lags, the ratio is less than one. Kurz et al. (2005a) show there exists a significant higher-order autocorrelation in simulated stock returns and hence a long-run predictability that is consistent with U.S. data on stock returns, as in Poterba and Summers (1988). In Table 8.3 Kurz et al. (2005a) report the computed values of the ratios for $k = 1, 2, \ldots, 10$ and compare them with the ratios in the empirical record reported by Poterba and Summers (1988, Table 8.2, line 3) for $k = 1, 2, \ldots, 8$. The model's prediction is close to the U.S. empirical record.

TABLE 8.3 Variance Ratios for NYSE 1926 to 1985

k	1	2	3	4	5	6	7	8	9	10
$VR(k)$	1.00	0.85	0.73	0.64	0.57	0.51	0.46	0.41	0.38	0.34
U.S.	1.00	0.96	0.84	0.75	0.64	0.52	0.40	0.35	—	—

The Autocorrelation of Log-Returns and Price–Dividend Ratios In Table 8.4 we report the Kurz et al. (2005a) autocorrelation function of log annual returns. The model

TABLE 8.4 Autocorrelation of Log-Returns

corr(ϱ_t, ϱ_{t-k})	Model	Empirical Record
$i = 1$	−0.154	0.070
$i = 2$	−0.094	−0.170
$i = 3$	−0.069	−0.050
$i = 4$	−0.035	−0.110
$i = 5$	−0.040	−0.040

TABLE 8.5 Autocorrelation of Price–Dividend Ratio

corr(q_t, q_{t-k})	Model	Empirical Record
$i = 1$	0.695	0.700
$i = 2$	0.485	0.500
$i = 3$	0.336	0.450
$i = 4$	0.232	0.430
$i = 5$	0.149	0.400

predicts negatively autocorrelated returns at all lags. This implies a long horizon mean reversion of the kind documented by Poterba and Summers (1988), Fama and French (1998a), and Campbell and Shiller (1988). Thus, apart from the very short returns that exhibit positive autocorrelation, the model reproduces the empirical record.

In Table 8.5 we report the autocorrelation function of the price-dividend ratio reported by Kurz et al. (2005a). The table shows the model generates a highly autocorrelated price/dividend ratio, which matches reasonably well the behavior of U.S. stock market data. The empirical record in Tables 8.4 and 8.5 is for NYSE data for 1926–1995 as reported in Barberis et al. (2001).

Dividend Yield as a Predictor of Future Stock Returns The papers cited previously show that the price–dividend ratio is the best explanatory variable of long returns. To test this fact Kurz et al. (2005a) consider the following regression model:

$$\varrho_{t+k}^{k} = \zeta_k + \eta_k(e^{v_t}/q_{t-1}) + \vartheta_{t,k} \tag{8.31}$$

e^{v_t}/q_{t-1} is the dividend yield since it is the ratio between dividend paid at t and the stock price at $t-1$.

Fama and French (1988b) report that the ability of the dividend yield to forecast stock returns, measured by regression coefficient R^2 of Eq. 8.31, increases with the return horizon. Kurz et al. (2005a) find that the model captures the main features of the empirical evidence as reported in Table 8.6.

TABLE 8.6 The Behavior of the Regression Slopes in Eq. 8.31

Time Horizon	Model		Empirical Record	
k	η_k	R^2	η_k	R^2
1	5.03	0.08	5.32	0.07
2	8.66	0.14	9.08	0.11
3	11.16	0.18	11.73	0.15
4	13.10	0.21	13.44	0.17

To conclude the discussion of predictability, observe that the empirical evidence reported by Fama and French (1998a, 1998b), Campbell and Shiller (1988), Poterba and Summers (1998), and others is consistent with asset price theories in which time-varying expected returns generate predictable, mean-reverting components of prices (see Summers, 1986). The important question left unresolved by these papers is, what drives the predictability of returns implied by such mean-reverting components of prices? Part of the answer is the persistence of the dividend growth rate. Kurz et al. (2005a) offer a second and stronger persistent mechanism. It shows that these results are primarily driven by the dynamics of market state of beliefs, which exhibit correlation across agents and persistence over time. Agents go through bull and bear states, causing their perception of risk to change and expected returns to vary over time. Equilibrium asset prices depend on states of belief that then exhibit memory and mean reversion. Hence returns exhibit these same properties.

GARCH Behavior of the Price–Dividend Ratio and of the Risky Returns

Stochastic volatility in asset prices and returns is well documented (e.g., Bollerslev, Engle, and Nelson, 1994; Brock and LeBaron, 1996). In partial equilibrium finance it is virtually standard to model asset prices by stochastic differential equations, *assuming* an exogenously driven stochastic *volatility*. But where does stochastic volatility come from? Dividends certainly do not exhibit stochastic volatility. We now show that models with diverse beliefs *can explain why asset prices and returns exhibit stochastic volatility*.

To formally test the GARCH property of the price–dividend ratio and of the risky returns, Kurz et al. (2005a) use the 20,000 simulated observations noted in the "Equity Premium Puzzle" section. With that data they estimate the following econometric model of the dynamics of the log of the price–dividend ratio:

$$\begin{aligned} \log(q_{t+1}) &= \kappa^q + \mu^q \log(q_t) + \zeta^q_{t+1} \\ \zeta^q_t &\sim N(0, h^q_t) \\ h^q_t &= \xi^q_0 + \xi^q_1(\zeta^q_{t-1})^2 + \nu^q_1 h^q_{t-1} \end{aligned} \tag{8.32a}$$

Since the price–dividend ratio is postulated to be an AR(1) process, the process in Eq. (8.32a) is GARCH(1,1). Similarly, for the risky rates of return, they postulated the model

$$\begin{aligned} \varrho_{t+1} &= \kappa^{\varrho} + \mu^{\varrho}\log(q_{t+1}) + \varsigma^{\varrho}_{t+1} \\ \varsigma^{\varrho}_t &\sim N(0, h^{\varrho}_t) \\ h^{\varrho}_t &= \xi^{\varrho}_0 + \xi^{\varrho}_1(\varsigma^{\varrho}_{t-1})^2 + v^{\varrho}_1 h^{\varrho}_{t-1} \end{aligned} \tag{8.32b}$$

For a specification of Eqs. 8.32a and 8.32b, they also tested ARCH(1) and GARCH(2,1) but concluded that the proposed GARCH(1,1) describes best the behavior of the data. Due to the large sample they ignore standard errors and report that the estimated model for the log of the price–dividend ratio satisfies the GARCH(1,1) specification:

$$\begin{aligned} \log(q_{t+1}) &= 0.99001 + 0.69384\log(q_t) + \varsigma^{q}_{t+1} \\ \varsigma^{q}_t &\sim N(0, h^{q}_t) \\ h^{q}_t &= 0.00592 + 0.02370(\varsigma^{q}_{t-1})^2 + 0.73920 h^{q}_{t-1},\ R^2 = 0.481 \end{aligned}$$

For risky rates of return, the estimated model satisfies the GARCH(1,1) specification

$$\begin{aligned} \varrho_t &= 1.13561 - 0.33355\log(q_t) + \varsigma^{\varrho}_t \\ \varsigma^{\varrho}_t &\sim N(0, h^{\varrho}_t) \\ h^{\varrho}_t &= 0.00505 + 0.01714(\varsigma^{\varrho}_{t-1})^2 + 0.77596 h^{\varrho}_{t-1}, \quad R^2 = 0.180 \end{aligned}$$

To explain we observe that stochastic volatility is a direct consequence of the dynamics of beliefs, defined by (z^1_t, z^2_t) in Kurz et al. (2005a). Persistence of beliefs and correlation across agents introduce these patterns into prices and returns. When agents disagree (i.e., $z^1_t z^2_t < 0$), they offset the demands of each other and as that pattern persists, prices do not need to change by very much for markets to clear. During such periods prices exhibit low volatility: persistence of belief states induce persistence of low volatility. When agents agree (i.e., $z^1_t z^2_t > 0$) they compete for the same assets and prices are determined by difference in belief intensities. Changes in the levels of bull or bear states generate high volatility in asset prices and returns. Persistence of beliefs cause such high-volatility regimes to exhibit persistence. Market volatility is then time dependent and has a predictable component as in Eqs. 8.32a and 8.32b.

The virtue of the above argument is that it explains stochastic volatility *as an endogenous consequence of equilibrium dynamics*. Some "fundamental" shocks (i.e., an oil shock) surely cause market volatility, but it has been empirically established that market volatility cannot be explained *consistently* by "fundamental" exogenous shocks (e.g., Schwert, 1989; Pesaran and Timmermann, 1995; and Beltratti and Morana, 2006). The Kurz et al. (2005a) explanation of stochastic volatility is thus consistent with the empirical evidence.

8.4.3. Volatility of Foreign Exchange Rates and the Forward Discount Bias

The relevance of foreign exchange markets to our discussion in this chapter is motivated by the following problem. Estimate a regression of the form

$$\frac{ex_{t+1} - ex_t}{ex_t} = c + \zeta(r_t^D - r_t^F) + \epsilon_{t+1} \tag{8.33}$$

where $(ex_{t+1} - ex_t)$ is the change of the exchange rate between t and $t+1$ and $(r_t^D - r_t^F)$ is the difference between the short-term nominal interest rates in the domestic and the foreign economies. Under rational expectations $(r_t^D - r_t^F)$ is an unbiased predictor of $(ex_{t+1} - ex_t)$. It is motivated by a standard arbitrage argument: If there is a differential in nominal rates, agents can borrow in one country and invest in the other and gain from the difference if the exchange rate does not move against them by the amount of the differential. In a no-arbitrage REE, a rationally expected change in the exchange rate must then be equal to the interest rate differential. This means that apart from a technical correction for risk aversion, the parameter ζ should be close to 1. In 75 empirical studies ζ was estimated to be significantly less than 1 and in many studies it was estimated to be negative (see Froot and Frankel, 1989; Frankel and Rose, 1995; and Engel, 1996, for an extensive survey).

The failure of ζ to exhibit estimated values close to 1 is known as the *forward discount bias* in foreign exchange markets. The empirical fact is that exchange rates are far more volatile than can be explained by differentials in nominal interest rates or inflation rates between countries. But changes in foreign exchange rates are not predictable and interest rate differentials account only for a small fraction of the movements in foreign exchange rates. However, it is not surprising that this lack of predictability decreases with the length of time involved. That is, long-run differentials in nominal interest rates do exhibit better predictive power of long-run movements in foreign exchange rates, since long-run differentials in nominal rates reflect differentials in inflation rates. Since the problem at hand is the nature of market expectations and exchange rate volatility, it is a natural for us to consider it here, and the model of diverse belief is an obvious candidate to be used to solve the problem.

Applying the rational belief theory to this problem, Kurz (1997b) and Black (1997), (2005) developed a model that is similar to the Kurz and Motolese (2001) model except for treating the second agent as a second country and adding two nominal debt instruments. A similar model was also reformulated by Kurz and Motolese (2001). Limitation of space makes it impossible to review all technical details of these models here. Instead, we outline the key points of the model construction and note the results. Hence, the central model construction elements in these papers are as follows:

- Consider the first agent as the "domestic United States," which is the home country, and the second agent as a "foreign economy"
- Introduce a second shock that is associated with productivity in the foreign economy and is different from the first shock defined for productivity in the domestic economy

- Introduce a monetary system for both countries and a second currency
- Introduce two nominal interest rates and two different monetary policies
- There are the two standard financial assets: (1) ownership shares of a domestic firm with stochastic dividend whose stock trades freely in both currencies across the countries, and (2) a zero net supply riskless bond that pays a unit of consumption and that trades in both currencies across the countries
- Introduce a simple production structure for the foreign economy

Note that these models do not aim to simulate the United States or world economies; they merely aim to explain via simulations why a model with diverse beliefs implies $\zeta < 1$. And indeed, all models produce estimated parameters ζ, that are significantly less than 1: In the *rational belief equilibrium* of Kurz (1997b) the estimated ζ is around 0.25, in Black (1997, 2005) it is around 0.15, and in Kurz and Motolese (2001) it is around 0.45. More realistic results could be obtained by formulating more realistic models, but the key result $\zeta < 1$ is virtually independent of the model formulation. We now provide an explanation for this strong conclusion.

Why do diverse beliefs predict that ζ is less than 1? If $\zeta < 1$ in an REE, agents can make an *expectational* arbitrage: They can borrow today in one currency, invest in the other, and *expect* that the net return on their investment *next period* will be larger than the depreciation of the currency. In such an equilibrium all agents hold the same self-fulfilling expectations; the *expectational* arbitrage becomes a *real* arbitrage and consequently this implies that $\zeta < 1$ cannot hold in equilibrium.

In a world with diverse beliefs, equilibrium exchange rate depends on the distribution of beliefs and hence exchange rates exhibit excess volatility, reflecting the variability of investors' beliefs. Indeed, volatility of foreign exchange rates is dominated by endogenous uncertainty. The implication is that *regardless of the information today*, to forecast future exchange rates agents must forecast future market states of belief, rendering exchange rates virtually unpredictable. Hence, if a condition of differential nominal interest rates across countries arises, it can never be the only factor that will determine the exchange rate next period. With risk-averse agents who are unable to predict the exchange rate, a condition of differential interest rates will not generate the beliefs of traders that the exchange rate will, in fact, adjust. Failing to expect the exchange rate to adjust, they will not undertake such arbitrage and the exchange rate will, in fact, not adjust. This mechanism ensures that a differential of nominal interest rates between the two countries is not an unbiased estimate of the rate of depreciation of the exchange rate one period later, hence $\zeta < 1$. This reasoning does not hold in the long run since a long-term differential of nominal interest rates will persuade the markets that the exchange rate must adjust in the long run and this will persuade them to engage in such arbitrage.

8.4.4. Macroeconomic Applications

Although there is a wide range of potential applications in macroeconomics, so far only limited questions have been studied with the model of diverse beliefs. Motolese (2001, 2003) shows that in an economy with diverse beliefs, money is not neutral. To see why, it is important to observe that before *rational expectations* the case for money neutrality was based on the quantity theory of money. The main contribution of Lucas (1972) was to show that money neutrality can be proved only by an exploration of the structure of expectations. In a model with heterogeneous beliefs, agents hold diverse beliefs about the relative effects of productivity growth and money shocks; hence they hold diverse beliefs about future inflation. With diverse expectations money cannot be neutral.

Kurz et al. (2005b) is a comprehensive study of the efficacy of monetary policy in an economy with diverse beliefs. The authors show that diverse beliefs constitute an important propagation mechanism of fluctuations, money nonneutrality, and efficacy of monetary policy. Since expectations affect demand, the theory shows that economic fluctuations are driven mostly by varying demand, not supply shocks. Using a competitive model with flexible prices in which agents hold rational beliefs, the authors arrive at six conclusions:

1. The model economy replicates well the empirical record of fluctuations in the United States.
2. Under monetary rules without discretion, monetary policy has a strong stabilization effect, and an aggressive anti-inflationary policy can reduce inflation volatility to zero.
3. The statistical Phillips Curve changes substantially with policy instruments, and activist policy rules render it vertical.
4. Although prices are flexible, money shocks result in less than proportional changes in inflation; hence the aggregate price level is "sticky" with respect to money shocks.
5. Discretion in monetary policy adds a random element to policy and increases volatility. The impact of discretion on the efficacy of policy depends on the structure of market beliefs about future discretionary decisions. The paper studies two rationalizable beliefs. In one case, market beliefs weaken the effect of policy; in the second, beliefs bolster policy outcomes. Therefore, in this case, discretion is a desirable attribute of the policy rule. That is, social gains from discretion arise only under special structures of belief of the private sector about future bank discretionary acts, and such requirement complicates the bank's problem. Hence, the weight of the argument leads Kurz et al. (2005b) to conclude that a bank's *policy should be transparent and abandon discretion except for rare and unusual circumstances*. This analysis is in contrast to the recent literature initiated by Morris and Shin (2005) and others who suggest that due to asymmetric private information, central bank transparency has inherent *cost of failing to retrieve useful private information* by the bank. We have rejected the applicability of the private information model for the study of economic aggregates such as interest rates, inflation rate, or GDP growth. Hence, the Morris and Shin (2005) model does not address the real problem associated with the objective of central bank transparency, which is the coordination of expectations.
6. One implication of the model suggests that the present-day policy is only mildly activist and aims mostly to target inflation.

8.5. CONCLUSION AND OPEN PROBLEMS

Rational expectations and irrational behavior are extreme hypotheses; with REE one cannot explain the observed data on market dynamics and with irrational behavior one can prove anything. We highlight here the merit of an intermediate concept of belief rationality that emerges from the fact that the economy is a nonstationary system with time-varying structure. This prevents agents from ever learning true structural relations and probability laws. All they learn are the empirical frequencies from which emerge a common knowledge of a stationary probability reflecting the long-term dynamics. Belief rationality requires agents to hold only beliefs that are not contradicted by the empirical evidence. But since it is irrational to believe in a fixed deviation from the stationary probability, such belief rationality implies belief dynamics: Individual beliefs must be time varying, and correlation across agents generates a new aggregate force in market dynamics, which is the dynamics of market belief.

The main observation made in this chapter is that the dynamics of market beliefs are a central market force that is as important to asset pricing and allocation as the dynamics of productivity or public policy. Indeed, the dynamics of market beliefs explain well the four recessions Samuelson noted that the market predicted but that did not happen. It shows that a rational market makes forecasting mistakes and rational investors are not infallible. They may use wrong forecasting models. Once we recognize that being rational and being wrong are not incompatible and that no psychological impulses are needed for this proposition, we are open to a new paradigm of market dynamics. This paradigm provides a coherent explanation to most dynamical phenomena of interest as outlined in this chapter. We thus sum up our five central conclusions:

1. Diverse beliefs without any private information are an empirical fact, and such diversity provides a strong motive to trade assets and hedge subjectively perceived risks.
2. Financial markets are the great arena for agents to trade differences in beliefs.
3. The dynamics of market beliefs are a central component of asset price volatility, and this component of risk has been named *endogenous uncertainty*. Market belief is observable.
4. Asset markets exhibit large excess volatility of prices, returns, and high volume of trade due to the dynamics of beliefs.
5. Risk premia reflect the added market risk due to the dynamics of beliefs and in some markets the component of risk premia due to the dynamics of market beliefs is very significant.

Important problems that we have not discussed are still open, some of which are being researched at this time. Five examples are as follows:

1. *Pareto optimality*. The concept of ex-ante Pareto optimum is not a satisfactory concept for a market with diverse beliefs. To attain any Pareto improvement, all agents must believe it is an improvement, and that is not likely. Hence, most stabilization policies would not be Pareto improving. Following the idea of ex-post Pareto optimality (e.g., Starr, 1973, and Hammond, 1981), progress on this issue was made

by Nielsen (2003, 2006), who argued that a currency union is superior to multiple currencies since a union would eliminate endogenous uncertainty inherent in foreign exchange rates.

2. *Stabilization policy*. When the problem of Pareto optimality is resolved, the door will be open to a study of the desirability of stabilization public policies. Some start has been made by Kurz et al. (2005b) regarding stabilizing monetary policy. But the question is broader. Should the Fed target the stock market? Should countries cooperate to avoid an international financial crisis? What is the role of an international convention regarding bank reserve requirements? Under REE these types of questions are set aside since it is often argued that the market solution is best and no cooperative policy is needed. In a world of diverse beliefs, this is not true and the question is open.

3. *Continuous time reformulation*. A continuous time reformulation of the RB theory would open the door to a study of the decomposition of risk into fundamental and endogenous components. With such formulation available, we can formulate the decomposition of the values of derivative securities, using Black Scholes, into the fundamental and endogenous components. Such a decomposition is likely to provide an explanation to the Smile Curves in derivative pricing.

4. *Destabilizing speculation of futures markets*. Could the opening of a future's market increase the volatility of a spot market? This is an old question that has not been fully clarified. Our conjecture is that a proper formulation of the problem will show that if margin requirements and leverage conditions are sufficiently relaxed in a futures market, its opening could give rise to endogenous uncertainty, which cannot arise in the spot market if storage cost are high enough. This could increase the volatility of a spot market.

5. *Volume of trade and speculation*. Markets with diverse beliefs are the natural arenas for agents to trade differences in their beliefs, and we have reviewed recent progress made by Wu and Guo on this problem. However, the problem of speculation needs to be solved for economies with risk aversion. Also, a significant amount of empirical work has been done on patterns of the volume of trade, but much remains to be explained with formal models.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz, Learning and disagreement in an uncertain world, working paper, MIT, 2007.

Allen, F., S. Morris, and H. S. Shin, Beauty contests and iterated expectations in asset markets, *Review of Financial Studies*, 2006, 19, 719–752.

Amato, J., and H. S. Shin, Imperfect common knowledge and the information value of prices, *Economic Theory*, 2006, 27, 213–241.

Angeletos, G. M., and A. Pavan, Efficient use of information and social value of information, working paper, MIT, 2006.

Angeletos, G. M., and I. Werning, Crises and prices: Information aggregation, multiplicity and volatility, *American Economic Review*, 2006, 96, 1721–1737.

Bacchetta, P., and E. van Wincoop, Higher order expectations in asset pricing, working paper, Study Center Gerzensee, 2005.

Bacchetta, P., and E. van Wincoop, Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review*, 2006, 96, 552–576.

Barberis, N., M. Huang, and T. Santos, Prospect theory and asset prices, *Quarterly Journal of Economics*, 2001, 116 1–53.

Basci, E., S. Ozyildirim, and K. Aydogan, A note on price-volume dynamics in an emerging stock market, *Journal of Banking and Finance*, 1996, 20, 389–400.

Batchelor, R., and P. Dua, Blue chip rationality tests, *Journal of Money, Credit and Banking*, 1991, 23, 692–705.

Beltratti, A., and C. Morana, Breaks and persistency: Macroeconomic causes of stock market volatility, *Journal of Econometrics*, 2006, 131, 151–177.

Black, S., The forward discount puzzle in a rational beliefs framework, working paper, Department of Economics, Stanford University, 1997.

Black, S., The role of endogenous uncertainty in foreign exchange markets, Ph.D. dissertation submitted to the Department of Economics, Stanford University, 2005.

Blanchard, O. J., and C. M. Kahn, The solution of linear difference models under rational expectations, *Econometrica*, 1980, 48, 1305–1311.

Bollerslev, T., R. Engle, and D. Nelson, ARCH models, in *Handbook of Econometrics*, Vol. IV, R. Engle, and G. McFadden (eds.), North-Holland, 1994.

Brock, W. A., and C. H. Hommes, A rational route to randomness, *Econometrica*, 1997, 65, 1059–1095.

Brock, W. A., and C. H. Hommes, Heterogeneous beliefs and routes to chaos in a simple asset-pricing model, *Journal of Economic Dynamics and Control*, 1998, 22, 1235–1274.

Brock, W. A., and B. LeBaron, A dynamic structural model for stock return volatility and trading volume, *The Review of Economics and Statistics*, 1996, 78, 94–110.

Brown, D., and R. Jennings, On technical analysis, *Review of Financial Studies*, 1989, 2, 527–551.

Brunnermeier, M. K., *Asset Pricing Under Asymmetric Information*, Oxford University Press, 2001.

Cabrales, A., and T. Hoshi, Heterogeneous beliefs, wealth accumulation, and asset price dynamics, *Journal of Economic Dynamics and Control*, 1996, 20, 1073–1100.

Camerer, C., and D. Lovallo, Overconfidence and excess entry: An experimental approach, *American Economic Review*, 1999, 89, 306–318.

Campbell, J. Y., and R. J. Shiller, Stock prices, earnings, and expected dividends, *Journal of Finance*, 1988, 43, 661–676.

Dawid, A. P., Calibration-based empirical probability, *The Journal of Statistics*, 1985, 13, 1251–1273.

Detemple, J., and S. Murthy, Intertemporal asset pricing with heterogeneous beliefs, *Journal of Economic Theory*, 1994, 62, 294–320.

Diamond, D., and R. Verrecchia, Information aggregation in a noisy rational expectations equilibrium, *Journal of Financial Economics*, 1981, 9, 221–235.

Engel, C. M., The forward discount anomaly and the risk premium: A survey of recent evidence, *Journal of Empirical Finance*, 1996, 3, 123–192.

Fama, E. F., The behavior of stock-market prices, *Journal of Business*, 1965, 38, 34–105.

Fama, E. F., and K. R. French, Permanent and temporary components of stock prices, *Journal of Political Economy*, 1988a, 96, 246–273.

Fama, E. F., and K. R. French, Dividend yields and expected stock returns, *Journal of Financial Economics*, 1988b, 22, 3–25.

Fan, M., Heterogeneous beliefs, the term structure and time-varying risk premia, *Annals of Finance,* 2006, 2, 259–285.

Frankel, J. A., and A. K. Rose, A survey of empirical research on nominal exchange rates, in *Handbook of International Economics*, Vol. III, G. M. Grossman, and K. Rogoff (eds.), Chapter 33, pp. 1689–1729, North-Holland, 1995.

Freedman, D., On the asymptotic behavior of Bayes' estimates in the discrete case I, *Annals of Mathematical Statistics,* 1963, 34, 1386–1403.

Freedman, D., On the asymptotic behavior of Bayes' estimates in the discrete case II, *Annals of Mathematical Statistics*, 1965, 36, 454–456.

Froot, K. A., and J. A. Frankel, Forward discount bias: Is it an exchange risk premium?, *Quarterly Journal of Economics*, 1989, 104, 139–161.

Garmaise, M. J., Rational beliefs and security design, *Review of Financial Studies*, 2001, 14, 1183–1214.

Grossman, S. J., An introduction to the theory of rational expectations under asymmetric information, *Review of Economic Studies*, 1981, 154, 541–559.

Grossman, S. J., and J. Stiglitz, On the impossibility of informationally efficient markets, *American Economic Review*, 1980, 70, 393–408.

Grundy, B., and M. McNichols, Trade and revelation of information through prices and direct disclosure, *Review of Financial Studies*, 1989, 2, 495–526.

Guo, W. C., and H. M. Wu, Pyramiding/depyramiding and margin trades, working paper, July 2007.

Hammond, P., Ex-ante and ex-post welfare optimality under uncertainty, *Economica*, 1981, 46, 137–147.

Harris, M., and A. Raviv, Differences of opinion make a horse race, *Review of Financial Studies*, 1993, 6, 473–506.

Harrison, M., and D. Kreps, Speculative investor behavior in a stock market with heterogeneous expectations, *Quarterly Journal of Economics*, 1978, 92, 323–336.

He, H., and J. Wang, Differential information and dynamic behavior of stock trading volume, *Review of Financial Studies*, 1995, 8, 914–972.

Hellwig, C., Public announcements, adjustment delays and the business cycle, working paper, Department of Economics, UCLA, 2002.

Hellwig, C., Heterogeneous information and welfare effects of public information disclosures, working paper, Department of Economics, UCLA, 2005.

Hellwig, C., A. Mukherji, and A. Tsyvinski, Self-fulfilling currency crises: The role of interest rates, *American Economic Review*, 2006, 96, 1769–1787.

Judd, K. L., and A. E. Bernardo, Volume and price formation in an asset trading model with asymmetric information, working paper, Stanford University, 1996.

Judd, K. L., and A. E. Bernardo, Asset market equilibrium with general tastes, returns, and informational asymmetries, *Journal of Financial Markets*, 2000, 1, 17–43.

Kaldor, N., Speculation and economic stability, *Review of Economic Studies*, 1939, 7, 1–27.

Kandel, E., and N. D. Pearson, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy*, 1995, 4, 831–872.

Karpoff, J., The relation between price changes and trading volume: A survey, *Journal of Financial and Quantitative Analysis*, 1987, 22, 109–126.

Keynes, J. M., *The General Theory of Employment, Interest and Money*, Macmillan, 1936.

Kurz, M., The Kesten-Stigum Model and the treatment of uncertainty in equilibrium theory, in *Essays on Economic Behavior Under Uncertainty*, M. S. Balch, D. L. McFadden, and S. Y. Wu (eds.), North-Holland, 1974, 389–399.

Kurz, M., On the structure and diversity of rational beliefs, *Economic Theory*, 1994, 4, 877–900. (An edited version appears as Chapter 2 of M. Kurz (ed.), 1997a.)

Kurz, M. (ed.), Symposium: Rational beliefs and endogenous uncertainty, *Economic Theory*, 1996, 8, 383–553.

Kurz, M. (ed.), *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*, Studies in Economic Theory, No. 6, Springer-Verlag, 1997a.

Kurz, M., Asset prices with rational beliefs, 1997b, in M. Kurz (ed.), Chapter 9, 211–250, 1997a.

Kurz, M., On the volatility of foreign exchange rates, 1997c, in M. Kurz (ed.), Chapter 12, 317–352, 1997a.

Kurz, M., Private information and diverse beliefs: How different?, working paper, Stanford University, *Journal of Mathematical Economics*, 2007.

Kurz, M., and A. Beltratti, The equity premium is no puzzle, in M. Kurz (ed.), Chapter 11, 283–316, 1997a.

Kurz, M., H. Jin, and M. Motolese, Determinants of stock market volatility and risk premia, *Annals of Finance*, 2005a, 1, 109–147.

Kurz, M., H. Jin, and M. Motolese, The role of expectations in economic fluctuations and the efficacy of monetary policy, *Journal of Economic Dynamics and Control*, 2005b, 29, 2017–2065.

Kurz, M., and M. Motolese, Endogenous uncertainty and market volatility, *Economic Theory*, 2001, 17, 497–544.

Kurz, M., and M. Motolese, Risk premia, diverse beliefs and time variability of risk premia, working paper, Stanford University, November 2007, available on SSRN: *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1040421*.

Kurz, M., and M, Schneider, Coordination and correlation in Markov rational belief equilibria, *Economic Theory*, 1996, 8, 489–520.

Kurz, M., and H. M. Wu, Endogenous uncertainty in a general equilibrium model with price contingent contracts, *Economic Theory*, 1996, 8, 461–488. (Appears as Chapter 2 of M. Kurz (ed.), 1997a.)

Lindahl, E., *Studies in the Theory of Money and Capital*, George Allen and Unwin Ltd, 1939.

Lucas, R. E., Expectations and the neutrality of money, *Journal of Economic Theory*, 1972, 4, 103–124.

Mehra, R., and E. C. Prescott, The equity premium: A puzzle, *Journal of Monetary Economics*, 1985, 15, 145–162.

Milgrom, P., and N. Stokey, Information, trade and common knowledge, *Journal of Economic Theory*, 1982, 26, 17–27.

Miller, E. M., Risk, uncertainty and divergence of opinions, *Journal of Finance*, 1977, 32, 1151–1168.

Morris, S., Speculative investor behavior and learning, *Quarterly Journal of Economics*, 1996, 111, 1113–1133.

Morris, S., and H. S. Shin, Social value of public information, *American Economic Review*, 2002, 92, 1521–1534.

Morris, S., and H. S. Shin, Central bank transparency and the signal value of prices, *Brookings Papers on Economic Activity*, 2005, 2, 1–66.

Motolese, M., Money non-neutrality in a rational belief equilibrium with financial assets, *Economic Theory*, 2001, 18, 97–16.

Motolese, M., Endogenous uncertainty and the non-neutrality of money, *Economic Theory*, 2003, 21, 317–345.

Myrdal, G., *Monetary Equilibrium* (1939), reprints of economic classics, Augustus M. Kelley, 1962.

Nakata, H., A model of financial markets with endogenously correlated rational beliefs, *Economic Theory*, 2007, 30, 431–452.

Nielsen, K. C., Rational belief structures and rational belief equilibria, *Economic Theory*, 1996, 8, 339–422.

Nielsen, K. C., Floating exchange rates vs. a monetary union under rational beliefs: The role of endogenous uncertainty, *Economic Theory*, 2003, 21, 293–315.

Nielsen, K. C., On rationally confident beliefs and rational overconfidence, working paper, Catholic University of Milan, 2005.

Nielsen, K. C., Optimal economic institutions under rational overconfidence, *Quaderni dell'Instituto di Politica Economica*, No. 49, Universita Cattolica del Sacro Cuore, Milano, 2006.

Nielsen, K. C., On non-stationary stable Markov processes on a continuous state space, working paper, Catholic University of Milano 2007.

Paye, B., and A. Timmerman, Instability of return prediction models, working paper, UCSD, 2003.

Pesaran, H., and A. Timmerman, Predictability of asset returns: Robustness and economic significance, *Journal of Finance*, 1995, 50, 1201–1228.

Phelps, E., Introduction: The new microeconomics in employment and inflation theory, in *Microeconomic Foundations of Employment and Inflation Theory*, Norton, 1970.

Pigou, A. C., *Employment and Equilibrium*, Macmillan and Co, 1941.

Poterba, J. M., and L. H. Summers, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics*, 1988, 22, 27–29.

Russo, J. E., and P. J. H. Schoemaker, Managing overconfidence, *Sloan Management Review*, Winter, 1992.

Saari, D., Parts, whole, and evolution. Lecture notes, University of California– Irvine, 2006.

Schwert, G. W., Why does stock market volatility change over time?, *Journal of Finance*, 1989, 64, 1115–1163.

Shiller, R. J., Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economics Review*, 1981, 71, 421–436.

Shilling, G. A., Market timing: Better than buy-and-hold strategy, *Financial Analysis Journal*, 1992, 48, 46–50.

Singleton, K., Asset prices in a time-series model with disparately informed, competitive traders, in *New Approaches to Monetary Economics, Proceedings of the Second International Symposium in Economic Theory and Econometrics*, W. Barnett, and K. Singleton (eds.), Cambridge University Press, 1987.

Starr, R., Optimal production and allocation under uncertainty, *Quarterly Journal of Economics*, 1973, 87, 81–95.

Stock, H. J., and W. M. Watson, Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 2003, 41, 788–829.

Stock, H. J., and W. M. Watson, Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics*, The American Statistical Association, 2002, 20, 147–162.

Stock, H. J., and W. M. Watson, An empirical comparison of methods for forecasting using many predictors, Department of Economics, Harvard University, January 2005.

Summers, L. H., Does the stock market rationally reflect fundamental values?, *Journal of Finance*, 1986, 41, 591–601.

Svenson, O., Are we all less risky and more skillful than our fellow drivers?, *Acta Psychologica*, 1981, 47, 143–148.

Thornton, H., *An Inquiry into the Nature and Effects of the Paper Credit of Great Britain*, 1802 edited with an introduction by F. A. v. Hayek, Augustus K. Kelley, 1962.

Tirole, J., On the possibility of speculations under rational expectations, *Econometrica*, 1982, 50, 1163–1181.

Toukan, A., Privately held or publicly owned? Evolutionary game theoretic analysis, working paper, University of California–Irvine, 2006.

Townsend, R., Market anticipations, rational expectations and Bayesian analysis, *International Economic Review*, 1978, 19, 481–494.

Townsend, R., Forecasting the forecasts of others, *Journal of Political Economy*, 1983, 91, 546–588.

Varian, H. R., Divergence of opinion in complete markets: A note, *Journal of Finance*, 1985, 40, 309–317.

Varian, H. R., Differences of opinion in financial markets, in *Financial Risk: Theory, Evidence and Implications*, Proceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis, C. C. Stone (ed.), Kluwer/Academic Publishers, 1989.

Wang, J., A model of intertemporal asset prices under asymmetric information, *Review of Economic Studies*, 1993, 60, 249–282.

Wang, J., A model of competitive stock trading volume, *Journal of Political Economy*, 1994, 102, 127–167.

Wu, H. M., and W. C. Guo, Speculative trading with rational beliefs and endogenous uncertainty, *Economic Theory*, 2003, 21, 263–292.

Wu, H. M., and W. C. Guo, Asset price volatility and trading volume with rational beliefs, *Economic Theory*, 2004, 23, 795–829.

Woodford, M., Imperfect common knowledge and the effect of monetary policy, Chapter 1 in *Knowledge, Information and Expectations in Modern Macroeconomics, in Honor of Edmund S. Phelps*, P. Aghion, R. Frydman, J. Stiglitz, M. Woodford (eds.), Princeton University Press, 2003.

CHAPTER 9

Evolutionary Finance

Igor V. Evstigneev
Economic Studies, University of Manchester

Thorsten Hens
Swiss Banking Institute, University of Zurich

Klaus Reiner Schenk-Hoppé
Leeds University Business School and School of Mathematics, University of Leeds

9.1. Introduction 509
9.1.1. Motivation and Background 509
9.1.2. Applications and Real-World Implications 511
9.1.3. Structure of Chapter 511
9.1.4. Dynamics and Evolution 512
9.1.5. Horse Races and the Kelly Rule 515
9.2. Evolutionary Models of Financial Markets 518
9.2.1. Components of the Models 519
9.2.2. Discussion of the Assumptions 521
9.2.3. Outline of the Dynamics 523
9.3. An Evolutionary Model with Short-Lived Assets 524
9.3.1. The Model 524
9.3.2. Analysis of Local Dynamics 528
9.3.3. An Example 530
9.3.4. The Generalized Kelly Rule 532
9.3.5. Global Dynamics with Adaptive Strategies 534
9.4. An Evolutionary Stock Market Model 537
9.4.1. Local Dynamics 540
9.4.2. Global Dynamics with Constant Strategies 543
9.4.3. Kelly Rule in General Equilibrium 546

Note: We thank the National Center of Competence in Research, "Financial Valuation and Risk Management," Switzerland for financial support. Klaus Reiner Schenk-Hoppé gratefully acknowledges financial support by the Finance Market Fund, Norway (project "Stability of Financial Markets: An Evolutionary Approach").

9.5. Applications 547
9.5.1. Simulation Studies 548
9.5.2. Dynamics of Strategies: Genetic Programming 552
9.5.3. Empirical Tests of Evolutionary Asset Pricing 557
9.6. Continuous-Time Evolutionary Finance 560
9.7. Conclusion 563
References 564

Abstract

Evolutionary finance studies the dynamic interaction of investment strategies in financial markets. This market interaction generates a stochastic wealth dynamic on a heterogenous population of traders through the fluctuation of asset prices and their random payoffs. Asset prices are endogenously determined through short-term market clearing. Investors' portfolio choices are characterized by investment strategies that provide a descriptive model of decision behavior. The mathematical framework of these models is given by random dynamical systems.

This chapter surveys the recent progress made by the authors in the theory and applications of evolutionary finance models. An introduction to and the motivation of the modeling approach is followed by a theoretical part that presents results on the market selection (and coexistence) of investment strategies, discusses the relation to the Kelly Rule and implications for asset-pricing theory, and introduces a continuous-time mathematical finance version. Applications are concerned with simulation studies of market dynamics, empirical estimation of asset prices and their dynamics, and evolution of investment strategies using genetic programming.

Keywords: evolutionary finance, financial markets, investment strategies, dynamic interaction, market selection, wealth dynamics, asset pricing, market stability, Kelly Rule, simulation, stochastic dynamics, random dynamical systems

9.1. INTRODUCTION

9.1.1. Motivation and Background

Evolutionary finance aims at improving our understanding of the causes and effects of the dynamic nature of financial markets through the application of Darwinian ideas. Market places for risky assets exhibit an unparalleled degree of dynamics and evolution in the behavior and interaction of its participants. The innovations in investment styles, products, and the regulatory framework appear to be limitless. All of these changes can be traced back to human endeavor, which tries to achieve intended aims; however, to a similar extent, they are caused by the adaptive, self-organizational, and endogenous dynamics of the decisions and interaction of the market participants, sometimes with unintended consequences. It is this "life on their own," which financial markets are often claimed to possess, that our evolutionary approach strives to capture. This chapter surveys the progress made during the last seven years by the authors and their collaborators in this direction of inquiry within the recently established field of evolutionary finance.

Our approach is rooted in several (quite diverse) lines of research: evolutionary economics, financial economics, economic theory, mathematical finance, and dynamical systems theory. The application of evolutionary ideas in the social sciences has a long history. It goes back at least to Malthus, who played an inspirational role for Darwin; see Hodgeson (1993) for a review of this subject. The 1950s saw a renewed interest in this approach with the publications of Alchian (1950), Penrose (1952), and others.

This area experienced tremendous developments through the interdisciplinary research conducted in the 1980s and 1990s under the auspices of the Santa Fe Institute that brought together researchers of different backgrounds—economists, mathematicians, physicists and biologists—to study evolutionary dynamics in biology, economics, and finance (Arthur et al., 1997; Farmer and Lo, 1999; LeBaron et al., 1999; Blume and Easley, 1992; Blume and Durlauf, 2005). Their research provided the main source of inspiration and motivation for our work on evolutionary finance.

Evolutionary finance has two defining characteristics: a descriptive approach to the specification of investors and a focus on the dynamics of the wealth distribution. The descriptive modeling of investors shuns any notion of utility and its maximization. The dynamics of investors' wealth is driven by the market interaction of investors and the randomness of asset payoffs. This approach lets actions speak louder than intentions and money speak louder than happiness. Financial practitioners at the cutting-edge of active investment are mainly concerned with beating a benchmark, which is rewarded a bonus, rather than in pursuing some more elusive goals. Evolutionary finance attempts to develop models that reflect this hands-on view to financial markets, where the interaction of the investors plays a major role.

Evolutionary modeling overcomes the need to use sophisticated equilibrium concepts; it dispenses with the assumption of a high degree of rationality on the part of market participants. Both of these assumptions play an important role in classic finance and financial economics despite the fact that they have attracted so much criticism from different quarters over the last century. Evolutionary models of financial markets, in contrast, rest on a very different view of the behavior of market participants and the interdependence of investment decisions and their performance and, thus, the interaction of the traders. The emphasis in this approach is on a descriptive model of investors to allow for behaviors driven by heuristic reasoning and/or behavioral biases (e.g., myopic optimization, dependence of decisions on past performance, and other forms of bounded rationality).

The choice of the equilibrium concept marks another main shift in the paradigm of how markets work: rather than assuming that all of the investors share the same opinion about the possible future contingencies (and the price of each asset in every possible state), market equilibrium is only invoked in the short term through market clearing at the current date. The advantage of this approach is twofold: computational and conceptual. Heterogeneity of investors represents the diversity of opinions and types of behavior; short-run goals shift the focus from discounted expected utility to the wealth of investors and its dynamics. A main object of study is the performance of investment styles, in particular within a specific set of strategies. Evolutionary finance opens the door to the study of this line of inquiry without invoking a notion of an equilibrium that requires the agreement of market participants on future price systems.

The investors populating our models can be viewed as heterogeneous agents pursuing particular investment goals. Agent-based models in finance are, however, typically restricted to a very narrow set of investment strategies. Usually agent types are defined through myopic mean-variance optimization—the application of technical trading rules such as chartists and fundamentalists (see Chiarella et al., 2009; Hommes and Wagener, 2009; Lux, 2009). Such an approach has the advantage of explicit demand-and-supply functions derived from standard utility maximization. As demonstrated in this chapter, however, the class of investment strategies considered might be too restrictive because it ignores better-performing investment strategies. It is our aim to maintain the largest degree of freedom in the choice of investment strategies without sacrificing the applicability of random dynamical systems as a modeling framework.

9.1.2. Applications and Real-World Implications

Our research aims to contribute to the portfolio choice of investors and to the valuation of financial assets. Both are highly relevant topics for practitioners.

The approach to portfolio choice pursued here is quite different from most of those found in the literature on financial decision making. Rather than offering investment advice for particular tastes of risk, we seek to select investment strategies through the optimality of their asymptotic performance. This performance is dependent on the investors' market interaction, which describes the price impact of their strategies. Portfolio choice therefore is informed by objective criteria. The investment recommendation derived in this fashion is closely related to the Kelly Rule. (The term "generalized Kelly Rule" will be used in this chapter to honor his original contribution.) As for any good guide to investment, practitioners have been aware of similar concepts for quite some time—though typically lacking a theoretical foundation.

The concept of value investment, which goes back at least 75 years to Graham and Dodd (1934), or stock picking according to relative dividend yield, which even made it as a book title, share the feature that portfolio choice is guided by fundamentals (dividends). The empirical results of evolutionary portfolio theory presented here lend some support to the validity of our approach. The ultimate question—whether this investment recommendation is normative and should guide individual investors' decisions—is left to be decided by those who bet their money; we believe the 30-year debate has raged long enough.

A benchmark for asset prices is obtained through the long-term outcome of market interaction—the Kelly Rule. This finding provides a framework for the valuation of financial assets. The rationale behind this valuation approach is as follows: Only if the relative prices correspond to the Kelly benchmark, the asymptotically optimal investment strategy, does not achieve excess growth. Otherwise, a Kelly investor will reap above-average returns. The economic foundation of the evolutionary finance model implies that the benchmark is only meaningful for tradeable assets. While our approach provides a prediction on the price of one asset relative to that of some other, it does not allow for assessing the "correctness" of the overall valuation of the market. The model therefore provides relative fundamental values that are of particular interest to long–short hedges such as in pairs trading.

9.1.3. Structure of Chapter

The introductory part of this chapter, Section 9.1.4, discusses the role of dynamics and evolution for evolutionary finance. Section 9.1.5 explains and demonstrates the basic elements of this approach within Kelly's famous model of horse-betting markets. Evolutionary models of financial markets are introduced in detail in Section 9.2. The theoretical analysis of this class of models is organized in two parts: Section 9.3 covers models with short-lived assets and Section 9.4 discusses those with long-lived assets. In both cases the study moves from local dynamics to (more demanding) global dynamics. Section 9.4.3 briefly discusses the role of the Kelly rule in dynamic general equilibrium models. A range of applications is presented in Section 9.5. These comprise simulation studies of the wealth dynamics and the evolution of strategies in combination with genetic programming, as well as an empirical study of evolutionary finance and its asset-pricing implications. Section 9.6 highlights recent advances in continuous-time models in evolutionary finance. Section 9.7 summaries this chapter.

Throughout, preference will be given to the heuristic derivation of results. Readers interested in the technical details will be provided with references to relevant articles.

9.1.4. Dynamics and Evolution

Our evolutionary finance approach employs a mathematical framework tailored to the description of dynamics in physical and social systems: the theory of random dynamical systems (Arnold, 1998); see Schenk-Hoppé (2001) for a survey of applications in economics. The main challenge in the quest for dynamic models of market evolution and trader interaction is the need to break away from the usage of sophisticated equilibrium concepts that are prevalent in economic theory. Standard equilibrium approaches, for instance, rule out disagreements among agents about future events (e.g., it is common to assume agreement of economic agents about the prices in each future contingency (Laffont, 1989) and render bankruptcy as the outcome of an agent's deliberate decision. A genuine dynamic and evolutionary model will not remove all surprises the future might hold. These models live from the blunders and unintended consequences of the actions of the individuals populating the model. Survival in an evolutionary struggle is a matter of life and death (though, thankfully, traders do not anticipate their demise). The application of evolutionary reasoning requires careful modeling and analysis if one wants to avoid the pitfalls of semantics as forcefully demonstrated by Friedman's argument (1953) on the price efficiency of markets which was (mistakenly—see De Long et al., 1990) attributed to the absence of supposedly loss-making irrational traders.

Our aim in advancing evolutionary finance is, in particular, to impose as few restrictions as possible on the specification of investors and their behavior while, at the same time, accommodating markets with several risky assets. Both these goals shall further be achieved in a truly dynamic model to capture the Darwinian origin of this evolutionary approach. A brief description of the defining characteristics of our evolutionary finance models follows.

Heterogeneity

Diversity in individual's investment behavior is a cornerstone of evolutionary finance. The variety of participants' strategies of the market (the ecology of the market) makes it possible to analyze the performance of specific investment styles in light of the interdependence of traders through endogenous prices. In the terminology of evolutionary biology, investment types are associated with different species. Two evolutionary forces affect the diversity in the population of investors: On the one hand, variety is reduced by the mechanism of selection; on the other hand, mutation creates novelty in behavior. Since identical behavior in financial markets entails an identical return, it is often possible to choose a representative agent for every investor type. In a finance context it is not important who does what but how much capital is behind a particular investment style. Research aiming at creating descriptive models with heterogeneous agents is, in a sense, perpendicular to the classic financial economics approach in which a single representative agent governs the relations of prices through indifference.

Strategies

Our approach builds on a model of investment behavior that is purely descriptive. This is at odds with the usual approach in economics in which theories abound to describe the behavior of investors as expressed through their decisions about holding and trading assets and/or consumption: expectations, beliefs, preferences, heuristic decision processes, and so. Investment decisions in evolutionary finance are characterized by investment strategies: budget shares allocated to the wealth invested in the available assets. As long as an investor's total funds are nonzero (e.g., if some collateral is required to borrow) and asset prices are nonzero (e.g., excluding futures), budget shares correspond to portfolio holdings if asset prices are given. In this respect investment strategies are a more primitive concept because they can be defined independent of price systems. Moreover they are easily observable, unlike preferences or behavioral biases. This modeling approach is flexible enough to capture, for instance, agent-based models, general equilibrium models (with and without incompleteness of markets), and individual's behavioral biases. Investment strategies are widely used in mathematical finance under the labels "relative portfolio" (Björk, 2004) or "trading strategies" (Pliska, 1997), and they also appear in monetary economics as "fiscal rules" (Shapley and Shubik, 1977). Their descriptive nature allows for many different interpretations of the behaviors exhibited by the investors. Investment strategies can be constant or, more generally, adapted to some information filtration. They can be governed by a process of selection and mutation using, for instance, genetic programming (Lensberg and Schenk-Hoppé, 2007).

Dynamic Interaction

The performance of strategies is interdependent through their interaction in the market. The action of one investor affects the other investors only through its impact on asset prices. Market-clearing is ensured by a pricing rule that gives investors a price impact that is proportional to their wealth. This mechanism implies that the market is shaped to a larger extent by rich rather than poor investors. Two aspects of financial markets are implicit in our model: the flow of capital between different investment strategies and the social interaction of investors. Both can be "accommodated" through the interpretation of the dynamics of investment strategies, which leaves plenty of scope to address these issues. For instance, the equilibrium of any general stochastic dynamic equilibrium model with incomplete markets can be reproduced by our evolutionary finance model. This merely requires an appropriate specification of the investment strategies. In line with Darwinian ideas we rather prefer to view investors as being "hardwired" to their strategy while the wealth tied to each investment strategy evolves through market interaction. This perspective highlights the wealth dynamics that act across investment strategies. The evolutionary finance models discussed in detail in this chapter are like laboratories populated by investment strategies.

Selection and Stability

The distribution of wealth across investment strategies exhibits stochastic dynamics. The dynamics of investors' wealth is endogenous because it is driven by random asset payoffs, the trade of assets and consumption goods, as well as by the changes that trade entails in portfolio holdings and investments. The wealth dynamics is the most prominent feature of evolutionary finance models. *Selection*, an elementary Darwinian force, acts through the wealth dynamics in a financial market. Successful investment strategies are those gathering more wealth while strategies losing wealth are rendered unsuccessful by the selection pressure. This interpretation relates to the market selection hypothesis in that the interaction in the market selects strategies through wealth dynamics. Selection is an asymptotic property of a model (i.e., an outcome that can only be observed in the long term). Whether selection occurs is a feature related to the stability of dynamical systems (e.g., in a steady state). If a market is characterized by a single strategy (in evolutionary terminology: an incumbent), stability refers to the local dynamics of the wealth distribution when some strategy with little wealth (a mutant) is introduced. The incumbent would constitute a stable market if the mutant is wiped out, because the strategy it represents loses all of its wealth. Instability of a market corresponds to the opposite situation in which the mutant gains wealth.

Evolutionary finance provides a novel approach to *asset pricing*. The stability of markets represented by particular investment strategies provides the foundation for an evolutionary asset-pricing theory. Suppose there is a model with a unique investment strategy that is stable against any mutant strategy. Then a market in which assets are priced accordingly exhibits a strong (evolutionary) stability property. Wealth dynamics provide an actual, rather than fictitious, convergence process for the investors' wealth and therefore for asset prices. In this sense evolutionary finance can provide an asset-pricing theory with sensible stability features. Empirical applications of evolutionary finance are currently at the cutting edge of research in this field—first results are presented in this chapter.

In contrast to most research related to agent-based modeling of financial markets, the pool of permissible strategies is kept as general as possible. The analytic results will impose different restrictions on the set of investment strategies but simulation studies are, by and large, free of these constraints. The combination of the wealth dynamics with the type-switching behavior of investors is straightforward. The major advantage to other approaches is that many assets and a richer market ecology can be studied.

In economics the trade-off between immediate and future consumption in intertemporal models plays a major role in an agent's saving–investment decision, while in finance the main focus is the allocation of wealth across investment opportunities. Our evolutionary finance models allow for a strict separation of the consumption and the investment decision through an exogenous (i.e., modeler's) choice of the investors' saving rates. This provides a level playing field for the competition of investors to avoid artifacts such as oversaving. Saving "too much" (i.e., a disproportional amount relative to other investors) due to holding consistently wrong beliefs about future returns is a trait (e.g., of general equilibrium models with incomplete markets; see Blume and Easley, 2006). We feel a more narrow view will benefit the study of financial investment. Rather than measuring performance by taking into account consumption amounts, which would fix this problem, our evolutionary finance approach controls consumption through an exogenous and common saving rate. This rate determines the proportion of wealth consumed during each period of time. Every investor spends the same amount per unit of wealth owned which in turn entails a level playing field.

9.1.5. Horse Races and the Kelly Rule

The main model components and concepts of our evolutionary approach to the study view to financial markets are best introduced and illustrated in a simple betting market model. These considerations can be traced back to Kelly (1956) who, among other things, studied optimal investment in pari-mutuel betting markets in which players repeatedly reinvest their wealth over an infinite time horizon in win-only bets. The ideas for this line of inquiry on optimal investment were developed by Claude Shannon, the founder of information theory (Cover, 1998).

Consider a race of $K \geq 2$ horses. The odds of the bet "horse k win" are given by $1 : \alpha_k$ (i.e., every \$1 bet on horse k pays \$$\alpha_k$ if this horse wins, and nothing otherwise). The odds correspond to the market's estimate of horse k's chances to win. In a pari-mutuel betting market without track-take, one has

$$\frac{1}{\alpha_1} + \cdots + \frac{1}{\alpha_K} = 1 \tag{9.1}$$

A risk-free payoff in this betting market is obtained by betting the fraction $1/\alpha_k$ of one's wealth on horse k—$k = 1, \ldots, K$. According to Eq. 9.1, the total expenditure is given by $w/\alpha_1 + \cdots + w/\alpha_K = w$—the bettor's wealth. If, say, horse k wins, the payoff is $\alpha_k \cdot w/\alpha_k = w$, which is equal to the invested fortune.

In a financial market setting, betting corresponds to the holding of assets. The preceding model can be rephrased as follows. There are $K \geq 2$ assets with prices $p_1, \ldots, p_K$. Each asset's payoff $A_k(s) \geq 0$ (per unit of the asset) depends on the state of the world, $s = 1, \ldots, S$, which is revealed after all asset purchases are carried out. In a betting market, assets correspond to bets on win and, therefore, $S = K$ and $A_k(s) > 0$ if and only if $s = k$. In other words, these assets are Arrow securities. The odds are given by $p_k : A_k(k)$ or, equivalently, $1 : (A_k(k)/p_k)$; this shows that $\alpha_k = A_k(k)/p_k$. The relation (9.1) holds if, in each state, the total payoff is equal to the total amount invested. Denoting by q_k, the number of asset k held, this condition means that for all k, $q_k A_k(k) = q_1 p_1 + \cdots + q_K p_K =: A > 0$. This relation implies Eq. 9.1.

Consider an infinite sequence of horse races in which, for simplicity, the outcome of each race is independent of the previous one. (Horses can have different probabilities of winning a race, though.) Denote the probability of the event that horse k wins by π_k and let $\pi = (\pi_1, \ldots, \pi_K)$. The outcome of race t is denoted by s_t, where $s_t \in \{1, \ldots, K\}$ has a probability distribution π for every $t = 1, 2, \ldots$. Consider a bettor who fixes (once and for all) the share of her wealth to be placed on each particular bet and, moreover, always invests all payoffs received in the previous race. This investment strategy can be formally described by a vector $\lambda = (\lambda_1, \ldots, \lambda_K)$ with $\lambda_k \geq 0$ and $\sum_{k=1}^{K} \lambda_k = 1$. ($\lambda$ is a vector of portfolio weights.) Starting with initial wealth $w_0 > 0$, the wealth of the bettor after race t is given by

$$w_t = \left(\alpha_{s_t} \lambda_{s_t}\right) \ldots \left(\alpha_{s_1} \lambda_{s_1}\right) w_0 \tag{9.2}$$

The average logarithmic growth rate over t periods therefore is

$$\frac{1}{t} \ln\left(\frac{w_t}{w_0}\right) = \frac{1}{t} \sum_{u=1}^{t} \ln\left(\alpha_{s_u} \lambda_{s_u}\right) \tag{9.3}$$

The strong law of large numbers implies that, as $t \to \infty$, the t-period growth rate (9.3) converges almost surely to

$$E \ln\left(\alpha_s \lambda_s\right) = \sum_{s=1}^{K} \pi_s \ln\left(\alpha_s \lambda_s\right) \tag{9.4}$$

The highest logarithmic growth rate is achieved by the vector of portfolio weights for which $E \ln(\alpha_s \lambda_s)$ is maximal. The Lagrange approach implies that $\lambda_k^* = \pi_k$ for all k.

The vector of portfolio weights $\lambda^* = \pi$ is called the *Kelly Rule*. Remarkably, this optimal betting rule does not depend on the odds of the bets. It is clear from (Eq. 9.3) that, in the case of independent outcomes, the Kelly Rule also maximizes the expected value of all *average* logarithmic growth rates. The Kelly investor's wealth will experience a strictly positive growth rate if the odds do not coincide with the probabilities of paying off (i.e., $1/\alpha_k \neq \pi_k$ for some k). In this case betting with the Kelly portfolio weights yields excess growth because this investor's wealth growth was faster than the average investor, who has a growth rate of 0. This effect does not occur only if all market's estimates are equal to the objective probabilities (i.e., if $1/\alpha_k = \pi_k$ for all k).

That the expected logarithmic growth rate is a sensible measure of success can be seen as follows. Consider two bettors with portfolio weights λ^1 and λ^2, respectively. Then the wealth of bettor 1 relative to that of bettor 2 evolves as

$$\frac{1}{t}\ln\left(\frac{w_t^1}{w_0^1}\Big/\frac{w_t^2}{w_0^2}\right) = \frac{1}{t}\sum_{u=1}^{t}\ln\left(\frac{\alpha_{s_u}\lambda_{s_u}^1}{\alpha_{s_u}\lambda_{s_u}^2}\right) = \frac{1}{t}\sum_{u=1}^{t}\ln\left(\frac{\lambda_{s_u}^1}{\lambda_{s_u}^2}\right) \xrightarrow[t\to\infty]{} E\ln\left(\frac{\lambda_s^1}{\lambda_s^2}\right) =: I_{\lambda^2}(\lambda^1) \qquad \textbf{(9.5)}$$

The term $I_{\lambda^2}(\lambda^1)$ is called the relative entropy of λ^1 with respect to λ^2. If $I_{\lambda^2}(\lambda^1) > 0$, bettor 1's wealth grows exponentially faster than that of bettor 2. In particular, one finds that $w_t^1/w_t^2 \to \infty$ (almost surely) as $t \to \infty$ (i.e., bettor 1 overtakes bettor 2).

Eq. 9.5 tells the intriguing lesson that the odds (or, equivalently asset prices and payoffs) do not matter for optimal long-term investment. Regardless of the particular odds, the advantage (in terms of the growth rate) of one investor over the other is given by $I_{\lambda^2}(\lambda^1)$. Only objective probabilities and the investors' portfolio rules matter. Moreover, whether one investment strategy is superior to some other one can be judged by a pairwise comparison. The total number of active bettors does not play any role.

Interpretation

The Kelly Rule has several remarkable properties that allow for the interpretation of the result in different contexts.

Equilibrium Asset Pricing If the odds of at least one bet do not coincide with the objective probability, there are excess returns (i.e., a strictly positive growth rate) for an investor using the Kelly Rule. That is, this investor's wealth will, in the long term, overtake that of any other investor who does not employ the rule. The equilibrium prices are those that equate the odds and the true probabilities of this event. At these "fair prices" there is no excess return. Every investor who employs the Kelly Rule has a growth rate of wealth equal to zero, and any other investor experiences a negative growth rate.

Market Selection The wealth dynamics of investors provides a mechanism for comparison of their performance in the market. An investor with a higher growth rate than a competitor is selected by the market in the sense that the relative wealth of investors with lower growth rates tends to zero. The market selects for investors employing the Kelly Rule. The analysis here is greatly facilitated because the interaction of investors through prices does not play any role: performance can be quantified solely by using the objective probabilities (see the relation in Eq. 9.5).

Betting Your Beliefs The best decision of an investor, who strives to maximize his growth rate but is not informed about the true probabilities, is to choose a portfolio rule according to his estimate (or partial knowledge) of the true probabilities (i.e., to "bet his beliefs"). An investor with a better estimate or knowledge of the vector π (in terms of the entropy) than other investors will achieve a higher growth rate and thus overtake the investor with inferior estimates. In the preceding model, Bayesian updating presents the optimal way of learning about the true probabilities.

Log-Optimum Investment The fact that the Kelly Rule maximizes the expected logarithmic growth rate, as well as *any* expected average logarithmic growth rate, can be used to characterize the rule as the one that maximizes, at any point in time, the logarithmic growth rate. We will see later that this trait is specific to win-only betting markets (i.e., a financial market consisting only of Arrow securities). A general mathematical theory on log-optimum investment has been developed by Breiman (1961), Thorp (1971), Algoet and Cover (1988), Browne (1998), and Hakansson and Ziemba (1995).

Generalizations

Consumption, as a share of wealth, is easily accommodated. Suppose bettor i reinvests the constant fraction $0 < \delta^i \leq 1$ of her wealth in every one race. Then considerations completely analogous to the previous model show that bettor 1 overtakes bettor 2 if and only if $I_{\delta^2\lambda^2}(\delta^1\lambda^1) = E \ln\left(\delta^1\lambda_s^1/\delta^2\lambda_s^2\right) > 0$. Even if λ^1 is closer to λ^* than λ^2, a too-small δ^1 (relative to δ^2) can ensure that bettor 2 overtakes bettor 1. It is clear that more economic content can be added to the specification of investors in the model (e.g., econometric learning models). The simple link between the absolute performance of an investment strategy and the consumption rate, as well the closeness to the Kelly Rule, enables a study of the market-selection hypothesis in this framework. A detailed coverage is given by Blume and Easley (2009).

Kelly's contribution to portfolio choice has stirred an amazing controversy within financial economics. The main adversary in this debate is Samuelson (1979) who questioned the value of the Kelly Rule as an investment advice on the grounds that "we should not make mean log of wealth big though years to act are long." In essence, the critique is that you should maximize your utility function rather than base your investment decision on some other criterion. This is certainly correct, but it fails to appreciate that Kelly's results are not necessarily normative but rather descriptive. This is true in particular if the issue of selection of investment strategies (in connection with the market-selection hypothesis) is discussed. In this view, the Kelly Rule, as well as the growth rates, provides a benchmark that is available to an outside observer.

9.2. EVOLUTIONARY MODELS OF FINANCIAL MARKETS

There are two main classes of models, both with proven potential. This section provides an outline of them by introducing their principal components, followed by a thorough discussion of the underlying assumptions. The section concludes with an outline of wealth and price dynamics.

The fundamental difference between these two model classes is the life span of the assets: assets either live for one period (short-lived assets) or infinitely many periods (long-lived assets). Short-lived assets are entitlements to a random payoff. They are issued at some point in time, pay out at the beginning of the subsequent period, and then become worthless (i.e., disappear and are issued again). Examples are bets at the horse track or options with a one-period maturity. A detailed discussion of this model with Arrow-type securities is provided in Section 9.5. Long-lived assets produce a random payoff stream from the day of issue that lasts until eternity. Since these assets do not expire or disappear, their (future) value is positive, and they are traded among the investors. The classic example is that of dividend-bearing stocks. In the first case an investor's income is only from asset payoffs, while for long-lived assets investors receive dividend income as well as capital gains (or losses) from price changes in the assets.

9.2.1. Components of the Models

Both types of evolutionary finance models (short- and long-lived assets) use the same components, which are explained in detail in the following subsections.

Time

All models discussed here are placed within a discrete-time framework. Time is indexed with $t = 0, 1, 2, \ldots$, with $t = 0$ being the initial time period.

Randomness

The randomness of asset payoffs is modeled through a sequence of random variables—s_t, $t = 0, 1, \ldots$—with a finite state space S. $s_t = 1, \ldots, S$ describes the "state of the world" at time t. It is convenient (and without loss of generality) to assume that there is an infinite past as well (i.e., states of the world s_t are also defined for $t = -1, -2, \ldots$). The state is either an (IID). (independent and identically distributed) process with distribution $\pi(s) = P\{s_t = s\} > 0$ or, more generally, a time-homogenous Markov process with transition probabilities $\pi(s|\hat{s}) = P\{s_{t+1} = s|s_t = \hat{s}\} \geq 0$. The state s_t should be seen as a proxy of a rather complex set of variables characterizing investors' information. At each point in time t, the vector $s^t = (\ldots, s_{t-1}, s_t)$ denotes the history of events.

Assets

There are $K \geq 1$ assets, each in unit supply. Asset k's payoff at time t is given by $A_k(s_t)$. The asset payoff is in terms of a (perishable) consumption good—just as in Lucas (1978). This assumption in particular ensures that the assets are the only store of value.

The dependence on s_t is responsible for the randomness of the payoff. Throughout the remainder of this chapter we will assume that

$$A_k(s) \geq 0 \quad \text{and} \quad \sum_{k=1}^{K} A_k(s) > 0 \tag{9.6}$$

for all k and all s. It is further convenient (and customary) to assume the absence of redundant assets.

This condition ensures that different portfolios have different payoff streams and thus a unique relationship. The functions $A_1(\cdot), \ldots, A_K(\cdot)$ restricted to the set $\{s \in S : \pi(s) > 0\}$ are linearly independent. Since $\pi(s) > 0$ for all $s = 1, \ldots, S$, there are no redundant assets if and only if matrix $(A_1(\cdot), \ldots, A_K(\cdot))$ has full rank.

Assets are called short-lived if they pay off only once and then become worthless. They are called long-lived if they produce a payoff stream that, during each period in time, has a strictly positive probability of being strictly positive.

Strategies and Investors

There are $I \geq 1$ investors who can trade in the K assets at every point in time t. Investor i's wealth at time t is denoted by w_t^i, the initial endowment being $w_0^i \geq 0$. An investor's wealth can change because of (1) receipts of asset payoffs, (2) changes in asset prices, and (3) expenditures for consumption. Each investor is characterized by an investment strategy—a time- and history-dependent vector of portfolio weights. Investor i's investment strategy is denoted by

$$\lambda_t^i = (\lambda_{1,t}^i, \ldots, \lambda_{K,t}^i), \;\; \lambda_t^i = \lambda_t^i(s^t), \quad t \geq 0, \tag{9.7}$$

with

$$\lambda_{k,t}^i > 0 \quad \text{and} \quad \sum_{k=1}^{K} \lambda_{k,t}^i = 1 \tag{9.8}$$

The value of $\lambda_{k,t}^i$ is investor i's budget share allocated to the investment in asset k (obtained either through purchases or reduction of a position). Non-negativity of budget shares means that short-selling is not permitted.

It will be assumed throughout the following that the "pool" of the I strategies only contains strategies that are different from each other. As usual in evolutionary theory, the focus is on parts of a population pursuing a particular type of behavior rather than on the individual. In a finance context this identification is straightforward. All individuals who follow the same investment strategy are considered owners of an investment fund pursuing that strategy. Each individual's wealth is equal to a fraction (the person's share of an initial contribution) of the fund's current wealth.

Budget

The budget of investor i available for the purchase of assets at time t is denoted by b_t^i. This budget depends on the investor's income and consumption. If investor i has a saving rate $0 \leq \rho^i \leq 1$, her budget is $b_t^i = \rho^i w_t^i$. The expenditure on consumption is $(1 - \rho^i)w_t^i$. It will be assumed that there is a common (constant) saving rate ρ for all investors. The endowment is in wealth.

Prices

Asset prices $p_{k,t}$ at any point in time t are determined by market-clearing. Given every investor's portfolio weights λ_t^i and the vector $b_t = (b_t^1, \ldots, b_t^I)$ of the budget of investors that is available for investment, the price of asset k is

$$p_{k,t} = \langle \lambda_{k,t}, b_t \rangle := \sum_{i=1}^{I} \lambda_{k,t}^i b_t^i \qquad \textbf{(9.9)}$$

where $\lambda_{k,t} = (\lambda_{k,t}^1, \ldots, \lambda_{k,t}^I)$. Given a common saving rate ρ, the price of asset k at time t is given by $p_{k,t} = \rho\langle \lambda_{k,t}, w_t \rangle$.

Portfolios

After transaction at prices $p_{k,t} > 0$, investor i's portfolio is given by

$$\theta_{k,t}^i = \frac{\lambda_{k,t}^i b_t^i}{\langle \lambda_{k,t}, b_t \rangle} \qquad \textbf{(9.10)}$$

(i.e., $\theta_{k,t}^i$ is equal to the budget of investor i for the purchase of asset k divided by the price of asset k). Aggregating Eq. 9.10 over investors can verify that the total demand is equal to the total supply: $\sum_i \theta_{k,t}^i = 1$. With a common saving rate ρ, $\theta_{k,t}^i = \rho\lambda_{k,t}^i w_t^i / \langle \lambda_{k,t}, \rho w_t \rangle = \lambda_{k,t}^i w_t^i / \langle \lambda_{k,t}, w_t \rangle$.

9.2.2. Discussion of the Assumptions

A few comments relating these definitions to the literature are in order.

Investment strategies are specified as non-negative budget shares. This precludes short-selling of assets. The assumption is necessary to rule out bankruptcy as well as undefined asset prices. In particular, bankruptcy (i.e., negative net worth) would be prevalent in a dynamic model in which perfect foresight is absent (see also De Giorgi, 2008). The absence of demand functions further prevents the usual mechanism that yields strictly positive asset prices. This assumption can therefore be seen as a necessary limitation when considering a behavioral model in a dynamical systems setting.

Asset prices are determined with a market-clearing mechanism that, surprisingly perhaps, does not require demand functions. Remarkably, this pricing rule simultaneously clears any number of markets. This is in stark contrast to general equilibrium models and even to most agent-based models. An economic interpretation of this market-clearing approach is that of fiscal rules as introduced by Shapley and Shubik (1977). In financial mathematics, the relation (Eq. 9.9) between prices and strategies is a consequence of the self-financing constraint on portfolios.[1] Prices are linear combinations of the investors' strategies with weights determined by the their wealth. The prices will therefore resemble rich investors' strategies rather than those of the poor ones. In the extreme case in which all investors but one have no wealth, prices will be determined by the single investor with capital. The price rule (Eq. 9.9) governs the market interaction of investors. Each investor has an impact on the price proportional to her wealth.

Trade between agents takes place as an exchange of assets and the consumption good. An investor will therefore become richer if he has above-average dividend income and capital gains or if the superior performance in one source of income outweighs inferiority in the other.

The absence of a market-clearing mechanism for the consumption good is explained by Walras's Law: All asset markets clear, investors exhaust their budgets, and thus the remaining market for the consumption good also clears. It will be convenient to use the price of the consumption good as the numeraire (and thus set it equal to one).

Asset payoffs are made in a perishable consumption good, an assumption common in financial economics (Lucas, 1978). Its main advantages in the present context are that only the assets can be used for the intertemporal transfer of wealth and that there is no growing stock of money that could inflate prices. In agent-based models with one stock and money, an increase in the money supply does not affect the return on the stock because of agents' CARA utility functions. This specification of preference ensures the independence of investors' appetite for risk from the level of wealth; see, for example, Hommes and Wagener (2009).

The careful treatment of dividends as consumption good is inspired by economics. There is a clear preference for closed models in the sense that every good is accounted for (and equations balance). Assets can be interpreted as firms endowed with an initial capital stock that is worked to produce goods. Here the produce is a generic consumption good. Each asset could be viewed as a sector of the economy, with the aggregate payoff being the economy's gross domestic product. Our analysis will focus on the case in which the relative payoffs possess some degree of stationarity.

Whether the assets are short- or long-lived has a substantial impact for the wealth dynamics in evolutionary finance models. Since the agents are boundedly rational, capital and dividend gains play different roles. The presence of capital gains (or losses) strengthens the link between market dynamics and individual investor's performance. Both models therefore display different dynamics. In general equilibrium models in which economic agents have perfect foresight (Laffont, 1989), these two cases essentially coincide. Payoffs $A_{k,t+1}(s^{t+1})$ and prices $p_{k,t+1}(s^{t+1})$ can be replaced, in

[1]See, for example, Section 6.2 in Björk (2004) and Sections 2.5 and 5.6 in Pliska (1997) as well as the discussion in Section 9.6 of this chapter.

equilibrium, by cum-dividend prices (long-lived assets) or cum-price dividends (short-lived assets): $p_{k,t+1}(s^{t+1}) + A_{k,t+1}(s^{t+1})$. Then the same allocation can be obtained after appropriate change of the agents' portfolios. What matters for the equilibrium dynamics is the span of the dividend matrix.

9.2.3. Outline of the Dynamics

A brief description of the dynamics in the two models (short- and long-lived assets) follows. The purpose of this section is to provide some intuition for this modeling approach without going into technical detail—this is reserved for later.

If assets are short-lived, investment income only consists of dividends. The wealth dynamics of investor i can be written as

$$w_{t+1}^i = \sum_{k=1}^{K} A_{k,t+1}(s_{t+1})\, \theta_{k,t}^i \tag{9.11}$$

For long-lived assets, changes in asset prices will affect the investor's wealth, in addition, through capital gains and losses. One has the dynamics

$$w_{t+1}^i = \sum_{k=1}^{K} \left(A_{k,t+1}(s_{t+1}) + p_{k,t+1}\right) \theta_{k,t}^i \tag{9.12}$$

Clearly, if assets do not have a resale value (i.e., $p_{k,t+1} = 0$), then Eq. 9.12 is identical to Eq. 9.11.

The market interaction of all the investors is via their impact on asset prices. According to Eq. 9.9, $p_{k,t} = \langle \lambda_{k,t}, b_t \rangle$, where b_t^i denotes investor i's budget. If there is a common saving rate ρ, one has $p_{k,t} = \rho\langle \lambda_{k,t}, w_t \rangle$. The price of each asset therefore depends on the wealth distribution $w_t = (w_t^1, \ldots, w_t^I)$. In fact the price represents a wealth-weighted strategy. When adjusting the portfolio, the number of shares held (relative to wealth) by an investor will depend linearly on the person's strategy. If the budget share of the strategy exceeds the price, the investor will have a higher exposure to that asset than to those assigned a smaller share.

Let us assume for the time being that both dynamics (9.11) and (9.12) are well defined. (Details are left for later.) Then, for a given set of strategies, a wealth distribution $w_t = (w_t^1, \ldots, w_t^I)$ is mapped into a new distribution of wealth across the investors $w_{t+1} = (w_{t+1}^1, \ldots, w_{t+1}^I)$ simply by drawing a state of nature s_{t+1} and applying Eq. 9.11 (respectively Eq. 9.12). The evolution of the wealth distribution is defined by a dynamical system with a random component (the state of nature) outside the control of the economic agents. In mathematical terms, each of these equations defines a random dynamical system (Arnold, 1998).

Selection, survival, and stability will all be defined in terms of this wealth dynamic. Evolutionary stability will, in addition, allow for the enlargement of the number of investors. The concept of incumbents and mutants is embedded, for instance, as follows.

With two investors with wealth (w_t^1, w_t^2), the incumbent–mutant situation corresponds to the case in which w_t^1/w_t^2 is either very large (investor 1 being the incumbent) or very small (investor 1 being the mutant). Stability refers to the convergence of the wealth distribution after a small perturbation of the wealth distribution that moves it away from a steady state.

9.3. AN EVOLUTIONARY MODEL WITH SHORT-LIVED ASSETS

This section introduces an evolutionary finance model with short-lived assets. It is a direct generalization of the Kelly model of a pari-mutuel betting market discussed in Section 9.1.5. The presentation draws on Amir et al. (2005), Evstigneev et al. (2002), and Hens and Schenk-Hoppé (2005b).

The main innovation is the introduction of incomplete markets in this framework. In contrast to the preceding, it turns out that the growth rate of an investor's wealth depends on asset prices. Since prices matter in this setting, the issue of market interaction (and the price dynamics it entails) becomes important. The distinctive property of short-lived assets is the absence of a resale value. Each asset pays off one period after its issue and then becomes worthless. This requires, at every period in time t, the (re-)issue of new assets. These assets are just like lottery tickets or, as discussed in detail earlier, bets on win. After the winners received their payoff, the tickets are worthless.

The approach is incremental, with the most simple version of the evolutionary model with short-lived assets presented first. This basic setting already provides a good intuition for evolutionary finance models without burdening the reader with too much notation and technicalities. Section 9.3.5 discusses a much more general case with adapted investment strategies and the state of the world following a Markov process. The assumptions introduced in Section 9.2.1 are assumed to hold.

9.3.1. The Model

Suppose the state of the world follows an IID process. Asset payoffs depend only on the current state of nature; that is, one unit of asset k bought at time t pays out $A_k(s_{t+1})$ with s_{t+1} being the state of nature revealed after all trade is completed in period t. Investors employ constant proportions strategies (i.e., portfolio weights are fixed once and for all, $\lambda^i = (\lambda_1^i, \ldots, \lambda_K^i)$, $i = 1, \ldots, I$). There is no consumption but investors reinvest all receipts in any one period.

The wealth in period $t + 1$ of investor i is determined by the portfolio purchased in period t and the realization of the random asset payoffs. The relation between portfolio $\theta_{k,t}^i$ and wealth w_{t+1}^i is given by

$$w_{t+1}^i = \sum_{k=1}^{K} A_k(s_{t+1})\,\theta_{k,t}^i \qquad \textbf{(9.13)}$$

Inserting the definition of the portfolio (9.10), one obtains the random dynamics of the wealth of investor i as

$$w_{t+1}^i = \sum_{k=1}^{K} A_k(s_{t+1}) \frac{\lambda_k^i w_t^i}{\langle \lambda_k, w_t \rangle} \tag{9.14}$$

These dynamics exhibit the market interaction of investors. The right side of Eq. 9.14 depends on the distribution of wealth, $w_t = (w_t^1, \ldots, w_t^I)$, across investors as well as every investor's strategy. The evolution of investors' wealth is therefore interdependent, with the dependence being caused by each investor's impact on asset prices.

The particular circumstances under which analysis for the Kelly Rule applies is apparent from Eq. 9.14. Testing for overtaking by calculating w_{t+1}^i / w_{t+1}^j yields the expression

$$\frac{w_{t+1}^i}{w_{t+1}^j} = \frac{\sum_{k=1}^{K} A_k(s_{t+1}) \lambda_k^i / p_{k,t}}{\sum_{k=1}^{K} A_k(s_{t+1}) \lambda_k^j / p_{k,t}} \, \frac{w_t^i}{w_t^j} \tag{9.15}$$

The asset prices only cancel if, in every state of the world, exactly one asset has a strictly positive payoff (and all the others have zero); in other words, if the market consists only of Arrow securities which, in particular, implies completeness of the market. In this sense, incomplete market prices matter (for relative growth and thus for survival). The study of the long-term dynamics will be more involved. Let us first transform the problem into one that is more convenient to analyze.

An investor with initial wealth $w_0^i > 0$ has, by our assumption on strictly positive budget shares, Eq. 9.8, strictly positive wealth at every point in time (i.e., $w_t^i > 0$ for all $t \geq 0$; see Eq. 9.14). Further, since the aggregate supply of each asset is equal to one, the aggregate, or total, wealth of investors is

$$W_{t+1} = \sum_{i=1}^{I} w_{t+1}^i = \sum_{k=1}^{K} A_k(s_{t+1}) \tag{9.16}$$

To this end we obtain the dynamics of relative wealth $r_t^i = w_t^i / W_t$ as

$$r_{t+1}^i = \sum_{k=1}^{K} R_k(s_{t+1}) \frac{\lambda_k^i r_t^i}{\langle \lambda_k, r_t \rangle} \tag{9.17}$$

where

$$R_k(s) = \frac{A_k(s)}{\sum_{n=1}^{K} A_n(s)}$$

The random functions $R_k(s)$, the *relative asset payoff*, inherit all properties from the original payoff process $A_k(s)$ and thus satisfy our assumptions. In addition, one has

$\sum_k R_k(s) = 1$ for all $s = 1, \ldots, S$. In matrix notation this is

$$R = \begin{pmatrix} R_1(1) & \ldots & R_K(1) \\ \vdots & \ldots & \vdots \\ R_1(S) & \ldots & R_K(S) \end{pmatrix} \quad \textbf{(9.18)}$$

The dynamics (Eq. 9.17) lives on the simplex

$$\Delta^I = \left\{ x = (x^1, \ldots, x^I) \in \mathbb{R}^I : \ x^i \geq 0, \ \sum_{i=1}^{I} x^i = 1 \right\}$$

The initial state $r_0 = (r_0^1, \ldots, r_0^I) \in \Delta_+^I$ is given by $r_0^i = w_0^i / W_0$. The vector of wealth shares $r_t = (r_t^1, \ldots, r_t^I)$ at time t depends on the entire history of states of the world; that is, $r_t = r_t(s^t)$—see Eq. 9.14. As explained before, $r_0 \in \Delta_+^I$ implies $r_t \in \Delta_+^I$ for all t and s^t. Indeed, Eq. 9.14 defines a random dynamical system (Arnold, 1998) on a simplex. Illustrations are provided in Figures 9.1 and 9.2.

Each vertex of the simplex corresponds to a state in which all but one component of the vector of wealth shares is equal to zero. An investment strategy λ^i is therefore associated with the corresponding vertex e^i of the simplex. It is straightforward from Eq. 9.17 that every vertex is a *fixed point* because $r_0^i = 0$ implies $r_t^i = 0$ for all t. In the steady state e^i only investor i has positive wealth and asset prices coincide with her portfolio shares. The fixed points of these dynamics are of central interest because they correspond to particular investment strategies. The absence of redundant assets ensures that there are no fixed points in the interior of the simplex—that is, the dynamics cannot "get stuck" (Hens and Schenk-Hoppé, 2005b, Proposition 1). The same considerations show that every face of the simplex Δ^I is invariant under the dynamics in Eq. 9.14. A face of the simplex corresponds to a situation in which certain investors have no wealth and do not impact prices. There is, however, a (nontrivial) wealth dynamic among the remaining investors.

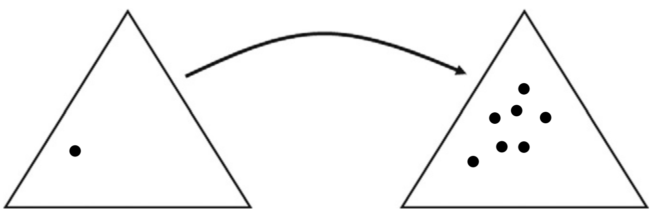

FIGURE 9.1 Graphical representation of the map defined in Eq. 9.17 for $I = 3$. The state r_{t+1} depends on the realization of the state of the world s_{t+1}. The simplex on the right shows all possible future states r_{t+1} for a given vector of wealth shares r_t; the actual state r_{t+1} observed depends on the realization of the random event s_{t+1}.

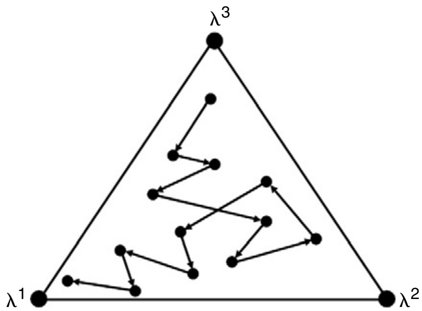

FIGURE 9.2 Random dynamics of the relative wealth r_t in Eq. 9.17 with $I = 3$. The vertices correspond to steady states in which the respective investment strategy's wealth share is equal to 1.

Selection

The criterion of overtaking, where investor i's wealth grows faster than that of investor j ($w_t^i/w_t^j \to \infty$ as $t \to \infty$), translates into the convergence of the vector of wealth shares r_t toward a face of the simplex or a vertex. For instance, if $w_t^1/w_t^j \to \infty$ for every $j \neq 1$, then $r_t \to (1, 0, \dots, 0) = e^1$. As a formal definition, one says investment strategy λ^i (represented by investor i) is *selected* if $\lim_{t\to\infty} r_t \to e^i$ almost surely, where e^i is the ith vertex (i.e., the vector with all components equal to zero except for the ith component which is one). The qualifier "almost surely" is mostly dropped in the following.

Selection and the stability of fixed points are closely linked. In essence, a fixed point is *stable* if the wealth shares converge back to the steady state after a small perturbation. This small displacement of the fixed point is used as the initial state. Selection will often happen exponentially fast. In this case stability can be detected through linearization at the fixed point.

Our analysis of the model considered in Kelly (1956), Section 9.1.5, revealed that a *pairwise* comparison of investors suffices to analyze the issue of selection. This might not be appropriate in the general case in which the pool of investors matters for the wealth dynamics of each market participant. A notion of stability of strategies (i.e., the vertex that it represents) in markets with a different number of investors being present is required. Such a market will be referred to as a "pool of strategies."

Definition of Stability

An investment strategy λ^i is called:

- *Globally stable* in a given pool of strategies if the fixed point e^i is globally stable: for every $r_0 \in \Delta$ with $r_0^i > 0$, $\lim_{t\to\infty} r_t = e^i$.
- *(Locally) stable* in a given pool of strategies if the fixed point e^i is (locally) stable: there exists a (random) neighborhood of e^i such that $\lim_{t\to\infty} r_t = e^i$ for each initial r_0 in this neighborhood.
- *Globally evolutionary stable* if λ^i is globally stable in any pool of investment strategies. In line with our assumptions, all investment strategies in this pool have to be different to λ^i. (Local) evolutionary stability is defined analogously.

All of these notions relate to the idea of mutant strategies entering the market. *Local* concepts correspond to mutants possessing little wealth initially, while *global* refers to a perturbation of the wealth distribution that is not necessarily small. Market selection, often referred to in contexts similar to the one considered here, can be interpreted as both a local or a global property of wealth dynamics. The most demanding requirement is the globally evolutionary stability of an investment strategy.

9.3.2. Analysis of Local Dynamics

A mathematical analysis of the local stability properties of investment strategies requires advanced methods from random dynamical systems theory (Hens and Schenk-Hoppé, 2005b). The local stability of a fixed point can be derived, under certain assumptions, from the linearization of a random dynamical system at this point (analogous to deterministic dynamical systems). The linearization allows one to infer (local) logarithmic growth rates, called Lyapunov exponents (or eigenvalues for deterministic systems), of the original system. If all Lyapunov exponents of the linearized system are strictly negative, local dynamics drives the state back to the fixed point. But if at least one Lyapunov exponent is strictly positive, the dynamics do not provide this pull—the fixed point is unstable.

Fortunately a heuristic derivation of the local stability analysis is available. It is presented in the following. To derive a criterion for the local stability of a constant investment strategy λ^i, suppose that r_t is close to e^i. Then the (relative) price of asset k is given by

$$q_{k,t} = \langle \lambda_k, r_t \rangle = \sum_{j=1}^{I} \lambda_k^j r_t^j \approx \lambda_k^i$$

Inserting this approximation into Eq. 9.17, one finds

$$r_{t+1}^j \approx \sum_{k=1}^{K} R_k(s_{t+1}) \frac{\lambda_k^j r_t^j}{\lambda_k^i} = \left(\sum_{k=1}^{K} R_k(s_{t+1}) \frac{\lambda_k^j}{\lambda_k^i} \right) r_t^j \tag{9.19}$$

for every $j = 1, \ldots, I$. Arranging these approximations in the form of a linear equation (with vector r_t) gives the variational equation, which is stochastic. The logarithmic growth rate of investor j's wealth share is therefore approximated by

$$\begin{aligned} \frac{1}{t} \ln \left(r_t^j / r_0^j \right) &= \frac{1}{t} \sum_{u=0}^{t-1} \ln \left(\frac{r_{u+1}^j}{r_u^j} \right) \approx \frac{1}{t} \sum_{u=0}^{t-1} \ln \left(\sum_{k=1}^{K} R_k(s_{u+1}) \frac{\lambda_k^j}{\lambda_k^i} \right) \\ &\underset{t\to\infty}{\longrightarrow} E \ln \left(\sum_{k=1}^{K} R_k(s) \frac{\lambda_k^j}{\lambda_k^i} \right) =: g_{\lambda^i}(\lambda^j) \end{aligned} \tag{9.20}$$

The growth rate has a straightforward interpretation. Any mutant competes in a market in which the prices are determined by the incumbent's strategy. From the perspective of a potential entrant to the market, the person will act in a market in which the dividend yields, which correspond to the asset returns here, are given, and not influenced by his actions. The application of this finding to the evolutionary stability of strategies is detailed next.

The analysis also led us full circle back to the pairwise comparison of investment strategies. This finding is surprising because, as already pointed out, a direct attack using the overtaking criterion fails to work. Indeed the considerations show that locally the impact of mutants on the price is negligible (a second-order effect, in economic terms). Close to a steady state—where selection happens, if it does—a one-to-one comparison of strategies suffices to gauge the dynamics of a multiple-investor setting. In mathematical terms, the dynamics is locally decoupled; that is, the growth rate $g_{\lambda^i}(\lambda^j)$ depends only on the strategies of investors i and j. This is a consequence of the fact that the matrix appearing in the variational equation (the dynamics of the linearization) is diagonal.

The information contained in Eq. 9.20 is easy to extract. If the growth rate is strictly negative, investor j's wealth share declines and eventually goes to zero. On the other hand, if the equation is strictly positive for some j, investor j's wealth share increases. From the perspective of investor i, if there is one $j \neq i$ such that Eq. 9.20 is strictly positive, then investor i's wealth share decreases (i.e., it does not converge to one). A potential shortcoming of this approach is that it only measures speed at an exponential scale. Slower convergence and divergence speeds will not be detected. For our purpose, however, this plays no role.

The following proposition summarizes our discussion (see Hens and Schenk-Hoppé, 2005b, Proposition 2).

Theorem 9.1. *Consider the growth rate*

$$g_{\lambda^i}(\lambda^j) = E \ln\left(\sum_{k=1}^{K} R_k(s)\, \frac{\lambda_k^j}{\lambda_k^i}\right) \tag{9.21}$$

The investment strategy λ^i is

(i) stable, if $g_{\lambda^i}(\lambda^j) < 0$ for all $j \neq i$;
(ii) unstable, if $g_{\lambda^i}(\lambda^j) > 0$ for some $j \neq i$.

A criterion for the local evolutionary stability of an investment strategy is a direct application of Theorem 9.1. The investment strategy λ is locally evolutionary stable if

$$E \ln\left(\sum_{k=1}^{K} R_k(s)\, \frac{\mu_k}{\lambda_k}\right) < 0 \text{ for all } \mu \neq \lambda \tag{9.22}$$

where μ is an investment strategy satisfying the assumptions in Eq. 9.8.

Stability of Markets

The preceding result provides a simple criterion to test for the stability (or instability) of a market characterized by particular asset prices. Using that, any price system can be represented by a situation in which just one investment strategy owns all of the wealth, the stability properties of the corresponding fixed point reflect that of the system of asset prices.

Coexistence of Strategies

The coexistence of strategies corresponds to a situation with the feature that, in a given pool of strategies, *all* of the investment strategies are locally unstable. Then selection fails to hold and no strategy can wipe out (or be wiped out by) its competitors. As Theorem 9.1 asserts, coexistence of investment strategies is linked to the growth rates in the neighborhood of steady states. A negative growth rate of investor i close to the steady state e^i means that prices turn against the richest investor's strategy. The person's wealth cannot grow at the prices induced by her investment strategy.

The particular role played by the price mechanism in the interaction of investors is made explicit by Theorem 9.1. Whether an investor (a mutant) can increase his wealth at the expense of the incumbent depends on how well the person's strategy performs at the prices induced by the incumbent's strategy. Interestingly perhaps, the concept of evolutionary spite does not have any "bite" here because the total payoff is independent of investors' decisions.

9.3.3. An Example

An illustration of the selection and coexistence of strategies is provided. This simple example also highlights the feature that prices "can turn against you."

Let the payoff matrix be given by

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 3 \end{pmatrix} \quad \Rightarrow \quad R = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}$$

The market is incomplete with two assets and three states. States of the world are IID with $\pi(s) = 1/3$ for $s = 1, 2, 3$. Consider two scenarios with two versus three investment strategies. The strategies are constant and given by

$$\lambda^1 = (1/2, 1/2),\ \lambda^2 = (1/4, 3/4),\ \lambda^3 = (1/3, 2/3) \tag{9.23}$$

In Scenario 1, λ^1 and λ^2 are present in the market—both endowed with equal initial wealth shares. In Scenario 2, the strategy λ^3 is added to this set of investment strategies; the initial wealth share of the new strategy is 10% while the two others share the remainder equally. A typical simulation run is depicted in Figure 9.3. The left panels

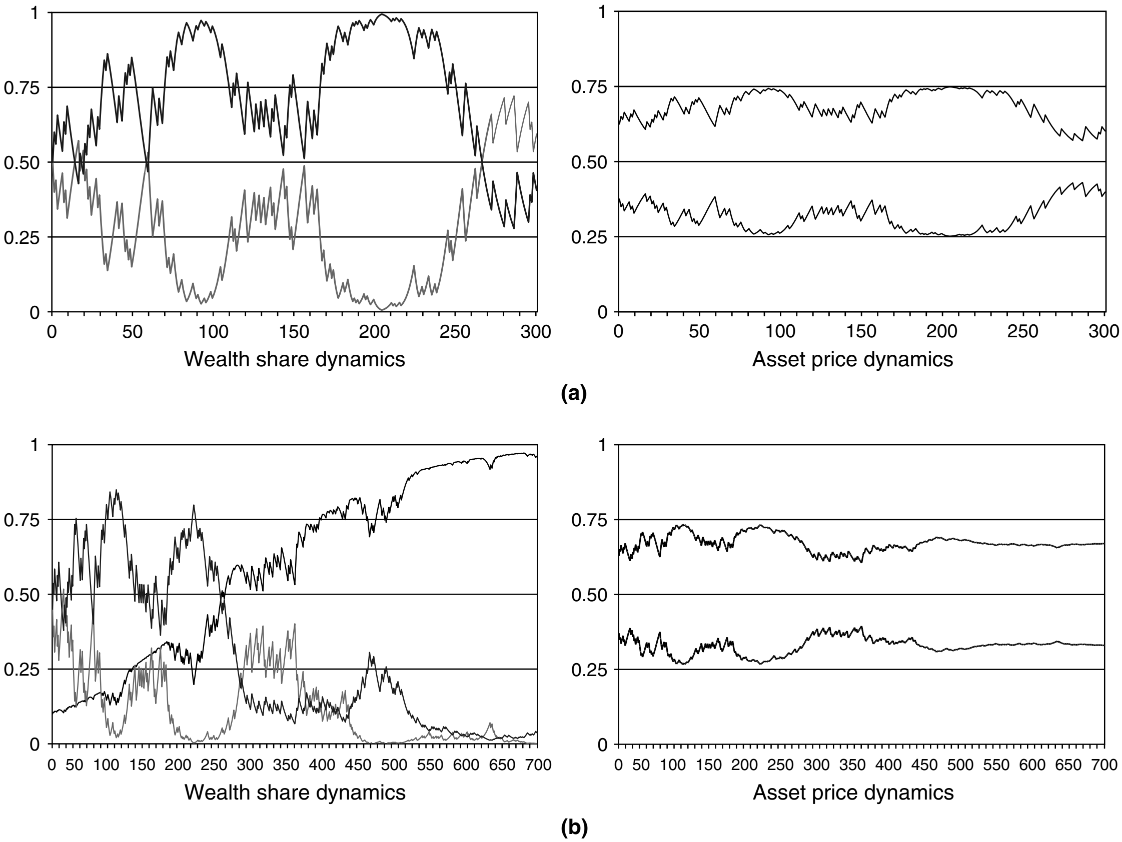

FIGURE 9.3 Dynamics of the evolutionary finance model with short-lived assets defined in Section 9.3.3. Strategies are defined in Eq. 9.23. Panel (a): Scenario 1. Two strategies (λ^1 and λ^2), time periods 0–300. Panel (b): Scenario 2. Three strategies (λ^1, λ^2 and λ^3), time periods 0–700. Both simulation runs use the same time series of states s_t, $t = 0, \ldots, 700$.

show all strategies' wealth shares and the right panels the relative prices of both assets for each case.

Coexistence of investment strategies occurs in Scenario 1 with just two strategies. The addition of the investment strategy λ^3 leads to a very different outcome: λ^3 is selected because the wealth of the two other strategies tends to zero.

The growth rates of investment strategies in Scenario 1 can be equated as

$$g_{\lambda^1}(\lambda^2) = \frac{1}{3}\left[2\ \ln\left(\frac{1}{2}\,\frac{1/4}{1/2} + \frac{1}{2}\,\frac{3/4}{1/2}\right) + \ln\left(0\,\frac{1/4}{1/2} + 1\,\frac{3/4}{1/2}\right)\right] \approx 0.13515 > 0$$

$$g_{\lambda^2}(\lambda^1) = \frac{1}{3}\left[2\ln\left(\frac{1}{2}\,\frac{1/2}{1/4} + \frac{1}{2}\,\frac{1/2}{3/4}\right) + \ln\left(0\,\frac{1/2}{1/4} + 1\,\frac{1/2}{3/4}\right)\right] \approx 0.056633 > 0$$

Both investment strategies λ^1 and λ^2 are locally unstable; selection cannot work and these two strategies coexist, as illustrated in Figure 9.3(a, left panel). The underlying cause for these dynamics can be traced to price dynamics.

If the investment strategy λ^1 owns almost all wealth, the price of asset 2 becomes too low, which is to the advantage of strategy λ^2 that places more wealth on asset 2. In the opposite situation in which λ^2 owns almost all of the wealth, asset 1 becomes too cheap. The investment strategy λ^1 benefits from this price system because it puts a higher share on that asset. This line of reasoning is confirmed by price dynamics. In the time period 150 to 225, during which strategy λ^1 is relatively poor, the price for asset 1 is lower than at any other point in time. One also observes consistent price fluctuations in Figure 9.3(a, right panel). These findings highlight the dynamic interaction of investment strategies—an interaction solely through the price system.

In the second scenario with the additional investment strategy $\lambda^3 = (1/3, 2/3)$ present, the dynamic is quite different. Since the initial wealth share of this strategy is small, the dynamic is very similar to the preceding case up to about period 250—see Figure 9.3(b, left panel). At that time period, the new strategy has gathered about half of the wealth and starts to impact prices. Over the remaining time horizon, first λ^1 quickly loses wealth to λ^3 and finally strategy λ^2 is wiped out. The dynamics of prices, Figure 9.3(b, right panel), differs from the preceding case: the fluctuations die out around time period 400 and prices converge to the values prescribed by strategy λ^3. The investment strategy λ^3 is selected by the market dynamics. Simulations with any number of constant strategies—and the strategy λ^3 being present in the pool—display the same selection outcome in every case tested. This leads to the conjecture of λ^3 being the unique locally evolutionary stable investment strategy in this example.

9.3.4. The Generalized Kelly Rule

The task of finding an analogue of the Kelly Rule in the case of short-lived assets and incomplete markets is closely related to the search for locally evolutionary stable investment strategies. Of course any such strategy is only a candidate for a *globally* evolutionary stable strategy, but it can be expected that the list of candidates will be short. The previous example provides a good motivation and further leads to the conjecture that there is indeed only one such candidate. This claim can be verified as follows.

Suppose there is some strategy λ satisfying Eq. 9.22; then

$$g_\lambda(\mu) = E \ln\left(\sum_{k=1}^{K} R_k(s)\,\frac{\lambda_k}{\mu_k}\right) \geq E \ln\left(\sum_{k=1}^{K} R_k(s)\,\frac{\mu_k}{\lambda_k}\right)^{-1} > 0$$

by the Jensen inequality because, for every s, $R_k(s)$ is a probability measure on $\{1, \ldots, K\}$. This finding implies that there is at most one candidate for a locally (and therefore globally) evolutionary stable strategy (Hens and Schenk-Hoppé, 2005b, Corollary 1).

Key to the problem of finding the locally evolutionary stable strategy is to achieve an understanding of the properties of the function $\mu \to g_\lambda(\mu)$ with $g_\lambda(\cdot)\colon \Delta^K \to \mathbb{R}$, defined in Eq. 9.20. The function $g_\lambda(\mu)$ is well defined for every $\mu \in \Delta^K$ because all components of λ are strictly positive by our assumptions in Section 9.2.1.

Obviously $g_\lambda(\cdot)$ is a concave function (on $\mathbb{R}^K$) and, by the absence of redundant assets, it is even strictly concave. Therefore for every given λ, there is a unique μ that maximizes $g_\lambda(\mu)$ on the set Δ^K. The quest for a locally evolutionary strategy is the search for a fixed point on this map. The first-order condition for a maximum, which is necessary and sufficient, is given by

$$\sum_{n=1}^{K} E \frac{R_n(s)/\lambda_n}{\sum_{k=1}^K R_k(s)\frac{\mu_k}{\lambda_k}} \alpha_n = 0$$

for every $\alpha \in \mathbb{R}^K$ with $\sum_{n=1}^K \alpha_n = 0$ (because the maximization is over the elements of a simplex). Any fixed-point λ^* of the argmax problem therefore solves

$$\sum_{n=1}^{K} E \frac{R_n(s)}{\lambda_n^*} \alpha_n = 0$$

because $\sum_{k=1}^K R_k(s) = 1$ for each $s = 1, \ldots, S$. This condition implies

$$\lambda_k^* = ER_k(s) = \sum_{s=1}^{S} \pi(s) R_k(s), \quad k = 1, \ldots, K \tag{9.24}$$

The investment strategy (9.24) is the only candidate for a locally (and, thus, globally) evolutionary stable strategy. $\lambda_k^* > 0$ by the assumptions $\pi(s) > 0$ for $s = 1, \ldots, S$ and (9.6). Our analysis shows that λ^* is locally stable against every other constant strategy. With some more work, it can be shown that this result holds true for any stationary strategy. The restriction to strategies that are stationary processes in Δ^K can be justified on grounds of the stationarity of asset payoffs. A more general class would be investment strategies that are adapted—that is, functions of the (entire) history.

Summarizing the analysis, one can state (Hens and Schenk-Hoppé, 2005b, Theorem 2) the following.

Theorem 9.2. *Suppose the state of the world s_t follows an IID process. Then the strategy λ^* defined in Eq. 9.24 is locally evolutionary stable in every pool of stationary investment strategies.*

If the state of the world follows a Markov process, considerations analogous to the preceding show that $\lambda_k^*(\hat{s}) = \sum_{s=1}^S \pi(s|\hat{s}) R_k(s)$ is the only locally evolutionary stable strategy. In the Markov case, this strategy depends on the current state of nature, and the expectation of the relative payoff $R_k(s)$ is calculated under the transition probabilities (i.e., conditional on the current event). Similar to Theorem 9.2, one needs to impose assumptions on the payoffs and transition probabilities to ensure strict positivity of $\lambda_k^*(\hat{s})$ for all $k = 1, \ldots, K$.

Interpretation

The interpretation of this result is similar to that of the Kelly Rule as supplied in Section 9.1.5. There are some notable exceptions however.

The most striking observation is that the investment strategy λ^* is given by the (conditional) expected value of the relative asset payoffs. This recipe is similar to the Kelly principle of "betting your beliefs" as detailed in Section 9.1.5. Only the (objective) probabilities and the relative payoffs are needed in the calculation of λ^*. Moreover, if the assets are Arrow securities, $R_k(s) \in \{0, 1\}$ and $R_k(s) = 1$ if and only if $k = s$. In this case $\lambda^*_{k=s} = \pi(s)$ coincides with the original Kelly Rule.

The locally evolutionary stable investment strategy λ^* derived in Theorem 9.2 yields a superior growth rate at its own prices, and it is the only strategy with this property. The result holds in complete as well as in incomplete asset markets, which is remarkable given that a simple analysis using the overtaking criteria does not apply in the latter case. In general, however, this rule will not maximize the one-period logarithmic growth rate because, away from a steady state, the composition of the market matters. The wealth distribution and the particular strategies employed by all investors impact the price and thus the log-optimum investment. For Arrow securities λ^* possesses the previously discussed optimality properties. In light of these properties, it is appropriate to call λ^* the (generalized) *Kelly Rule* for the short-lived asset market model with incomplete markets.

It might be of interest to inquire whether the Kelly Rule λ^* can be linked to utility maximization. Indeed there is a strong connection to logarithmic utility functions in a competitive equilibrium. Suppose prices are given by λ^* and an investor maximizes log utility given these prices (such as in a competitive equilibrium). Then the person's optimal strategy is λ^*. This is actually part of the reasoning in the proof of Theorem 9.2, which studies a strategy's logarithmic growth rate at given prices.

The result can also be interpreted in light of *market selection*. If a λ^* investor is present in the market, this strategy is the only one that can be selected by the market dynamics. No other investment strategy can gather all the wealth in the market. As explained before, whether selection can occur is related to the performance of a "mutant strategy" against that of an incumbent: The incumbent's strategy "sets" prices and the mutant has to play against these prices. This interaction highlights the role of the price mechanism.

9.3.5. Global Dynamics with Adaptive Strategies

The previous result leaves open two questions. The first is whether the Kelly Rule λ^* is *globally* evolutionarily stable in a pool of stationary strategies as well as for more general payoff matrices. Second, whether this demanding stability property holds true if general, adaptive strategies are permitted. This case is studied in Amir et al. (2005).

In what follows we consider a more general specification of assets, where asset k's payoff at time t is given by $A_k(s_t, s_{t-1})$. As before, the dependence on s_t is responsible for the randomness of the payoff while the entry s_{t-1}, which is observed at the time of decision making, allows for changes in the asset's payoff structure. The latter might be caused, for instance, by the issuer's exposure of the business cycle or through other macroeconomic events. Throughout the remainder of the chapter we assume that

$$A_k(s,\hat{s}) \geq 0 \quad \text{and} \quad \sum_{k=1}^{K} A_k(s,\hat{s}) > 0 \tag{9.25}$$

for all k and all $s, \hat{s}$. It is further convenient (and customary) to assume the absence of (conditionally) redundant assets. This condition ensures that different portfolios have different payoff streams and thus a unique relationship. For each $\hat{s} = 1, \ldots, S$, the functions $A_1(\cdot,\hat{s}), \ldots, A_K(\cdot,\hat{s})$ restricted to the set $\{s \in S : \pi(s\,|\,\hat{s}) > 0\}$ are linearly independent.

We will impose two additional assumptions. First, the functions

$$R_k^*(\hat{s}) := \sum_{s=1}^{S} \pi(s|\hat{s})\, R_k(s,\hat{s}) = E[R_k(s_{t+1}, s_t) \mid s_t = \hat{s}] \tag{9.26}$$

$k = 1, 2, \ldots, K$, take on strictly positive values for each $\hat{s} = 1, \ldots, S$. Eq. 9.26 is the conditional expectation of the relative payoff of every asset k given $s_t = \hat{s}$. Second, the condition in Eq. 9.8 is tightened. The coordinates $\lambda_{k,t}(s^t)$ of every investment strategy are bounded away from zero by a nonrandom constant $\gamma > 0$ (i.e., $\inf_{i,k,t,s^t} \lambda_{k,t}^i(s^t) \geq \gamma > 0$). The constant γ might depend on the strategy λ, but not on k, t, and s^t.

As for payoff functions $A(s)$, the condition on the absence of redundant assets for $A(s,\hat{s})$ implies the same property for the relative payoff functions $R(s,\hat{s})$. Therefore the assumption in Amir et al. (2005, A.2) is satisfied.

The Kelly Rule is defined as a function of the conditional expectation of the relative payoffs (Eq. 9.26):

$$\lambda_{k,t}^*(s_t) = R_k^*(s_t) \tag{9.27}$$

The dynamics of relative wealth of investment strategies is given by

$$r_{t+1}^i = \sum_{k=1}^{K} R_k(s_{t+1}, s_t) \frac{\lambda_{k,t}^i(s^t) r_t^i}{\langle \lambda_{k,t}(s^t), r_t \rangle} \tag{9.28}$$

where (see Eq. 9.17),

$$R_k(s,\hat{s}) = \frac{A_k(s,\hat{s})}{\sum_{n=1}^{K} A_n(s,\hat{s})}$$

The availability of general adaptive strategies enables investors to buy the market portfolio which, in the model (9.28), entails a payoff equal to the invested wealth. An investment strategy always buying the market portfolio will therefore possess a constant wealth share (equal to its initial fortune).

Suppose investment strategies $\lambda_t^i(s^t)$, $i = 2, \ldots, I$ are given. We can then define an adapted investment strategy for investor 1 by

$$\lambda_{k,t}^1 = \frac{1}{1 - r_t^1} \sum_{j=2}^{I} \lambda_{k,t}^j r_t^j \tag{9.29}$$

This strategy's portfolio shares are equal to the price of each asset because

$$\lambda_{k,t}^1 = \sum_{j=1}^{I} \lambda_{k,t}^j r_t^j = q_{k,t}$$

According to Eq. 9.28, the wealth dynamics of the investment strategy $\lambda_{k,t}^1$ is

$$r_{t+1}^1 = \sum_{k=1}^{K} R_k(s_{t+1}, s_t) \frac{\lambda_{k,t}^1 r_t^1}{\langle \lambda_{k,t}, r_t \rangle} = \sum_{k=1}^{K} R_k(s_{t+1}, s_t)\, r_t^1 = r_t^1 \tag{9.30}$$

Investor 1's wealth share remains constant over time (regardless of the states of the world revealed). The portfolio positions can be equated as

$$\theta_{k,t}^1 = \frac{\lambda_{k,t}^1 r_t^1}{\sum_{j=1}^{I} \lambda_{k,t}^j r_t^j} = r_t^1 \tag{9.31}$$

which means the portfolio θ_t^1 is proportional to the market portfolio (or, equivalently, the total supply) given by $(1, \ldots, 1)$.

These considerations highlight the importance of the market portfolio in this modeling framework. It also follows that market selection can only occur if none of the investment strategies (asymptotically) coincides with the market portfolio. The portfolio in this model provides protection against extinction.

A "virtual" investment strategy $\zeta_t = (\zeta_{1,t}, \ldots, \zeta_{K,t})$, which would lead to the market portfolio, can be defined through Eq. 9.29:

$$\zeta_{k,t} = \frac{1}{1 - r_t^1} \sum_{j=2}^{I} \lambda_{k,t}^j r_t^j \tag{9.32}$$

The surprising result (Amir et al., 2005, Theorem 1) is that the Kelly strategy is selected by the market dynamics if it stays asymptotically distinct from the market portfolio. In other words, if investor 1 uses the Kelly Rule, while all the others use strategies distinct from the rule and the Kelly Rule does not converge to the market portfolio, investor 1 is almost surely the *single survivor* in the market selection process.

Theorem 9.3. *Let investor 1 use the Kelly strategy $\lambda^1 = \lambda^*$ defined by Eq. 9.27. Suppose with probability 1, one has*

$$\liminf_{t\to\infty} |\lambda^*(s_t) - \zeta_t| > 0 \tag{9.33}$$

Then Kelly investor 1 is a single survivor. Moreover,

$$\liminf_{t\to\infty} \frac{1}{t} \ln \frac{r_t^1}{1 - r_t^1} > 0 \tag{9.34}$$

almost surely.

The symbol $|\cdot|$ denotes the sum of the absolute values of the coordinates of a finite-dimensional vector. The convergence property of the Kelly investor's wealth share means that it tends to one at an exponential rate, while the wealth share of the other investors vanishes at the same rate. The strategy λ^* dominates the other investors exponentially.

In the case of constant strategies, payoffs $R(s_{t+1})$ and IID states of the world, Theorem 9.3 makes the identical statement as the result (Evstigneev et al., 2002, Theorem 3.1), which says that if investor 1 uses the Kelly Rule, $\lambda^1 = \lambda^* = \sum_{s=1}^{S} \pi(s) R_k(s)$, while all the other investors $j \geq 2$ use constant strategies $\lambda^J \neq \lambda^*$. Then investor 1 is the single survivor. The proof relies on the (nontrivial) observation that

$$E \ln \sum_{k=1}^{K} R_k(s) \frac{\lambda_k^*}{\lambda_k^* r + \mu_k (1-r)} > E \ln \sum_{k=1}^{K} R_k(s) \frac{\mu_k}{\lambda_k^* r + \mu_k (1-r)}$$

for any $\mu \in \Delta^K$ with $\mu > 0$ and $\lambda^* \neq \mu$, and any $r \in [0, 1]$ (see Evstigneev et al., 2002, Lemma 3.1). This result asserts the superiority of the Kelly investor's growth rate per unit invested.

The general case covered in Theorem 9.3 rests on a similar property for the conditional expected value.

9.4. AN EVOLUTIONARY STOCK MARKET MODEL

This section introduces an evolutionary finance model of a stock market. This framework overcomes the main shortcoming of the model with short-lived assets discussed in the preceding sections. Whereas short-lived assets pay off and disappear and new assets have to be issued during each period, a stock in a company entitles its holder to a (risky) payoff stream. Stocks can experience capital gains and/or losses. The availability of such a model is of particular importance for applications to real markets. The following is based on Evstigneev et al. (2006, 2008). This model has been used to study Tobin's liquidity preference argument from an evolutionary perspective in Hens and

Schenk-Hoppé (2006). An application to insurance markets (in which liquidity shocks are present) is discussed in De Giorgi (2008).

The main difference to the previous model is that assets are issued at time zero and "live" forever. In each period of time, these long-lived assets have a market price and can be traded among all investment strategies. While short-lived assets paid in terms of wealth spent on new assets, long-lived assets require a different approach. The idea of Lucas (1978) is applied: the asset pays off in units of a perishable consumption good (whose price is also taken as the numeraire). Consumption will be modeled through a common consumption rate to provide a level playing field for the investment strategies as they separate investment and consumption decisions. In the context of stocks, these payoffs can be interpreted as dividend payments. This specification of asset payoffs implies that investment strategies with above average dividend income (relative to their wealth) will sell dividends in exchange for assets to the underperforming investment strategies. It turns out that the model with short-lived assets can be accommodated by a particular choice of the consumption rate.

The model is derived step by step. We start with a simple accounting identity linking two successive periods in time. An investment strategy's wealth in period $t+1$ is derived from this strategy's portfolio holdings $\theta^i_{k,t}$, the realized asset payoffs $A_k(s_{t+1})$, and the resale prices of assets $pk, t+1$. One has

$$w^i_{t+1} = \sum_{k=1}^{K} \left(A_k(s_{t+1}) + p_{k,t+1}\right) \theta^i_{k,t} \tag{9.35}$$

For shortness, the notation $\lambda_t = \lambda_t(s^t)$ is used in the following. Inserting Eqs. 9.9 and 9.10 into Eq. 9.35, one obtains the dynamics

$$w^i_{t+1} = \sum_{k=1}^{K} \left(A_k(s_{t+1}) + \langle \lambda_{k,t+1}, b_{t+1}\rangle\right) \frac{\lambda^i_{k,t}\, b^i_t}{\langle \lambda_{k,t}, b_t\rangle} \tag{9.36}$$

where the budget b^i_t is defined by a strategy's saving rate and current wealth.

If all the investment strategies have a common saving rate ρ, the budgets are given by $b^i_t = \rho w^i_t$. Then Eq. 9.36 takes the form

$$w^i_{t+1} = \sum_{k=1}^{K} \left(A_k(s_{t+1}) + \rho\langle \lambda_{k,t+1}, w_{t+1}\rangle\right) \frac{\lambda^i_{k,t}\, w^i_t}{\langle \lambda_{k,t}, w_t\rangle} \tag{9.37}$$

The aggregate wealth $W_{t+1} = \sum_i w^i_{t+1}$ can be equated as (summation of Eq. 9.36 over i)

$$\begin{aligned} W_{t+1} &= \sum_{k=1}^{K} \left(A_k(s_{t+1}) + \langle \lambda_{k,t+1}, b_{t+1}\rangle\right) \left[\sum_{i=1}^{I} \frac{\lambda^i_{k,t}\, b^i_t}{\langle \lambda_{k,t}, b_t\rangle}\right] \\ &= A_{t+1} + \rho \sum_{j=1}^{I} \left[\sum_{k=1}^{K} \lambda^j_{k,t+1}\right] w^j_{t+1} = A_{t+1} + \rho W_{t+1} \end{aligned} \tag{9.38}$$

with $A_{t+1} = \sum_{k=1}^{K} A_k(s_{t+1})$. One finds

$$W_{t+1} = \frac{A_{t+1}}{1-\rho} \tag{9.39}$$

This finding in particular implies that the aggregate expenditure (demand) for the consumption good is equal to the value of the aggregate supply, $(1-\rho)W_{t+1} = A_{t+1}$. This is Walras's Law: the market for each asset clears, investors exhaust their budgets, and thus the market for the consumption good clears as well. This consideration shows that the price of a consumption good is set to one; no price variable is placed in front of the payoffs A_k in Eq. 9.36.

Employing Eq. 9.39, a relation for the investor's wealth shares $r_t^i = w_t^i / W_t$ can be obtained

$$r_{t+1}^i = \sum_{k=1}^{K} \big((1-\rho)R_k(s_{t+1}) + \rho\langle \lambda_{k,t+1}, r_{t+1}\rangle\big) \frac{\lambda_{k,t}^i\, r_t^i}{\langle \lambda_{k,t}, r_t\rangle} \tag{9.40}$$

$i = 1, \ldots, I$. Recall that

$$R_k(s) = \frac{A_k(s)}{\sum_{n=1}^{K} A_n(s)}$$

The system in Eq. 9.40 is linear in the vector r_t and can be written in matrix notation. Let $\lambda_{k,t} = (\lambda_{k,t}^1, \ldots, \lambda_{k,t}^I)$ and denote by $\Lambda_t^T = (\lambda_{1,t}^T, \ldots, \lambda_{K,t}^T) \in \mathbb{R}^{I\times K}$ the matrix of investment strategies. $\Theta_t \in \mathbb{R}^{I\times K}$ is the matrix of portfolios and $R(s_t)^T = (R_1(s_t), \ldots, R_I(s_t)) \in \mathbb{R}^I$ is the vector of dividend payments in period t. Then Eq. 9.40 can be written as

$$r_{t+1} = (1-\rho)\Theta_t R(s_{t+1}) + \rho\Theta_t\Lambda_{t+1} r_{t+1} \tag{9.41}$$

This equation is equivalent to

$$r_{t+1} = (1-\rho)\,[\mathrm{Id} - \rho\Theta_t\Lambda_{t+1}]^{-1}\,\Theta_t R(s_{t+1}) \tag{9.42}$$

The last step requires the existence of the inverse of the matrix $\mathrm{Id} - \rho\Theta_t\Lambda_{t+1}$. This is ensured by the fact that the matrix is a contraction for every $0 \le \rho < 1$ (see Evstigneev et al., 2008, Proposition 1).

Remark Setting the saving rate $\rho = 0$, one obtains the evolutionary finance model with short-lived assets (see Eq. 9.28). This observation leads to a comprehensive interpretation of the components of Eq. 9.42. $\Theta_t R(s_{t+1})$ gives the investment strategies' dividend gains (i.e., income from asset payoffs), while $[\mathrm{Id} - \rho\Theta_t\Lambda_{t+1}]^{-1}$ are the capital gains (i.e., changes in book value of asset holdings due to changes in asset prices). The factor $(1-\rho)$ stems from the normalization to express wealth in terms of investors' shares of the total wealth.

An alternative and computationally efficient method to solve (Eq. 9.40) is to determine the prices $q_{t+1,k} = \langle \lambda_{t+1,k}, r_{t+1} \rangle$ first. Then these prices are inserted on the right side and the vector of wealth shares r_{t+1} can be easily calculated. Rather than deriving the inverse of a matrix with dimension $I \times I$, one only needs to invert a (typically much smaller) $K \times K$ matrix. Eq. 9.40 gives

$$q_{t+1} = (1-\rho)\,[\mathrm{Id} - \rho \Lambda_{t+1}\Theta_t]^{-1}\,\Lambda_{t+1}\Theta_t R(s_{t+1}) \qquad \textbf{(9.43)}$$

where Id is the $K \times K$-dimensional identity matrix.

The dynamics of the investment strategies' wealth shares has several features in common with those of the evolutionary finance model with short-lived assets. Every vertex of the simplex Δ^K is a fixed point, the faces are invariant, the interior of the simplex (and each face) is invariant, and there are no deterministic fixed points in the interior of Δ^K.

9.4.1. Local Dynamics

The analysis of the long-term dynamics of the evolutionary finance model with long-lived assets is similar to the case of short-lived assets considered in Section 9.3.2. There are, however, several interesting features that are unique to this model. These properties will only surface for strategies that are stationary and time-variant rather than constant. In this section it is assumed that strategies can depend on the past; that is, for each i, $\lambda^i_t = \lambda^i(s^t)$. Again we provide a heuristic analysis of an investment strategy's growth rate close to a fixed point; see Evstigneev et al. (2006) for a mathematically precise derivation.

Suppose $r_t \approx e^i$ for all t. Then the price $q_{k,t} \approx \lambda^i_k$ and the dynamics of strategy j's wealth share can be approximated (see Eq. 9.40) by

$$r^j_{t+1} \approx \left[\sum_{k=1}^{K} \frac{(1-\rho)R_k(s_{t+1}) + \rho\lambda^i_{k,t+1}}{\lambda^i_{k,t}}\,\lambda^j_{k,t}\right] r^j_t \qquad \textbf{(9.44)}$$

which implies an approximate logarithmic growth rate

$$\frac{1}{t}\ln\left(r^j_t/r^j_0\right) \underset{t\to\infty}{\longrightarrow} E\ln\left(\sum_{k=1}^{K} \frac{(1-\rho)R_k(s_1) + \rho\lambda^i_k(s^1)}{\lambda^i_k(s^0)}\,\lambda^j_k(s^0)\right) =: g_{\lambda^i}(\lambda^j) \qquad \textbf{(9.45)}$$

Suppose the state of the world follows an IID process. Then a constant strategy of incumbent i will induce an IID returns process and additionally there are no capital gains. This case can be studied completely analogous to Section 9.3.2. Define $\lambda^*_k = \sum_{s=1}^{S} \pi(s)R_k(s)$ as in Eq. 9.24. Then Eq. 9.45 gives

$$g_{\lambda^*}(\lambda^j) = E \ln\left(\rho + (1-\rho)\sum_{k=1}^{K}\frac{R_k(s_1)}{\lambda_k^*}\lambda_k^j\left(s^0\right)\right)$$
$$= \int_{S^{\mathbb{N}}}\sum_{s=1}^{S}\pi(s)\ln\left(\rho + (1-\rho)\sum_{k=1}^{K}\frac{R_k(s)}{\lambda_k^*}\lambda_k^j\left(s^0\right)\right)dP^0\left(s^0\right)$$

where P^0 denotes the probability distribution for the histories s^0. For each fixed history s^0, the inner term is strictly negative if $\lambda^j(s^0) \neq \lambda^*$ and zero if both coincide, (see Section 9.3.2). Therefore λ^* is locally evolutionary stable against all stationary investment strategies.

If the state of the world follows a Markov process with transition probability $\pi(\cdot|\cdot)$, then Eq. 9.45 can be written as

$$g_{\lambda^i}(\lambda^j) = \int_{S^{\mathbb{N}}}\tilde{g}_{\lambda^i}(\lambda^j, s^0)dP^0\left(s^0\right) \tag{9.46}$$

with

$$\tilde{g}_{\lambda^i}(\lambda^j, s^0) = \sum_{s=1}^{S}\pi(s_0|s_1)\ln\left(\sum_{k=1}^{K}\frac{(1-\rho)R_k(s_1) + \rho\lambda_k^i(s^1)}{\lambda_k^i(s^0)}\lambda_k^j\left(s^0\right)\right)$$

being the expected logarithmic growth rate of strategy λ^j at λ^i prices for a given history s^0.

Even if the incumbent's strategy is constant, returns will follow a Markov process and evolutionary stability will fail. On the other hand, if the incumbent has a Markov strategy, then the returns are Markov as well. Indeed, it turns out locally evolutionary stable strategies are Markov.

For the analysis of local stability of a stationary strategy μ in a market with λ-price system, it suffices to study the integrant $\tilde{g}_\lambda(\mu, s^0)$ in Eq. 9.45. If this term is non-negative and strictly negative on a set of histories of positive measure, $g_\lambda(\mu) < 0$. A maximum is obtained at $\mu = \lambda$ if the first-order condition holds

$$\sum_{n=1}^{K}\left(\frac{\partial\tilde{g}_\lambda(\mu, s^0)}{\partial\mu_n}\bigg|_{\mu=\lambda}\right)\alpha_n = \sum_{n=1}^{K}\sum_{s=1}^{S}\pi(s_0|s_1)\frac{(1-\rho)\,R_n(s_1) + \rho\lambda_n(s^1)}{\lambda_n(s^0)}\alpha_n = 0$$

for every $\alpha \in \mathbb{R}^K$ with $\sum_{n=1}^{K}\alpha_n = 0$. This implies that the conditional expected return of each asset must be constant; that is,

$$\sum_{s=1}^{S}\pi(s_0|s_1)\frac{(1-\rho)\,R_n(s_1) + \rho\lambda_n(s^1)}{\lambda_n(s^0)} = constant$$

It is not too difficult to see that the only investment strategy with this property is given by the function $\lambda^* : S \times K \to [0, 1]$, defined as

$$\lambda^* = \frac{1-\rho}{\rho} \sum_{t=1}^{\infty} \rho^t \pi^t R, \tag{9.47}$$

where $\pi^t = \pi \ldots \pi$ denotes the t-period transition probability with $\pi_{s\tilde{s}} = \pi(s|\tilde{s})$. The investment strategy (9.47) will be referred to as the *Kelly Rule* for reasons explained in detail later.

The local stability versus instability of an investment strategy might not be determined by the first-order condition if this strategy is stationary rather than just constant. The condition is only sufficient if the rank of the $K \times S$-dimensional matrix of returns, with elements $(1-\rho)R_k(s) + \rho\lambda_k^*(s)$, is equal to K.

That the strategy λ^* is locally stable against all stationary investment strategies (i.e., $g_{\lambda^*}(\mu) < 0$ for all μ such that $\mu(s^0) \neq \lambda^*(s^0)$ on a set of positive measure) can be seen as follows. At the prices λ^* given by Eq. 9.47, the return matrix has full rank. One has

$$\begin{aligned}(1-\rho)R + \rho\lambda^* &= (1-\rho)R + (1-\rho)\sum_{t=1}^{\infty} \rho^t \pi^t R \\ &= (1-\rho)\sum_{t=0}^{\infty} \rho^t \pi^t R = (1-\rho)[\mathrm{Id} - \rho\pi]^{-1} R\end{aligned}$$

The inverse of $\mathrm{Id} - \rho\pi$ is well defined because $[\mathrm{Id} - \rho\pi]x = 0 \iff x = \rho\pi x$ and $\rho\pi$ is a contraction. Since R has full rank by assumption, the preceding relation implies that the matrix of returns has full rank.

If the incumbent pursues a strategy different from the Kelly Rule (9.47), we can construct strategies that have a strictly positive growth rate.

Summarizing, one has from Theorem 1 from Evstigneev et al. (2006).

Theorem 9.4. *The investment strategy λ^* defined in Eq. 9.47 is the only locally stable investment strategy. That is, for each stationary strategy $\mu \neq \lambda^*$ one has (1) $g_{\lambda^*}(\mu) < 0$, and (2) there exists a stationary investment strategy λ such that $g_\lambda(\mu) > 0$.*

Interpretation

The investment strategy defined in Eq. 9.47 derives its portfolio shares from the fundamental value assets. For a given state of the world, the term on the right side of the equation is the discounted expected relative payoff of each asset. The discount factor is given by the saving rate and the expected value is calculated with respect to the conditional expectation. As in the case of short-lived assets, the relative payoff of an asset is important, not the absolute payoff.

The investment strategy (9.47) merits the term Kelly Rule because it is a natural extension of “betting your beliefs” to the framework of long-lived assets with Markov state of the world. All that is needed in the calculation of Eq. 9.47 are the transition probabilities and the asset payoffs. If the state of the world is an IID process, then $\pi^t = \pi$ and therefore the equation collapses to $\sum_s \pi(s)R_k(s)$. For Arrow securities, this investment strategy coincides with the Kelly Rule in betting markets: $\lambda_k^* = \pi(k)$ (see Section 9.1.5).

The result shows that the only locally evolutionary stable investment strategy is the Kelly Rule (9.45). A market in which a Kelly investor is the incumbent, relative asset prices are given by their fundamental value in terms of their relative payoffs. The robustness of this market against any stationary mutant strategy implies that deviations from the fundamental relative valuation are corrected over time. This finding provides a novel asset-pricing hypothesis for dividend-bearing assets such as stocks traded on security exchanges (see Section 9.5.3).

9.4.2. Global Dynamics with Constant Strategies

The global dynamics of the evolutionary finance model with long-lived assets is considerably more demanding to analyze than the short-lived asset case. At present, the wealth dynamics of a market in which a Kelly investor is present is only fully understood when all investment strategies are constant and the state of the world is governed by an IID process. The following briefly summarizes the main findings obtained in Evstigneev et al. (2008).

Define the constant investment strategy $\lambda^* = (\lambda_1^*, \ldots, \lambda_K^*)$ by

$$\lambda_k^* = ER_k(s) = \sum_{s=1}^{S} \pi(s)R_k(s) \qquad \textbf{(9.48)}$$

for $k = 1, \ldots, K$. Each budget share λ_k^* is the *expected relative dividend* of the respective asset.

To formulate the main result, a couple of definitions are required. An investment strategy $\lambda^i = (\lambda_1^i, \ldots, \lambda_K^i)$ *survives* with probability one if $\lim_{t\to\infty} r_t^i > 0$ almost surely. It *becomes extinct* with probability one if $\lim_{t\to\infty} r_t^i = 0$ almost surely. The investment strategy $\lambda = (\lambda_1, \ldots, \lambda_K)$ is called *globally evolutionary stable* if the following condition holds. Suppose investor 1 uses the strategy λ, while all the other investors $j = 2, \ldots, I$ use portfolio rules $\hat{\lambda}^j$ distinct from λ; then investor 1 survives with probability one, whereas all the other investors become extinct with probability one. Thus, we have Theorem 1 from Evstigneev et al. (2008).

Theorem 9.5. *The Kelly investment strategy λ^* defined in Eq. 9.48 is globally evolutionary stable in the pool of constant strategies.*

The strategy λ^* can be interpreted as a generalization of the Kelly Rule because, in the case of Arrow securities, the portfolio shares λ_k^* are equal to the probability of the corresponding state of the world. The presence of a price dynamic which implies the potential for capital gains and losses however highlights the quite remarkable nature of the result in Theorem 9.5. Details are given after a brief discussion of the proof.

The proof of this result relies on the observation that the Kelly investor's wealth share has a positive expected logarithmic return. This growth rate is strictly positive if and only if current prices do not coincide with the Kelly Rule. In formal notation this statement can be expressed as follows. Let r be the distribution of wealth shares across investment strategies at some period in time t. Then the asset prices in period t are given by $p_k = \langle \lambda_k, r\rangle$. The solution to

$$F^i(s,r) = \sum_{k=1}^{K} \left(\rho\langle \lambda_k, F(s,r)\rangle + (1-\rho)R_k(s)\right)\frac{\lambda_k^i r^i}{p_k}, \; i = 1,\ldots,I \tag{9.49}$$

which corresponds to Eq. 9.40, defines the asset prices in the subsequent period in time

$$q_k(s) = \langle \lambda_k, F(s,r)\rangle$$

Theorem 3 in Evstigneev et al. (2008) asserts that for each $r \in \Delta^I$ one has

$$E \ln\left(\sum_{k=1}^{K} \frac{\rho q_k(s) + (1-\rho)R_k(s)}{p_k}\lambda_k^*\right) \geq 0 \tag{9.50}$$

with strict inequality if and only if $p_k \neq \lambda_k^*$ for at least one $k = 1,\ldots,K$.

A related result is employed in the analysis of the global dynamics of constant investment strategies for short-lived assets (see the discussion toward the end of Section 9.3.5). The impact of price dynamics, which stem from the appearance of the prices $q_k(s)$ in Eq. 9.50, considerably raises the level of difficulty in studying the growth rates of investment strategies.

The case of adapted investment strategies and a Markovian dividend process is still open.

Interpretation

The preceding result has the following interpretation and reveals several interesting implications.

Theorem 9.5 states that if all investors are constrained by being required to choose constant investment strategies, there is exactly one strategy that will do best in the long term. It is the rule that divides an investor's wealth in proportions given by the expected relative dividends. The same investment strategy was discovered in the case of short-lived assets (see Section 9.3.5). In the present case however assets are long lived and there is a price dynamic. The variations in the asset prices entails capital gains (and losses) in the investor's portfolio holdings, which is absent for short-lived assets. It is therefore not obvious whether a constant investment strategy can be globally evolutionary stable. Convergence of wealth dynamics moreover implies nonstationary prices. These observations show the depths of the finding in Theorem 9.5.

Referring to λ^* as a generalization of the Kelly Rule has some justification, as explained earlier. Most of the features the rule possesses in betting markets though do not carry over to the stock market model. Indeed only two features are preserved: the form (dividend payoffs and expected value) and the property of gathering all the wealth in the long run.

This investment strategy λ^* does not match the growth-optimal portfolio in general. The former is constant while the latter would depend on the price process and thus vary over time. The important exception is the case in which asset prices are constant and equal λ^*_k. Then the investment strategy λ^* maximizes the expected logarithmic growth rate (see Eq. 9.50). This implies that all strategies different from λ^*_k will have negative growth rates.

The growth optimality of λ^* at its "own" prices has been observed as well in the local analysis (see Section 9.4.1). Indeed this observation confirms that the linearized and the actual dynamics have the same qualitative properties close to the steady state in which the λ^* investor owns all wealth.

The long-term success of strategy λ^* is rooted in another property as well. If prices are not equal to vector λ^*, the prices dynamic is not trivial because it is driven by the wealth dynamics in the pool of strategies present in the market. In these circumstances it is the (expected logarithmic) growth rate of a λ^* investor's wealth share that matters for the long-term dynamics. This property is at the heart of the proof of Theorem 9.5. Eq. 9.50 ensures that the λ^* investor's relative wealth will, on average, grow: The investor's logarithmic growth rate is strictly positive if the current asset prices do not match λ^*. A positive growth rate can be interpreted as experiencing faster growth than the "average investor." It is straightforward from Eq. 9.50 that an investment strategy equal to current prices has a growth rate equal to zero because $\sum_{k=1}^{K} \rho q_k(s) + (1-\rho)R_k(s) = 1$ for every state of nature and every price vector $q(s)$. The positive growth rate of a λ^* investor's wealth is surprising because prices do vary over time.

Theorem 9.5 is a deep result in that it shows that the price dynamics induced in a pool of constant investment strategies (and IID dividend payoffs) favors a λ^* investor for *every* distribution of wealth shares. The above-average expected growth of the λ^* investor's wealth holds in every time period and for every current price system. The asset prices in the subsequent period in time, however, are tied down by the wealth dynamics; and, due to the investment strategies being constant, the possible outcomes of the price vector are linked to the random payoffs of the assets. It is important to emphasize that these price dynamics are nonstationary since prices converge.

The mechanism behind this growth stems from the fact that a λ^* investor holds more of those assets with a price lower than their expected relative dividend and fewer of those with prices exceeding λ^*_k. Viewing λ^*_k as benchmark, these positions can be characterized as being long versus short in relative terms. The potential capital gains and/or losses caused by the other investors' strategies and wealth dynamics do not have a systematically negative effect. The positivity of the λ^* investor's growth rate means she has excess returns; that is, her logarithmic return is higher than the market average given by the prices. The asset prices eventually converge to the λ^* benchmark because the λ^* investor will gradually increase her share of the total wealth because the expected logarithmic growth rate of relative wealth is positive.

It is obvious how to leverage this result. Identifying assets that are underpriced (or overpriced) relative to the λ^* benchmark, one could construct a self-financing portfolio by going long (respectively) short in these assets. This should potentially boost the growth rate, but, on the other hand, it increases the risk. Bankruptcy, which is absent in our framework because the λ^* investor only has long positions, becomes a real risk.

Surprisingly, perhaps, this investment advice is not new. It can be traced back at least to Graham and Dodd (1934) who claimed that excess returns can be reaped from the tendency of markets to converge toward fundamental values. Our approach provides a formal model to support this claim that is derived from empirical observations.

9.4.3. Kelly Rule in General Equilibrium

This section describes the Kelly Rule as an outcome of optimal investment and consumption behavior within a dynamic general equilibrium model in which agents have perfect foresight. This result is of interest because this framework is standard in the asset-pricing literature (as well as being the foundation for most of dynamic macroeconomics). The equilibrium concept goes back to Radner (1972) who called it an equilibrium in plans, prices, and price expectations. In such an equilibrium, every current decision requires knowledge of the result of all decisions in the future. This is the exact opposite of the approach followed in our evolutionary models. In evolutionary finance only historical observations influence current behavior; no agreement about future events is required. Time moves forward—in the sense of dynamical systems—in contrast to the simultaneity of past, present, and future in general equilibrium.

Assume the state of the world follows an IID process and asset payoffs at time t are given by $A_k(s_t)$, $k = 1, \ldots, K$. The stochastic structure modeling uncertainty about future states is identical to that in evolutionary models.

The plan of agent i is given by a consumption–investment process (ρ^i, λ^i) with a saving rate process $\rho^i = (\rho^i_t)$ and an investment strategy $\lambda^i = (\lambda^i_t)$, $t \geq 0$. A price system is a process $p = (p_t)$, $t \geq 0$, with $p_t(s^t) \in \mathbb{R}^K_{++}$. All processes have to be adapted to the filtration generated by the IID state of the world. Given a price system p and a plan (ρ^i, λ^i), the wealth w^i_t of agent i evolves (see Eq. 9.12) as

$$w^i_{t+1} = \left(\sum_{k=1}^{K} \frac{A_k(s_{t+1}) + p_{k,t+1}}{p_{k,t}} \lambda^i_{k,t} \right) \rho^i_t w^i_t \tag{9.51}$$

Each agent maximizes, for a given price process p, the expected discounted logarithmic utility from consumption

$$U^i = \mathbb{E} \sum_{t=0}^{\infty} (\beta^i)^t \ln(c^i_t) \tag{9.52}$$

with consumption given by $c^i_t = (1 - \rho^i_t) w^i_t$. The discount factor is $0 < \beta^i < 1$.

An equilibrium is given by a price process p and plans (ρ^i, λ^i), $i = 1, \ldots, I$ such that (1) the plans are optimal for the price process p (i.e., maximize, Eq. 9.52) and (2) markets clear; that is,

$$p_{k,t} = \sum_i \lambda^i_{k,t}(1 - \rho^i_t)w^i_t$$

for the plans (ρ^i, λ^i), $i = 1, \ldots, I$. Thus, we have the result in Theorem 9.6.

Theorem 9.6. *The preceding dynamic general equilibrium model has a competitive equilibrium in which each agent's optimal investment strategy is given by* λ^*.

Let us give some insight for the proof of this result that is related to Gerber et al. (2007). They consider a version with more general utility functions but in a model with finite-time horizon. Rewrite Eq. 9.51 as

$$w^i_{t+1} = w^i_0 \prod_{u=0}^{t} \rho^i_u \sum_{k=1}^{K} \frac{A_k(s_{u+1}) + p_{k,u+1}}{p_{k,u}} \lambda^i_{k,u} \tag{9.53}$$

for $t \geq 1$.

The first-order conditions for the saving-rate process shows that $\rho^i_t \equiv \beta^i$. Optimality of the investment strategy is derived from the first-order condition for $\lambda^i_{n,t}$:

$$E_t \sum_{u=t+1}^{\infty} (\beta^i)^u \frac{[A_n(s_{u+1}) + p_{n,u+1}]/p_{n,u}}{\sum_{k=1}^{K}[A_k(s_{u+1}) + p_{k,u+1}]\lambda^i_{k,u}/p_{k,u}} = \xi^i_t \tag{9.54}$$

where ξ^i_t is the Lagrange multiplier corresponding to the constraint $\sum_k \lambda^i_{k,t} = 1$.

The equilibrium specification in Theorem 9.6 gives $p_{k,u} = \lambda^*_k \sum_i (1 - \beta^i)w^i_u$ and $\lambda^i_{k,u} \equiv \lambda^*_k = E[A_k(s)/\sum_n A_n(s)]$. Since all agents follow the same strategy, one further has $w^i_u = (\beta^i/\beta^1)^u (w^i_0/w^1_0) w^1_u$. With these observations, one can prove that the left side of Eq. 9.54 is independent of n, which shows optimality of the Kelly Rule in this equilibrium.

Finally, the transversality condition needs to be verified. Logarithmic utility makes this a straightforward task:

$$\lim_{t\to\infty} (\beta^i)^t \frac{\partial \ln(c^i_t)/\partial c^i_t}{\partial \ln(c^i_0)/\partial c^i_0}\, c^i_t = \lim_{t\to\infty} (\beta^i)^t c^i_0 = 0$$

The proof of Theorem 9.6 implies a certain uniqueness property of the equilibrium: If all investors pursue the same strategy λ and if this strategy is constant, then $\lambda = \lambda^*$.

9.5. APPLICATIONS

This section discusses a range of applications of evolutionary finance theory. Numeric simulations of the evolutionary finance models introduced in Sections 9.3 and 9.4 allow the study of a variety of applied issues: dynamics of asset prices, long-term asset-pricing benchmarks, performance of agent-based portfolio choice, and coexistence of investment strategies.

All of the following studies use the same set of dividend data. The asset payoffs are modeled by the (annualized) dividends paid by firms in the Dow Jones Industrial Average (DJIA) index during the 26-year period 1981 to 2006. The data are obtained from the CRSP database. Each year is associated with a particular state of the world drawn according to an IID process distributed uniformly across the 26 potential outcomes. Related studies are Hens and Schenk-Hoppé (2004) and Hens et al. (2002). The following subsections present three topics.

- A simulation analysis of the wealth dynamics of a large set of common investment strategies (and the Kelly Rule)—Section 9.5.1.
- The possibility of the evolution of the Kelly Rule, rather than being present in a market from the beginning, in a framework in which the set of strategies is not a priori fixed but their evolution is modeled by genetic programming with tournament selection—Section 9.5.2.
- An empirical test of the predictions of evolutionary finance on asset pricing and the convergence of prices, closely related to the value premium puzzle—Section 9.5.3.

9.5.1. Simulation Studies

The numerical study is based on the dividend data of all firms that have been listed in the DJIA, without interruption, during the period 1981 to 2006. There are $K = 16$ firms with this property. Denote by $D(s) \in R_+^K$ the vector of firms' total dividend payment in year $s + 1980$, $s = 1, \ldots, 26$ is the state of the world. Define the relative dividend of firm k paid in state s by

$$d_k(s) = \frac{D_k(s)}{\sum_{n=1}^{K} D_n(s)}$$

Sample paths of firms' relative dividend payments are obtained by random draws from the set $\{1, \ldots, 26\}$ using a uniformly distributed IID process. This generates samples of infinite length by "randomizing the years."

Myopic Mean–Variance Optimization

Agent-based models often assume investors who plan just one period ahead and maximize a CARA utility function. This specification is prominent in the noise trader literature (De Long et al., 1990)—see also Hommes (2001) and Hommes and Wagener (2009). The evolutionary finance framework enables an assessment of the robustness of markets (with real-world background) in which myopic mean–variance traders are present. The dynamic is described by the evolutionary finance model with long-lived assets in Section 9.4. Only two investment strategies are present in the market: (1) A mean–variance optimizer that takes into account the statistics of the dividend process as well as the prices that will prevail in the long-term; (2) an investment strategy that corresponds to an investor who is a victim of illusionary diversification and distributes his wealth equally across assets. In both cases, all investors have constant strategies. Two cases of mean–variance maximizers are considered: the global minimum-variance portfolio (Figure 9.4) and the tangency portfolio with net interest rate set to zero (Figure 9.5).

The wealth dynamics depicted in Figures 9.4 and 9.5 illustrate the surprisingly poor performance of mean–variance optimization in competition with a rather unsophisticated investment strategy. The simple-minded investor, following the illusionary diversification rule, drives out a globally minimum mean–variance investor as well as the holder of the tangent portfolio. The figures show typical runs of the disadvantaged λ^{Illu}-investor who is only endowed with 10% of the wealth. These findings highlight the importance of studying financial market dynamics outside a mean–variance framework. We consider the simulation results as a major challenge to the literature on agent-based modeling with mean–variance investors.

Performance of Adaptive Strategies

The preceding study can be placed in a much broader context by increasing the pool of competing strategies in the market. To this end we consider a range of adaptive strategies to assess performance against the λ^* investment strategy. These competing strategies are time-invariant because they process observations on prices and dividends. Precursors to

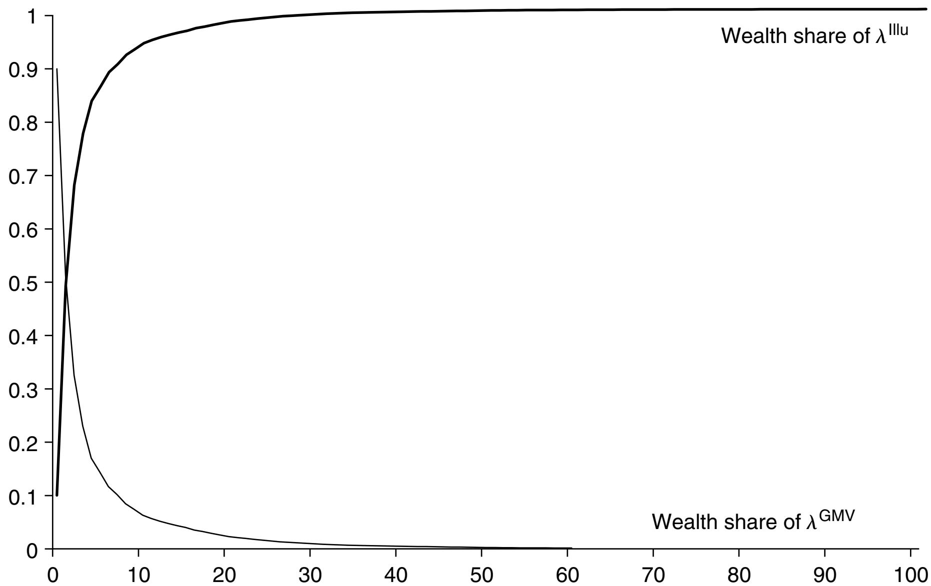

FIGURE 9.4 Dynamics of wealth shares in a market with an illusionary diversification strategy λ^{Illu} and the globally minimum mean-variance rule λ^{GMV}.

FIGURE 9.5 Dynamics of wealth shares in a market with an illusionary diversification strategy λ^{Illu} and the tangency mean-variance rule λ^{Tan}.

the numerical study presented here are Hens and Schenk-Hoppé (2004) and Hens et al. (2002).

The strategies considered in this simulation are defined as follows.

First, we have the usual suspects:

1. The Kelly strategy $\lambda_k^* = Ed_k(s) = \frac{1}{26}\sum_{s=1}^{26} d_k(s)$
2. An illusionary diversification strategy $\lambda_k^{\text{Illu}} = 1/K$
3. The weighted sample mean of the dividend payments $\lambda_{k,t}^{\text{SMean}} \sim \hat{d}_{k,t} := \sum_{\tau=1}^{t} \beta^{t-\tau} d_k(s_\tau)$ with $\beta = 0.95$
4. A strategy with behavioral bias in the sense of Kahneman and Tversky $\lambda_k^{\text{CPT}} \sim \sum_{s=1}^{26} h(d_k(s))$, where the function $h(x)$ is defined, as in Tversky and Kahneman (1992, Eq. 6, p. 309), with both parameters set to 0.65

Second, there are three "technical trading" strategies representing investors betting on the trend, or its reversal (contrarian strategy), as well as on the mean reversion of prices. The definition takes into account that short-selling is not permitted:

$$\lambda_{k,t+1}^{\text{Trend}} \sim \left[\frac{p_{k,t}}{p_{k,t-1}} - 1\right]^+, \quad \lambda_{k,t+1}^{\text{Contr}} \sim \left[1 - \frac{p_{k,t}}{p_{k,t-1}}\right]^+, \text{ and } \lambda_{k,t+1}^{\text{MRev}} \sim \left[1 - \frac{p_{k,t}}{\hat{p}_{k,t}}\right]^+$$

If any of the right side is identical to zero for all assets, the strategy is set to $(1/K, \ldots, 1/K)$. $\hat{p}_{k,t}$ denotes the sum of discounted realized prices with the discount parameter set to 0.95. Let the weighted sample mean gross return $\hat{Return}_{k,t}$ be defined analogously.

Finally, there are four adaptive investment strategies based on the solution of more-demanding optimization problems. Their initialization uses the annual returns of the 1986 to 2006 observation period. The optimization is under the constraint of no short-selling and subject to the "minimum required return" constraint

$$\sum_k (\lambda_k \hat{Return}_{k,t}) \geq 3/4 \ \max_k \hat{Return}_{k,t} \tag{9.55}$$

A mean–variance maximization investment strategy $\lambda_{t+1}^{\mu-\sigma}$ is defined as the solution to $\min_{\lambda \in \Delta^I} \lambda \hat{C}_t \lambda^{\mathrm{T}}$, where $\hat{C}_t^{k,j}$ is the weighted sample covariance. A growth-optimal investment strategy that maximizes the discounted logarithmic return is based on all realized returns. A "conditional value-at-risk" portfolio optimization as suggested by Rockafeller and Uryasev (2000) with the confidence level set to 5%. A mean–absolute deviation investment strategy as proposed by Konno and Yamuzaki (1991).

The simulation result presented in Figure 9.6 gives a clear message. The constant investment strategy λ^* prevails in the dynamics of wealth shares. The closest competitors are the adaptive strategy λ^{SMean}—based on past dividend payments—and the behavioral investment strategy λ^{CPT}—a "distorted" version of λ^*. The poor performance of the chartist strategies, as well as the quite sophisticated dynamic strategies, is

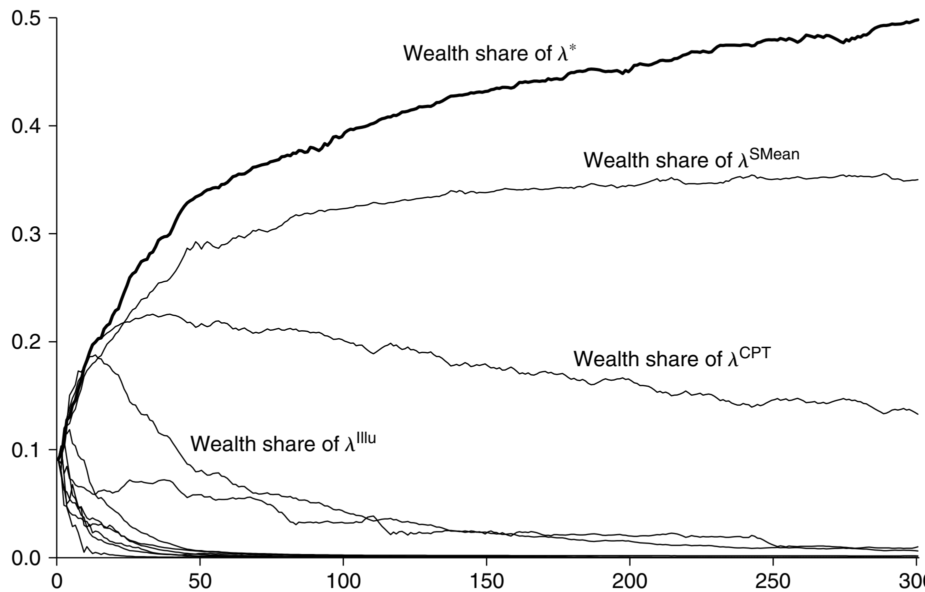

FIGURE 9.6 Typical realization of a sample path of the relative wealth of competing investment strategies. All strategies are endowed with the same wealth at time zero.

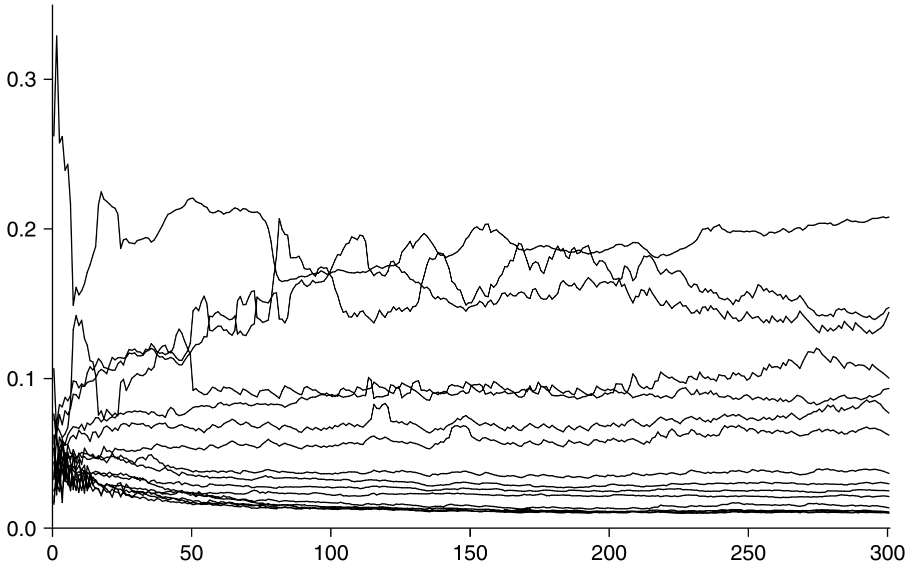

FIGURE 9.7 Price dynamics corresponding to Figure 9.6.

surprising. Another unexpected result is the excellent performance of the illusionary diversification strategy λ^{Illu}. The convergence is considerably slower than in the two-investor case studied earlier.

The (relative) asset price dynamics corresponding to the sample path of the dividend payments underlying Figure 9.6 is depicted in Figure 9.7. Asset prices converge but are more volatile than one might expect because the sample paths of the wealth dynamics are quite smooth. This observation is explained by the time variation of the adaptive strategy λ^{SMean}, which discounts dividend payments rather than just calculates the sample mean of the relative dividend—an unbiased estimator of λ^*.

These findings highlight the need for more simulation studies within this class of dynamic models. Despite extensive numerical work on agent-based models, we see this line of inquiry as a promising area for future research.

9.5.2. Dynamics of Strategies: Genetic Programming

The dynamics of investment strategies' wealth shares is the main focus in the preceding study. The strategies themselves played a rather static role that is at odds with, for instance, agent-based models. It is important to recall that strategies in the evolutionary finance models face no restrictions beside the absence of short selling and adaptiveness. Stationarity of strategies is useful in the local stability analysis because mathematical apparatus is tailored to this framework. The global convergence result for the model with short-lived assets, however, demonstrates that adapted strategies are a class of investment strategies for which wealth dynamics can be fully understood.

This leaves open the behavior of models in which the strategies and the dynamics of wealth shares coevolve. The study of this issue requires an explicit specification of the adaption and innovation of investment behavior and the entry of new strategies. Lensberg and Schenk-Hoppé (2007) pursue a Darwinian approach to the study of the evolution of investment strategies in the model with short-lived assets (which has the advantage of a relatively simple computation of the Kelly Rule). It adds the other main evolutionary process—reproduction—to the selection mechanism. The framework is that of genetic programming, which offers flexibility as well as full control on the data available to investment strategies. The latter is extremely useful in the interpretation of results.

In a genetic programming approach, the center stage is occupied by the population that embodies the investment skills of many individual strategies. The investors are simple-minded and unsophisticated in the sense that they follow preprogrammed behavior rules that are the result of mutations and crossovers. While the change in investment behavior is covered by the standard evolutionary finance model, the inflow of new investors requires an extension that is however straightforward.

Two questions are of particular interest. First, is the Kelly Rule—as the long-term equilibrium prediction of asset price—valid in a model that imposes much weaker assumptions on the market dynamics? The process of mutation generates a constant inflow of new traders, which generates a considerable amount of "noise" not present in the theoretical studies. Second, will the Kelly Rule emerge in the population of traders without strong assumptions on individuals' rationality or learning behavior and despite its absence in the initial population of investors? For instance, Bayesian learning is not an option available to the investors in this genetic programming approach.

Brief Description of Model

Lensberg and Schenk-Hoppé (2007) analyze four cases: complete/incomplete market and IID/Markov states of the world. In each case two different information scenarios are considered as specified next. The total number of investment strategies is limited to 2000. Each strategy is represented by a computer program that outputs numbers $\tilde{\lambda} : \mathcal{S} \times \{1, \ldots, K\} \to \mathbb{R}$ where the set of inputs $\mathcal{S}$ either contains the information on the current state ($\mathcal{S} = \{1, \ldots, S\}$) or, additionally, the last observed price corresponding to this state ($\mathcal{S} = \{1, \ldots, S\} \times \mathbb{R}^K$). The output is transformed to ensure that budget shares meet the conditions in Section 9.2.1. A computer program consists of up to 128 lines of instructions; see, for example, Lensberg and Schenk-Hoppé (2007, Table I).

The evolution of the strategies is driven by a tournament-selection process. In any one period in time, the following procedure is applied 20 times. Four randomly chosen programs are ranked according to their wealth (tournament). Then the two poorest programs are replaced by the two richest in this sample (reproduction). These two clones have, with some probability, a randomly selected instruction replaced by some random code (mutation). Finally, again with some probability, a randomly selected set of instructions is swapped between (crossover). Completely new behavior is introduced by adding a random draw that decides whether the programs are filled with random instructions (noise). All the investors in a tournament retain their wealth, except if it is zero, entitling the respective investor to an endowment of 1% of the total wealth. The simulations reveal a substantial number of investment strategies without wealth.

Consider a market consisting of Arrow-type securities in which the state of the world is Markovian and programs only have access to the current state of the world (Lensberg and Schenk-Hoppé, 2007, Section 3.2):

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad (\pi(s|\hat{s}))_{s,\hat{s}=1,2,3} = \begin{pmatrix} .7 & .2 & .1 \\ .1 & .7 & .2 \\ .2 & .1 & .7 \end{pmatrix}$$

If the only information given to the programs representing the investment strategies consists of the current state of the world, strategies' behavior is mainly a "bet their beliefs" style. For price-dependent strategies a more realistic scenario is obtained when providing programs with additional information about the last observed price system corresponding to the current state of the world. This in particular enables the purchase of an approximate market portfolio. The latter is achieved simply by outputting the prices as budget shares.

Results

The results for these two different specifications of the information set are summarized in Figure 9.8. The two top panels—(a) and (b)—illustrate the convergence to the Kelly Rule of both market and wealthiest investment strategy. The two middle panels—(c) and (d)—give the times series of asset prices. The bottom panels—(e) and (f)—provide some insight into the distribution of "investment skills" within the population. The graphs show how close investment strategies are to the Kelly Rule in the population of traders. g^* is the exponential of the expected logarithmic growth rate at λ^* prices.

For state-dependent strategies the distinctive features are almost monotone convergence, leapfrogging of the distance, and the observation that the market leads relative to the wealthiest investment strategy—determined in each period. The population quickly moves toward investment strategies that are quite close to the Kelly Rule. The emergence of nearly perfect matches however takes comparatively long. The effect of the noise (i.e., the continuing introduction of randomly generated strategies) is clearly documented by the persistence of a large number of poorly performing strategies (Figure 9.8(e)).

When strategies have access to the last-observed price system, which corresponds to the current state of the world, convergence still occurs but the pattern exhibits much more volatility. The long-term outcome is again the Kelly Rule for both wealthiest investment strategy and market prices. An analysis of wealthy investment strategies reveals a new type of behavior: almost all successful strategies use the (proxy) market portfolio. The fraction of wealth invested in the market portfolio, as well as the closeness of the budget shares to the market portfolio, vary with the difference between

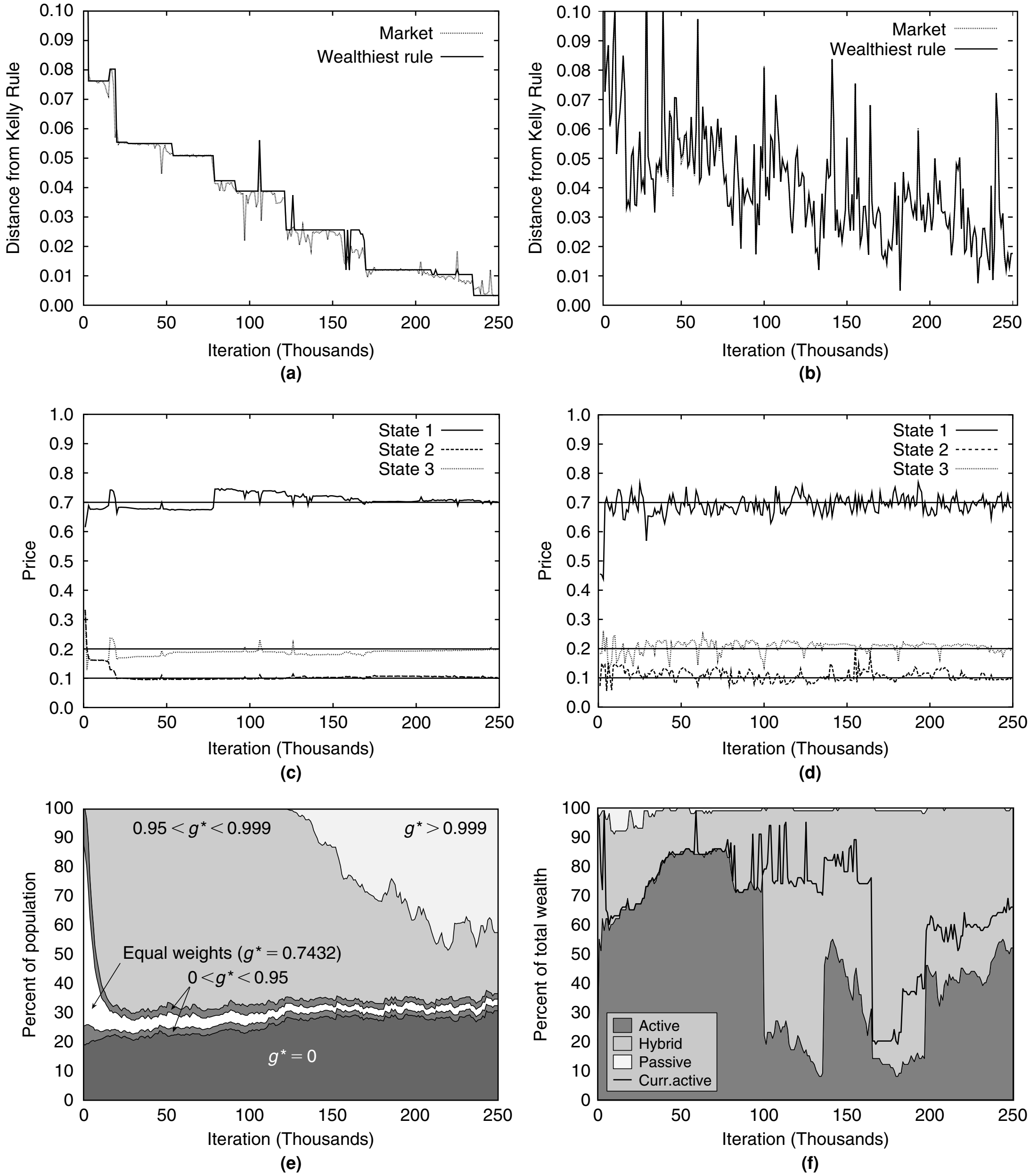

FIGURE 9.8 State-dependent strategies (left graphs) and price-dependent strategies (right graphs): (a) Distance between market prices, respectively, wealthiest strategy and Kelly Rule; (b) distance between market prices, respectively, wealthiest strategy and Kelly Rule (nearly identical); (c) relative price of Asset 1 for all three states (horizontal lines indicate the Kelly benchmark); (d) relative price of Asset 1 for all three states (horizontal lines indicate the Kelly benchmark); (e) distribution of growth rates in the population relative to the Kelly Rule; and (f) distribution of growth rates in the population relative to the Kelly Rule.

market prices and the Kelly Rule as well as specific prices. Figure 9.8(d) illustrates the composition of the pool of investment strategies. Strategies can be classified by their deviation from the market portfolio: active (always deviate), hybrid (deviate only for some asset prices), and passive (never deviate). For a particular run, the number of hybrid strategies currently deviating from the market portfolio varies.

Anatomy of Successful Strategies

Let us "dissect" the wealthiest strategy at the end of the simulation period, λ^{LW}. The behavior of this strategy in state $s = 3$ is summarized in Figure 9.9. The triangle is the set of all possible price vectors (q_1, q_2, q_3). The darker an area, the larger the distance of the budget shares to the market portfolio. White areas correspond to prices at which the strategy's portfolio is identical to the market portfolio.

The strategy λ^{LW} is hybrid and has a trigger that switches from active to passive investment. If the price of asset 1 is low, the market portfolio is played. This also happens if the price of asset 2 coincides with the Kelly Rule. For all other prices, the strategy deviates from the market portfolio. Its functional form in the active mode is given by

$$\lambda^{LW}(q) = (1 - \lambda_2^* - \epsilon(q_2)) \left(\frac{q_1}{q_1 + q_3}, \frac{\lambda_2^* + \epsilon(q_2)}{1 - \lambda_2^* - \epsilon(q_2)}, \frac{q_3}{q_1 + q_3} \right)$$

with a convex function $\epsilon(p_2)$, with values $\epsilon(0) \approx 0.008$, $\epsilon(\lambda_2^*) \approx 0.0$ and $\epsilon(1) \approx -0.016$. Strategy λ^{LW} makes bets on a reversal of the price of asset 2 to the Kelly Rule. The

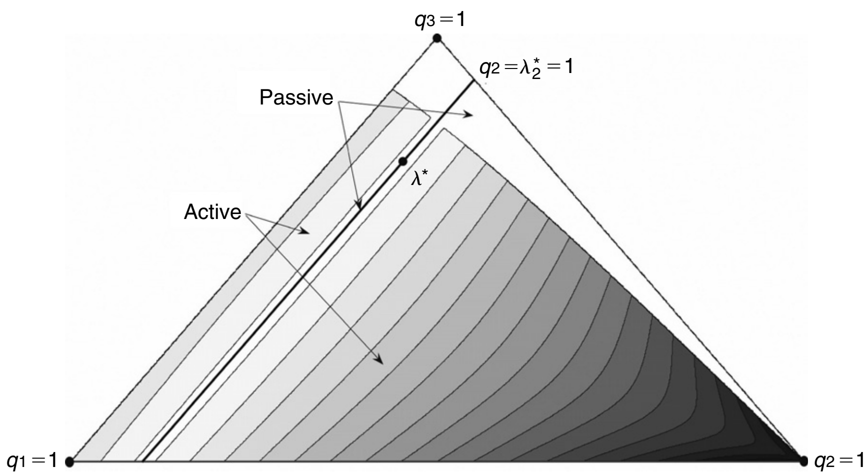

FIGURE 9.9 Price-dependent strategies. Contour plot of $\delta(q) := \|\lambda^{LW}(q) - q\|$ for relative prices $q = (q_1, q_2, q_3)$ in the unit simplex. $\delta(q)$ is the Euclidean distance between the market portfolio p and the portfolio weights of strategy λ^{LW} in state 3. Darker areas represent larger values of $\delta(q)$; and white areas represent those prices for which λ^{LW} is in perfect agreement with the market portfolio, (i.e., $\delta(q) = 0$). Kelly prices and Kelly portfolio weights are given by $\lambda^* = (0.2, 0.1, 0.7)$.

remaining wealth is invested in a market portfolio consisting only of the two assets, 1 and 3. The behavior leads to a convergence of the price of asset 2 to the Kelly benchmark if λ^{LW} becomes wealthy. Other strategies are specialized in very similar fashions but for different assets and states.

The optimality properties of investing part of the wealth so as to reduce the risk associated with the volatility of the portfolio return (the "fractional Kelly Rule") are discussed, for example, in MacLean et al. (1992).

Conclusion

The simulation results of the genetic programming approach to the dynamics and mutation of investment strategy confirms the pivotal role of the Kelly Rule. The long-term outcome of the market dynamics is fully described by the rule; this outcome is also robust against noise. The numerical study also highlights the importance of the market portfolio in this class of models. The market portfolio, even if it is only a proxy, provides insurance against severe losses. In the present genetic programming framework, volatility of returns that is too high is punished by the tournament process that annihilates poor strategies. Surprisingly, perhaps, typical trader types, assumed to populate the market in noise-trader or agent-based models, do not enter the stage.

9.5.3. Empirical Tests of Evolutionary Asset Pricing

In this section two of the theoretical evolutionary finance results are tested empirically. First, the prediction of asset prices derived from the long-term dynamics of the market: the Kelly Rule as a benchmark for the (relative) fundamental valuation of assets. Second, the market dynamics that, in the presence of a Kelly investor, describe the convergence of relative asset prices to the Kelly benchmark. The latter highlights the strength of evolutionary finance models that overcome the shortcomings of equilibrium models in which these convergence dynamics are mainly an exercise in semantics because this dynamic simply is not modeled.

Empirical support for this dynamic approach and its predictions has interesting implications. In this case evolutionary finance can shed light on the issue of excess returns in financial markets, a hot topic ever since these markets came into existence. A prominent example is excess returns from value investment (i.e., bets on the reversal of prices to some fundamental value such as price-to-book ratio or dividend yield). Graham and Dodd (1934), who were the main proponents of this investment advice, conjectured that excess returns from value investment originate from a tendency of markets to converge toward fundamental values. A simple approach to profit from this price dynamic is to go short in overvalued assets (the price of which falls) and long in undervalued assets (the price of which increases). This line of thought is explored in the empirical study of the value-premium puzzle by Hens et al. (2008).

The empirical test employs the evolutionary finance model with long-lived assets—Section 9.4. Each time period is interpreted as one year, and the asset payoffs are given by the vector with each firm's total dividend payment in that year. The data sample consists of all 16 firms listed in DJIA index during the time period 1981 to 2006. The data are taken from CRSP.

Hypothesis 1—Relative Asset Prices Determined by λ^*

Our results state that the relative market capitalization of an asset is (asymptotically) given by the expected value of its discounted relative payoffs. The relative market capitalization of a firm (denoted by $q_{k,t}$) is simply calculated from the stock prices and the number of shares issued for all firms in the sample. How to determine the relative fundamental value, however, is less straightforward and obviously leaves the econometrician with many options. We take the current relative dividend of each firm (denoted by $R_{k,t}$, k, the firm index) as a proxy for the relative fundamental value (the Kelly Rule $\lambda^*_{k,t}$).

Our (joint) hypothesis is that in the linear cross-sectional regression

$$q_{k,t} = a_0(t) R_{k,t} + a_1(t) + \epsilon_t, \quad k = 1, \ldots, 16 \tag{9.56}$$

$a_0(t) > 0$ and $a_1(t) = 0$ for $t = 1981, \ldots, 2006$. If this relation holds, then in each year, the relative market capitalization of a firm depends linearly on its current relative dividend payment.

Hypothesis 2—Convergence of Relative Asset Prices to λ^*

The convergence of prices to the Kelly prices λ^* is a consequence of market dynamics. If the previous hypothesis has sufficient empirical support, one can study the dynamics of small deviations from the benchmark $\lambda^*_{k,t}$. This empirical benchmark will be defined as the valuation derived in the study of Hypothesis 1. Suppose there is one $\lambda^*_{k,t}$ investor and a mutant investment strategy $\mu_{k,t}$ representing all the other investors in the market. Exponentially fast convergence of the Kelly investor's wealth share $r^*_t \to 1$ can be expressed as $[1 - r^*_{t+1}] = \alpha_t [1 - r^*_t]$ with some variable α_t, $0 < \alpha_t < 1$. The mean value of this parameter is determined by the exponential of the logarithmic growth rate $g_{\lambda^*}(\mu)$ as defined in Eq. 9.45. Since

$$q_{k,t} = \lambda^*_{k,t} r^*_t + \mu_{k,t}(1 - r^*_t)$$

one obtains (after some elementary calculations) the relation

$$[\lambda^*_{k,t+1} - q_{k,t+1}] = \alpha_t \frac{\lambda^*_{k,t+1} - \mu_{k,t+1}}{\lambda^*_{k,t} - \mu_{k,t}} [\lambda^*_{k,t} - q_{k,t}] \tag{9.57}$$

Our hypothesis is formalized as follows. Between any two consecutive years, t and $t+1$, $t = 1981, \ldots, 2005$, the linear regression

$$[\lambda^*_{k,t+1} - q_{k,t+1}] = a(t) [\lambda^*_{k,t} - q_{k,t}] + \epsilon_t, \quad k = 1, \ldots, K \tag{9.58}$$

has a least-squares estimator $0 < a(t) < 1$ and ϵ_t is a noise term with mean zero.

TABLE 9.1 Empirical Findings on the Two Hypotheses Derived from Evolutionary Finance

Coefficients, probabilities, and R^2 of the regression (Eq. 9.56) testing the asset pricing Hypothesis 1.

Year t	$a_0(t)$	P-value	$a_1(t)$	P-value	R^2 adj.
1981	0.550	0.000	0.028	0.011	0.671
1982	0.584	0.000	0.026	0.001	0.835
1983	0.613	0.000	0.024	0.005	0.799
1984	0.643	0.000	0.022	0.036	0.710
1985	0.622	0.000	0.024	0.027	0.713
1986	0.609	0.000	0.025	0.058	0.577
1987	0.474	0.000	0.033	0.009	0.485
1988	0.515	0.000	0.030	0.012	0.516
1989	0.549	0.000	0.028	0.013	0.561
1990	0.243	0.030	0.047	0.001	0.145
1991	0.280	0.030	0.045	0.002	0.146
1992	0.328	0.031	0.042	0.006	0.141
1993	0.508	0.003	0.031	0.026	0.315
1994	0.496	0.001	0.032	0.014	0.364
1995	0.575	0.000	0.027	0.020	0.474
1996	0.606	0.000	0.025	0.039	0.442
1997	0.596	0.000	0.025	0.041	0.421
1998	0.743	0.000	0.016	0.125	0.556
1999	0.795	0.000	0.013	0.212	0.518
2000	0.707	0.001	0.018	0.171	0.376
2001	0.891	0.000	0.007	0.154	0.917
2002	0.776	0.000	0.014	0.077	0.816
2003	0.692	0.000	0.019	0.009	0.832
2004	0.674	0.000	0.020	0.004	0.842
2005	0.803	0.000	0.012	0.061	0.873
2006	0.822	0.000	0.011	0.078	0.877

(a)

Coefficient, probabilities, and adjusted R^2 of the regression (Eq. 9.58) testing the convergence Hypothesis 2.

Year t	$a(t)$	P-value	R^2 adj.
1981	0.529	0.000	0.677
1982	0.865	0.000	0.524
1983	1.185	0.000	0.833
1984	0.973	0.000	0.943
1985	1.233	0.000	0.970
1986	0.814	0.000	0.852
1987	0.871	0.000	0.875
1988	0.907	0.000	0.967
1989	1.031	0.000	0.757
1990	1.049	0.000	0.945
1991	0.862	0.000	0.752
1992	0.716	0.000	0.673
1993	0.784	0.000	0.731
1994	0.815	0.000	0.886
1995	0.982	0.000	0.903
1996	1.002	0.000	0.885
1997	0.847	0.000	0.776
1998	1.076	0.000	0.901
1999	1.080	0.000	0.733
2000	0.115	0.113	0.089
2001	0.875	0.003	0.329
2002	0.612	0.000	0.546
2003	0.857	0.000	0.814
2004	0.884	0.000	0.748
2005	0.854	0.000	0.772

(b)

(a) Results on comparison of asset prices with the Kelly benchmark in cross-sections (last trading day in a given year t). (b) Results on the convergence of asset prices to the Kelly benchmark from year t to $t+1$.

The empirical results are summarized in Table 9.1. Hypothesis 1 on the relevance of the Kelly Rule as a pricing benchmark (for the relative valuation of firms) is strongly supported. In every year of the sample, the coefficient $a_0(t)$ is significantly positive. In addition, the coefficient $a_1(t)$ is not significantly different from zero. The adjusted R^2 values indicate that a considerable amount of the variation in the data is explained by the model (see Table 9.1(a)). Hypothesis 2 on the convergence of relative market capitalization toward the benchmark is supported by the empirical findings. Most of the coefficients $a(t)$ (see Table 9.1(b)), are between zero and one.

This finding is statistically significant on the 1% significance level. The adjusted R^2 values are quite high, indicating that the model has strong explanatory power. In seven of the 25 years of observation, the coefficient $a(t)$ is larger than one, implying divergence from the benchmark from the current to the next year. The hypothesis that the coefficient is less than one cannot, however, be rejected at the 1% level. In summary, both hypotheses are strongly supported by the empirical results.

The empirical analysis presented here is certainly not more than a preliminary assessment of the potential of evolutionary finance in explaining asset prices and their dynamics. This topic merits additional (and more thorough) inquiry.

9.6. CONTINUOUS-TIME EVOLUTIONARY FINANCE

This section presents recent progress in the advancement of the evolutionary finance approach in the direction of continuous-time financial mathematics. The development of such an approach is of interest because it builds on the workhorse model of financial mathematics, and it allows for different time scales for trading and changes in dividend payments. The main conceptual innovation is the introduction of the market interaction of heterogenous investors with self-financing investment strategies—and thus endogenous prices—in this framework. The model accommodates, for example, different time scales for the frequency and intensity of trades and dividend payments. This offers an alternative approach to the study of the price impact of large trades. The mathematical theory used to formulate the continuous-time evolutionary finance model is that of random dynamical systems with continuous time (Arnold, 1998).

The analysis focuses, as in the discrete-time model, on the asymptotic dynamics of wealth distribution and asset prices. The derivation of convergence results, however, requires the application of very different mathematical techniques. For simplicity of presentation, only the continuous case (without jumps) is considered here. Details and proofs can be found in Palczewski and Schenk-Hoppé (2008b).

There are K assets (stocks), each with a constant supply of one. Denote the price process, which will be described later, by $S(t) = (S_1(t), \ldots, S_K(t))$ and the cumulative dividend payment by $D(t) = (D_1(t), \ldots, D_K(t))$, $t \geq 0$. There are I investors. The portfolio of investor i is denoted by $\theta^i(t) = (\theta^i_1(t), \ldots, \theta^i_K(t))$, and the person's cumulative consumption process is given by $C^i(t)$.

For a self-financing portfolio-consumption process $(\theta^i(t), C^i(t))$, the dynamics of investor i's wealth, $V^i(t) = \sum_{k=1}^K \theta^i_k(t) S_k(t)$, is given by

$$dV^i(t) = \sum_{k=1}^{K} \theta^i_k(t)\big(dS_k(t) + dD_k(t)\big) - dC^i(t) \tag{9.59}$$

Self-financing means that changes in value can be attributed either to changes in asset prices, dividend income, or consumption expenditure. An investor's portfolio can be written as $\theta^i_k(t) = \lambda^i_k(t)V^i(t)/S_k(t)$ with a real-valued process $\lambda^i(t) = (\lambda^i_1(t), \ldots, \lambda^i_K(t))$ as investment strategy. Since assets have a net supply of one, market-clearing implies

$$S_k(t) = \lambda^1_k(t)\, V^1(t) + \cdots + \lambda^I_k(t)\, V^I(t) = \langle \lambda_k(t), V(t)\rangle \qquad \textbf{(9.60)}$$

that is, every asset's market value is equal to the aggregate investment in that asset, (see Eq. 9.9). This defines a market-clearing price for given investment strategies and wealth distribution. One obtains

$$dV^i(t) = \sum_{k=1}^{K} \frac{\lambda^i_k(t)V^i(t)}{\langle \lambda_k(t), V(t)\rangle}\big(d\langle \lambda_k(t), V(t)\rangle + dD_k(t)\big) - dC^i(t) \qquad \textbf{(9.61)}$$

for all $i = 1, \ldots, I$. Indeed, continuous-time dynamics describes the limit of the discrete-time model, Eq. 9.37, when the length of the time step tends to zero (Palczewski and Schenk-Hoppé, 2008a).

Suppose there are $I = 2$ investors with time-invariant investment strategies (i.e., $\lambda^i_k(t) \equiv \lambda^i_k$) and consumption process $dC^i(t) = c\, V^i(t)dt$. The constant $c > 0$ is the consumption rate; it is assumed to be the same for all investment strategies. Assume that the cumulative dividend process of each asset can be written using an intensity process (i.e., $dD_k(t) = \delta_k(t)dt$ with $\delta_k(t) \geq 0$) and that $\bar{\delta}(t) = \sum_{k=1}^{K} \delta_k(t) > 0$. Then the dynamics of the relative wealth of the investment strategy λ^1, $w^1(t) = V^1(t)/[V^1(t) + V^2(t)]$, is given by

$$dw^1(t) = cw^1(t)\frac{\sum_{k=1}^{K} \frac{\lambda^1_k}{[\lambda^1_k - \lambda^2_k]w^1(t) + \lambda^2_k}\rho_k - 1}{\sum_{k=1}^{K} \frac{\lambda^1_k \lambda^2_k}{[\lambda^1_k - \lambda^2_k]w^1(t) + \lambda^2_k}}dt \qquad \textbf{(9.62)}$$

with $\rho_k(t) = \delta_k(t)/[\delta_1(t) + \delta_2(t)]$, $k = 1, 2$—the relative dividend intensity. The dynamics (Eq. 9.62) is well-defined if $\rho_1(t)$—and thus $\rho_2(t)$—is locally integrable (see Palczewski and Schenk-Hoppé, 2008b, Lemma 2). (The relative wealth of the other investment strategy is given by $w^2(t) = 1 - w^1(t)$.)

Suppose further that there are $K = 2$ assets. Then Eq. 9.62 can be factorized as (provided $\lambda^1 \neq \lambda^2$)

$$dw^1(t) = c\frac{-w^1(t)\big(1 - w^1(t)\big)\big((\lambda^2_1 - \lambda^1_1)^2 w^1(t) + (\lambda^2_1 - \lambda^1_1)(\rho_1(t) - \lambda^2_1)\big)}{\big(\lambda^1_1\lambda^1_2 - \lambda^2_1\lambda^2_2\big)w^1(t) + \lambda^2_1\lambda^2_2}dt \qquad \textbf{(9.63)}$$

We finally assume that

$$\lambda^*_1 = \lim_{t\to\infty} \frac{1}{t}\int_0^t \rho_1(u)du \qquad \textbf{(9.64)}$$

is well-defined ($\lambda_2^* = 1 - \lambda_1^*$). Then one has Theorems 1 and 2, from Palczewski and Schenk-Hoppé (2008b).

Theorem 9.7. *Let $\lambda^2 = \lambda^*$, and assume that $\lambda^1 \neq \lambda^2$. Fix any initial value $w_1(0) \in (0, 1)$.*

(a) $\lim_{t\to\infty} \frac{1}{t} \int_0^t w_1(u)\, du = 0$

(b) Suppose there is a real number γ such that

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{(-\infty,\gamma)}\Big(\operatorname{sgn}(\lambda_1^* - \lambda_1^1) \int_s^t (\rho_1(u) - \lambda_1^*)\, du \Big) ds > 0 \qquad \textbf{(9.65)}$$

Then the relative wealth of investor 1 converges to 0 (while that of investor 2 converges to 1); that is,

$$\lim_{t\to\infty} w_1(t) = 0 \quad \textit{and} \quad \lim_{t\to\infty} w_2(t) = 1$$

Theorem 9.7 states that wealth dynamics selects the investor who divides wealth according to the time average of the relative dividend intensity. This finding is in line with our previous analysis. Part (a) asserts convergence in the Cesàro sense. Counter examples show that the stronger convergence in part (b) cannot be obtained without additional conditions.

It is of interest to note that the speed of convergence of $w^1(t) \to 0$ in Theorem 9.7 is not exponentially fast. This is at odds with the corresponding models in discrete time (see Sections 9.3.2 and 9.4.1). Suppose $\rho(t)$ is a stationary ergodic process with the stationary measure μ; then $\lambda^* = E^\mu \rho$. The linearization at the steady state $w^1(t) = 0$ gives the variational equation

$$dv(t) = c\frac{\lambda_1^1 - \lambda_1^2}{\lambda_1^2 \lambda_2^2}\big(\rho_1(t) - \lambda_1^2\big) v(t)dt$$

which shows that the exponential growth rate of $v(t)$ is equal to

$$c\frac{\lambda_1^1 - \lambda_1^2}{\lambda_1^2 \lambda_2^2}\big(E^\mu \rho_1 - \lambda_1^2\big)$$

If $\lambda^2 = \lambda^*$, the exponential growth rate is equal to zero for *every* investment strategy λ^1. For any time-invariant investment strategy $\lambda^2 \neq \lambda^*$, however, there is an investment strategy λ^1 such that the growth rate is strictly positive (i.e., $v(t)$ diverges from 0 exponentially fast). If $\lambda_1^2 < \lambda_1^*$, take any $\lambda_1^1 \in (\lambda_1^2, 1)$; otherwise, take $\lambda_1^1 \in (0, \lambda_1^2)$.

The condition Eq. 9.65 is satisfied for a large class of processes. Assume for instance that the dividend-intensity process $\rho_1(t)$ is a positively recurrent Markov process on a countable subset of $[0, 1]$. Denote the unique invariant probability measure by μ. Then, in Eq. 9.64, $\lambda_1^* = \bar{\rho} = E^\mu \rho_1(0)$ is well defined. Let P^μ denote the probability measure under which the distribution of $\rho_1(0)$ is given by μ.

Theorem 3 in Palczewski and Schenk-Hoppé (2008b) ensures that if

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t P^{\mu}\Big(\frac{1}{s}\int_0^s \rho_1(u)du < \bar{\rho}\Big) ds > 0$$
$$\lim_{t\to\infty} \frac{1}{t} \int_0^t P^{\mu}\Big(\frac{1}{s}\int_0^s \rho_1(u)du > \bar{\rho}\Big) ds > 0 \qquad \textbf{(9.66)}$$

then Eq. 9.65 holds for $\gamma = 0$. Therefore, $\lim_{t\to\infty} w^1(t) = 0$. For instance, if the process $\rho_1(t)$ has initial distribution μ and is symmetric around its expected value (and takes on at least two different values), then Eq. 9.66 is satisfied. It is also sufficient if the first return time is square integrable for at least one element of the state space E (see Palczewski and Schenk-Hoppé 2008b, Section 4).

An interesting topic for future research on evolutionary finance models in continuous time is the study of Eq. 9.61 with adapted, time-variant investment strategies (possibly more investors and more assets). Another line of inquiry is concerned with the corresponding diffusion-type model that requires the use of stochastic analysis.

9.7. CONCLUSION

This chapter surveyed current research on and applications of evolutionary finance inspired by Darwinian ideas and random dynamical systems theory. This approach studies the market interaction of investment strategies—and the wealth dynamics it entails—in financial markets. We were particularly interested in the long-term dynamics of the wealth distribution with the goal of identifying surviving investment strategies and the corresponding asset-price system. The emphasis in this survey was on the motivation and the heuristic justification of the results; technical details were avoided as much as possible. In contrast to the current standard paradigm in economic modeling, we pursued an approach based on random dynamical systems. Equilibrium holds only in the short term, which reflects the model of investment behavior explored in our evolutionary finance approach.

The motivation was derived in the context of a model of betting markets that goes back to Kelly's paper (1956). The modeling approach and its main components and assumptions were explained in detail in Section 9.2. The main part of the chapter was devoted to the two main modeling frameworks: models with short-lived assets (bets) and those with long-lived assets (stocks). In each case, the analysis moved from (relatively) simple to more demanding settings in which more advanced mathematical techniques were required and the proofs became more involved. In the simplest case considered here, investment strategies are constant vectors and asset payoffs are driven by an IID process. In the most advanced case, the first were adapted processes while the latter were governed by Markov processes.

Models with short-lived assets were covered in Section 9.3. Both local and global dynamics were studied, and some numerical simulations were presented. This model is a generalization of the betting market setting considered by Kelly. Surprisingly, results do not depend on whether the asset market is complete or incomplete (more states than assets). An evolutionary stock market model (with long-lived assets) was the subject of Section 9.4. In this class of models, investors are exposed to capital gains and losses induced by the price dynamics of the assets. This feature has a considerable impact on wealth dynamics and its quantitative study. All results obtained in this framework were presented and explained in detail.

Applications of evolutionary finance models for both short- and long-lived assets were presented in Section 9.5, which contains simulation and empirical studies. The numerical studies explored dynamics beyond the setting in which the analytical results were obtained. We simulated wealth and asset-price dynamics in scenarios with different types of investment strategies—and in the absence of the generalized Kelly Rule. The evolution (or mutation) of strategies, rather than just wealth dynamics of prespecified investment strategies, was numerically analyzed by combining the standard evolutionary finance model with genetic programming and tournament selection. The section closed with the presentation of recent empirical results on the explanatory power of evolutionary finance in real markets.

Continuous-time evolutionary finance models, presented in Section 9.6, are the latest development in this field. This approach can be seen as a generalization of the workhorse model of continuous-time financial mathematics. We introduced endogenous prices via short-term market clearing in this model using the same ideas as in discrete time. One advantage of this model is the flexibility to have different trade frequencies and changes of dividend payments.

Several proposals for future research topics within evolutionary finance were made throughout this chapter. One main task will be to study the game-theoretic perspective of evolutionary finance, which has not been satisfactorily explored yet—though a first step is taken in Amir et al. (2008). Among the challenging (as well as the most rewarding) subjects we highlighted is the need for additional empirical studies and the further development of continuous-time evolutionary finance.

References

Alchian, A., Uncertainty, evolution and economic theory, *Journal of Political Economy*, 1950, 58, 211–221.

Algoet, P. H., and T. M. Cover, Asymptotic optimality and asymptotic equipartition properties of log-optimum investment, *Annals of Probability*, 1988, 16, 876–898.

Amir, R., I. V. Evstigneev, T. Hens, and K. R. Schenk-Hoppé, Market selection and survival of investment strategies, *Journal of Mathematical Economics*, 2005, 41, 105–122.

Amir, R., I. V. Evstigneev, and K. R. Schenk-Hoppé, Asset market games of survival, NCCR Finrisk Working Paper No. 505, 2008.

Arnold, L., *Random Dynamical Systems*, Springer, 1998.

Arthur, W. B., J. H. Holland, B. LeBaron, R. G. Palmer, and P. Taylor, Asset pricing under endogenous expectations in an artificial stock market, in W. B. Arthur, S. Durlauf, and D. Lane (eds.), *The Economy as an Evolving Complex System II*, Addison-Wesley, 1997, pp. 15–44.

Balasko, Y., Temporary financial equilibria, *Economic Theory*, 2003, 21, 10–18.

Björk, T., *Arbitrage Theory in Continuous Time*, 2nd ed., Oxford University Press, 2004.

Blume, L., and S. Durlauf (eds.), *The Economy as an Evolving Complex System III*, Oxford University Press, 2005.

Blume, L., and D. Easley, Evolution and market behavior, *Journal of Economic Theory*, 1992, 58, 9–40.

Blume, L., and D. Easley, If you're so smart, why aren't you rich?: Belief selection in complete and incomplete markets, *Econometrica*, 2006, 74, 929–966.

Blume, L., and D. Easley, Market selection and asset pricing, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Breiman, L., Optimal gambling systems for favorable games, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 1961, pp. 65–78.

Browne, S., The return on investment from proportional portfolio strategies, *Advances in Applied Probability*, 1998, 30, 216–238.

Chiarella, C., R. Dieci, and X.-Z. He, Heterogeneity, market mechanisms, and asset price dynamics, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Cover, T. M., Shannon and investment, *IEEE Information Theory Society Newsletter*, Special Golden Jubilee Issue, Summer 1998, 10–11.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, Noise trader risk in financial markets, *Journal of Political Economy*, 1990, 98, 703–738.

De Giorgi, E., Evolutionary portfolio selection with liquidity shocks, *Journal of Economic Dynamics and Control*, 2008, 32, 1088–1119.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Evolutionary stable stock markets, *Economic Theory*, 2006, 27, 449–468.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Market selection of financial trading strategies: Global stability, *Mathematical Finance*, 2002, 12, 329–339.

Evstigneev, I. V., T. Hens, and K. R. Schenk-Hoppé, Globally evolutionarily stable portfolio rules, *Journal of Economic Theory*, 2008, 140, 197–228.

Farmer, J. D., and A. W. Lo, Frontiers of finance: Evolution and efficient markets, *Proceedings of the National Academy of Sciences*, 1999, 96, 9991–9992.

Friedman, M., *Essays in Positive Economics*, University of Chicago Press, 1953.

Gerber, A., T. Hens, and P. Woehrmann, Dynamic general equilibrium and T-period fund separation, *Journal of Financial and Quantitative Analysis*, forthcoming.

Graham, B., and D. L. Dodd, *Security Analysis*, McGraw-Hill, 1934.

Hakansson, N. H., and W. T. Ziemba, Capital growth theory, in R. A. Jarrow, V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in Operations Research and Management Science, Vol. 9: Finance*, Chapter 3, Elsevier, 1995, pp. 65–86.

Hens, T., and K. R. Schenk-Hoppé, Evolutionary finance: Introduction to the special issue, *Journal of Mathematical Economics*, 2005a, 41, 1–5.

Hens, T., and K. R. Schenk-Hoppé, Evolutionary stability of portfolio rules in incomplete markets, *Journal of Mathematical Economics*, 2005b, 41, 43–66.

Hens, T., and K. R. Schenk-Hoppé, Markets do not select for a liquidity preference as behavior towards risk, *Journal of Economic Dynamics and Control*, 2006, 30, 279–292.

Hens, T., and K. R. Schenk-Hoppé, Survival of the fittest on Wall Street, in M. Lehmann-Waffenschmidt et al. (eds.), *Institutioneller Wandel, Marktprozesse und dynamische Wirtschaftspolitik* (*Proceedings of the VI Buchenbach Workshop*), Marburg, Metropolis, 2004, pp. 339–367.

Hens, T., K. R. Schenk-Hoppé, and M. Stalder, An application of evolutionary finance to firms listed in the Swiss market index, *Swiss Journal of Economics and Statistics*, 2002, 138, 465–488.

Hens, T., K. R. Schenk-Hoppé, and P. Woehrmann, An evolutionary explanation of the value premium puzzle, NCCR Finrisk Working Paper No. 280, April 2008.

Hodgeson, G. M., *Economics and Evolution: Bringing Life Back into Economics*, Blackwell, 1993.

Hommes, C. H., Financial markets as nonlinear adaptive evolutionary systems, *Quantitative Finance*, 2001, 1, 149–167.

Hommes, C. H., and F. O. O. Wagener, Complex evolutionary systems in behavioral finance, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

Hicks, J. R., *Value and Capital*, 2nd edition, Clarendon Press, 1946.

Kelly, J. L., A new interpretation of information rate, *Bell System Technical Journal*, 1956, 35, 917–926.

Konno, H., and H. Yamuzaki, Mean absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, 1991, 37, 519–531.

Laffont, J.-J., *The Economics of Uncertainty and Information*, MIT Press, 1989.

LeBaron, B., W. B. Arthur, and R. Palmer, Time series properties of an artificial stock market, *Journal of Economic Dynamics and Control*, 1999, 23, 1487–1516.

Lensberg, T., and K. R. Schenk-Hoppé, On the evolution of investment strategies and the Kelly Rule—A Darwinian approach, *Review of Finance*, 2007, 11, 25–50.

Lo, A. W., The adaptive markets hypothesis, *Journal of Portfolio Management*, 2004, 30, 17–29.

Lucas, R., Asset prices in an exchange economy, *Econometrica*, 1978, 46, 1429–1445.

Lux, T., Stochastic behavioral asset pricing models and the stylized facts, *Handbook of Financial Markets: Dynamics and Evolution*, 2009.

MacLean, L. C., W. T. Ziemba, and G. Blazenko, Growth versus security in dynamic investment analysis, *Management Science*, 1992, 38, 1562–1585.

Palczewski, J., and K. R. Schenk-Hoppé, From discrete to continuous time evolutionary finance, NCCR Finrisk Working Paper No. 504, 2008a.

Palczewski, J., and K. R. Schenk-Hoppé, Market selection of constant proportions investment strategies in continuous time, NCCR Finrisk Working Paper No. 500, 2008b.

Penrose, E. T., Biological analogies in the theory of the firm, *American Economic Review*, 1952, 42, 804–819.

Perold, A. F., and W. F. Sharpe, Dynamic strategies for asset allocation. *Financial Analysts Journal*, 1988, 44, 16–27.

Pliska, S. R., *Introduction to Mathematical Finance—Discrete Time Models*, Blackwell Publishing, 1997.

Radner, R., Existence of equilibrium of plans, prices, and price expectations in a sequence of markets, *Econometrica*, 1972, 40, 289–303.

Rockafellar, R. T., and S. Uryasev, Optimization of conditional value-at-risk, *Journal of Risk*, 2000, 2, 21–41.

Samuelson, P. A., Why we should not make mean log of wealth big though years to act are long, *Journal of Banking and Finance*, 1979, 3, 307–309.

Schenk-Hoppé, K. R., Random dynamical systems in economics, *Stochastics and Dynamics*, 2001, 1, 63–83.

Shapley, L., and M. Shubik, Trade using one commodity as a means of payment, *Journal of Political Economy*, 1977, 85, 937–968.

Thorp, E. O., Portfolio choice and the Kelly criterion, in W. T. Ziemba and R. G. Vickson (eds.), *Stochastic Models in Finance*, Academic Press, 1971, pp. 599–619.

Tversky A., and D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 1992, 5, 297–323.

Author Index

A
Abhyankar, A., 178
Abreu, D., 29
Acemoglu, D., 444, 469
Adam, K., 255
Admati, A., 6
Aharony, J., 28
Aitken, B., 35
Alchian, A., 219, 387, 408, 408n, 509
Alevy, J. E., 36
Alfarano, S., 164, 190, 208, 209, 242, 315, 339
Algoet, P. H., 518
Allen, F., 28, 443, 445, 447, 449, 450, 482, 483
Allen, H., 178, 254, 280
Allingham, M., 355
Almgren, R., 88
Amato, J., 452
Amihud, Y., 28
Amir, R., 219n, 434, 435, 524, 534, 535, 536, 564
Anderson, L. R., 26, 26n24
Anderson, P. W., 218
Anderson, S. P., 224, 291, 348, 371
Ane, T., 127
Angeletos, G.-M., 347, 374, 443, 452
Anufriev, M., 236, 241, 262, 306n
Aoki, M., 193
Arifovic, J., 184, 206
Arnold, L., 331, 332, 338, 373, 512, 523, 526, 560
Arrowsmith, D. K., 392
Arthur, W. B., 2, 9, 184, 218, 219, 221, 241, 510
Ashiya, M., 21
Ausloos, M., 177
Avery, C., 31
Avnir, D., 177

B
Baak, S. J., 241
Bacchetta, P., 445, 452, 482
Baillie, R. T., 174, 320
Bala, V., 12n
Banerjee, A., 3, 6, 6n4, 8, 15
Barber, B. M., 39, 71
Barberis, N., 210, 218, 243n, 495
Barclay, M. J., 91
Barlow, N., 407
Barnsdorf-Nielsen, O., 177
Barry, F., 37
Barsade, S. G., 6, 34
Barsky, R. B., 307n
Bartholomew, D. J., 41
Bartolozzi, M., 209
Basci, E., 485
Batchelor, R., 456
Baumol, W. J., 179
Beach, R., 33
Beatty, R. P., 24n23
Beaudry, P., 13
Beirland, J., 170
Beja, A., 178, 179, 180, 180n, 182, 197, 199, 281, 284
Beker, P., 432
Beltratti, A., 443, 472, 497
Benartzi, S., 220
Beran, J., 78, 107, 174
Berg, A., 35
Berger, J. A., 41
Bernard, V. L., 280
Bernardo, A. E., 17, 27, 443
Bernhardt, D., 23
Bernheim, B. D., 17
Bessembinder, H., 117
Bhattacharya, S., 8n8, 27, 28
Bhushan, R., 23
Biais, B., 111
Bikhchandani, S., 3, 4n, 6, 6n, 8, 9, 10n, 12, 13, 26, 34, 35, 37
Bizjak, J. M., 39
Björk, T., 513, 522n
Black, F., 66, 111, 279, 353, 409
Black, S., 498, 499
Bloomfield, R., 36n
Blume, L. E., 219, 347, 348, 349, 369, 387, 406, 408, 409, 410, 416, 416n, 418n, 419, 421, 428, 432, 433, 434, 435, 510, 515, 518
Böhm, V., 264, 287, 322, 326n, 331, 348, 349, 353, 354, 355, 356, 357, 359, 363, 364, 374, 378, 387
Bollerslev, T., 137n, 174, 178, 320, 496
Bonabeau, E., 185
Boswijk, H. P., 242, 248n24
Bottazzi, G., 164, 236, 306n
Bouchaud, J.-P., 62, 66, 69, 77, 79, 86, 94, 98, 100, 101, 102, 104, 105, 107, 108, 129, 130, 131, 134, 137n, 206, 209, 241
Box-Steffensmeier, J., 178
Boyd, R., 17, 40
Braglia, M., 194
Branch, W. A., 236n, 242
Brandenburger, A. A., 6n7, 19
Brandt, A., 373
Breiman, L., 426, 434, 518
Brennan, M. J., 20, 35
Brock, W. A., 176, 183, 218, 219, 221, 222n3, 225, 226n, 227, 229n, 230, 236, 238, 241, 256, 265, 282, 285,

Brock, W. A. (*cont'd*)
288, 290, 291, 292, 294, 296, 316, 322, 326n25, 330, 348, 349, 359, 369, 443, 445, 455, 496
Brooks, C., 242
Brown, D., 443, 446, 449, 450
Brown, G., 164
Brown, J. R., 24n22, 25
Brown, K. C., 27
Brown, N. C., 27
Brown, P., 22
Browne, S., 518
Brunnermeier, M. K., 29, 445
Burguet, R., 11

C
Cabrales, A., 443, 455
Cacioppo, J. L., 34
Calomiris, C. W., 27, 28
Calvet, L.-E., 347, 374
Calvo, G., 12
Camerer, C., 480
Caminal, R., 12
Campbell, J. Y., 61, 246n23, 247, 256, 302, 302n, 494, 495, 496
Campello, M., 23
Cao, H. H., 8, 11, 12, 15, 32n, 33
Caplin, A., 13, 16, 37
Capstaff, J., 22
Casdagli, M., 75
Case, K. E., 39, 40
Celen, B., 26n24
Challet, D., 94, 137n
Chamberlain, G., 349
Chamley, C., 8n, 12, 19, 36n
Chamley, C. P., 6n, 8n, 13
Chan, L. K., 80, 88
Chang, S. K., 236
Chari, V. V., 6, 10, 14, 28
Chatterjee, K., 8n
Chattopadhyah, S., 432
Chavas, J. P., 241
Chen, S.-H., 184, 205
Chen, W., 123
Chevalier, J., 27, 220
Chiarella, C., 137n, 182, 183, 219, 222, 236, 264, 281, 284, 286n, 287, 291, 292, 294, 295, 298, 299, 300, 302, 303, 303n, 305, 306, 308, 308n, 309, 313, 314, 322, 325, 326n, 331, 331n, 339, 339n, 340, 348, 349, 353, 356, 359, 364, 374, 510
Chordia, T., 87
Chowdhry, B., 6
Cipriani, M., 30, 31, 36
Clark, P. K., 125, 126
Clement, M. B., 22
Cliff, M., 164
Cochrane, J. H., 280, 349
Cohen, K. J., 137n
Cohen, L., 16n
Cohn, R., 43
Condie, S., 410, 433
Conlisk, J., 17, 18
Cont, R., 69, 209, 241
Cootner, P. H., 409
Copeland, T. E., 353
Cote, I., 22
Coval, J. D., 32n, 33
Cover, T. M., 515, 518
Cox, J. C., 346, 347
Crato, N., 177
Curty, P., 69
Cuthbertson, K., 349
Cutler, D. M., 61, 218

D
Dacorogna, M., 178
dal Forno, A., 164
Dana, R. A., 349, 353, 354, 355, 357
Daniel, K. D., 4n, 18n, 29, 38n, 210
Daniels, M. G., 91, 132, 137, 137n, 141
D'Arcy, S. P., 37n
Dasgupta, A., 35
Davis, G. F., 23
Dawid, A. P., 458n, 472n
Dawid, H., 184
Dawkins, R., 4
Day, R. H., 178, 182, 281, 284
DeBondt, W. F. M., 218
Décamps, J.-P., 32n32
DeCoster, G. P., 37
Defond, M. L., 38
de Fontnouvelle, P., 236
De Giorgi, E., 521, 538
DeGrauwe, P., 183, 236, 257n
Deimling, K., 357
Del Guercio, D. D., 220
de Lima, P. J. F., 177, 185
DellaVigna, S., 16, 18n
DeLong, J. B., 17, 70, 218, 225n, 230n, 307n, 360, 387, 409, 416, 426, 428, 430, 512, 548
DeMarzo, P., 2n, 19, 35
Deneubourg, J.-L., 185
de Palma, A., 224, 291, 348, 371
Detemple, J., 398, 443, 455
Deutscher, N., 348, 359
Devenow, A., 4n
de Vries, C. G., 162, 172
Diamond, D. W., 6, 27, 443, 445
Dieci, R., 183, 236, 265, 284, 285, 286n, 291, 302, 303, 305, 308, 308n, 309, 313, 314, 322, 325, 339, 348, 359
Diks, C. G. H., 228, 241
Ding, Z., 79, 174, 315, 319
Distin, K., 41
Docking, D., 28
Dodd, D. L., 511, 546, 557
Doi, T., 21
Domowitz, I., 137n
Dorigo, M., 185
Drehmann, M., 30
Dua, P., 456
Duffie, D., 346
Duffy, J., 220, 254, 255
Duflo, E., 24
Dugatkin, L. A., 6n, 26n
Dumas, B., 431
Durlauf, S. N., 176, 349, 369, 510
Dutta, P., 408
Dwyer, G. P., 255
Dybvig, P. H., 6

E
Easley, D., 219, 347, 348, 387, 406, 408, 409, 410, 416, 416n, 418n, 419, 421, 428, 432, 433, 434, 435, 510, 515, 518
Eckmann, J.-P., 184
Eckwert, B., 347
Egenter, E., 207, 208
Eguiluz, V. M., 209
Ehrbeck, T., 21
Eichengreen, B., 28
Eisler, Z., 104, 105, 132, 133, 137, 138
Eleswarapu, V., 25
Eliezer, D., 137n
Ellison, G., 17, 27, 220
Ellsberg, D., 433
Embrechts, P., 79
Engel, C. M., 498
Engle, R., 61n, 173, 174, 315, 319, 320
Enke, S., 408, 408n
Evans, G. W., 236n
Evans, M. D. D., 87, 129
Evstigneev, I. V., 219, 219n, 348, 387, 434, 435, 524, 537, 540, 542, 543, 544

F
Fama, E. F., 168, 170, 171, 210, 211, 218, 248, 279, 409, 494, 495, 496
Fan, M., 444, 465
Farmer, J. D., 62, 75, 77, 79, 82, 84, 86, 91, 92, 92n, 94, 99, 101, 102, 107, 118, 123, 124, 126, 129, 132, 133,

134, 137, 137n, 139, 141, 146, 150, 152, 219, 241, 510
Feng, L., 26
Fergusson, K., 171
Figlewski, S., 432
Fisher, F. M., 61
Fisher, K. L., 250
Flaschel, P., 340
Föllmer, H., 191, 236n, 330, 349
Foresi, S., 37n
Foroni, I., 285, 291, 305
Foster, G., 22
Foucault, T., 132
Francis, J. R., 38
Franke, R., 290
Frankel, J. A., 178, 187, 254, 280, 281, 443, 498
Frankfurter, G., 162
Frazzini, A., 16n
Freedman, D., 469
Freixas, X., 353n, 398
French, K. R., 248, 248n, 494, 495, 496
Friedman, D., 255
Friedman, M., 218, 387, 408, 512
Froot, K. A., 21, 36, 178, 187, 254, 280, 281, 498
Fudenberg, D., 15, 17

G

Gabaix, X., 87, 90, 125, 126, 127, 174
Gale, D., 6n, 8n, 12, 15, 28, 357
Gallegati, M., 176, 339n
Gardini, L., 183, 236, 284, 285, 291, 302, 303, 305, 308, 308n, 309, 313, 314, 325
Gaunersdorfer, A., 183, 193, 222n, 227n, 241, 291, 316
Geanakoplos, J., 68
Geman, H., 127
Georges, C., 184
Gerber, A., 255, 547
Gerig, A., 82, 99, 101, 106, 107
Gervais, S., 31, 218
Giardina, I., 206
Gibson, R. M., 6n5
Gielens, G., 177
Gilbert, R. J., 37
Gilboa, I., 433
Gillemot, L., 127, 128
Gilli, M., 190, 241
Giraldeau, L.-A., 6n5
Givoly, D., 22, 24
Glaeser, E. L., 24
Glosten, L. R., 30, 34, 74, 111, 114, 136, 165
Godin, J. J., 26n24
Goeree, J., 17
Gohberg, I., 366
Goldbaum, D., 236
Goldfarb, B., 43
Goldman, M. B., 178, 179, 180, 180n, 182, 197, 199, 281, 284
Gompers, P., 36
Gonzalez, F. M., 13, 36n
Gopikrishnan, P., 83, 84, 125, 167, 172, 174
Gordon, M., 243
Görg, H., 37
Gorton, G., 27, 28
Goyal, S., 12n
Graham, B., 511, 546, 557
Graham, J. R., 23
Grandmont, J.-M., 348, 362
Granger, C. W. J., 79, 174, 178, 315, 319
Granovetter, M., 24, 35
Grenadier, S. R., 8n8, 13
Grencay, R., 206
Greve, H. R., 23, 37n
Griffiths, M., 26
Grimaldi, M., 183, 236
Grinblatt, M., 25, 26, 26n
Grossman, S. J., 34, 68, 69, 432, 443, 445
Grundy, B., 443, 445, 449, 450
Gu, G.-F., 123
Gu, M., 182
Guarino, A., 30, 31, 36
Guedj, O., 66
Gul, F., 10n10
Guo, W. C., 444, 472, 484
Gupta-Mukherjee, S., 16

H

Haag, G., 193
Haigh, M. S., 36
Hakansson, N. H., 518
Haltiwanger, J., 30
Hamao, Y., 37n
Hammond, P., 501
Han, B., 32n33
Handa, P., 75
Harlow, W. V., 27
Harris, L., 122
Harris, M., 443, 455
Harrison, M., 443, 455, 484
Hasbrouck, J., 60, 70, 86, 110, 122
Hatfield, E., 34
Haunschild, P. R., 37
Hauser, S., 28
Hausman, J., 86
Hayek, F. V., 45
He, H., 443, 445, 449, 452
He, T., 183
He, X., 184, 222, 236, 284, 285, 286n, 291, 292, 294, 295, 298, 299, 300, 302, 303, 303n, 305, 306, 314, 316, 319, 322, 325, 331n, 339, 339n, 348, 359
Heath, C., 39, 41
Heemeijer, P., 255
Hellwig, C., 443, 452
Hendricks, K., 8n
Henrich, J., 17, 40
Hens, T., 219, 331, 348, 349, 353, 354, 355, 357, 387, 434, 435, 524, 526, 528, 529, 533, 538, 548, 550, 557
Hey, J. D., 30, 255
Hillebrand, M., 340, 374, 375, 377, 381, 383, 385, 386, 395, 397, 398
Hillion, P., 26n24
Hirschey, M., 28
Hirshleifer, D., 3, 4n, 5, 6, 6n, 8, 9, 10n10, 11, 12, 13, 14, 15, 17, 18n, 21, 26, 29, 32n, 33, 34, 35, 37, 38, 44, 218
Ho, T., 287
Hodgeson, G. M., 509
Hoglund, J., 6n5
Holt, C. A., 36
Hommes, C. H., 183, 218, 219, 221, 222n3, 225, 226n, 227, 227n, 229n, 230, 236, 241, 255, 256, 259, 262, 265, 282, 284, 285, 288, 290, 291, 292, 294, 296, 298, 300, 316, 322, 326n, 330, 339n, 348, 349, 359, 369, 443, 455, 510, 522, 548
Hong, H., 24, 24n22, 25, 33, 218
Hong, H. G., 22
Hopman, C., 86, 87, 91, 129
Horgan, J., 178
Horst, U., 330, 348, 349, 359, 369, 374, 387, 392
Hoshi, T., 443, 455
Hosking, J. R. M., 79
Hu, X., 38
Huang, M., 210
Huang, W., 178, 182, 281, 284
Huberman, B. A., 241
Huberman, G., 32n, 347
Huffman, G., 347
Hung, A., 26n24
Hung, H., 340

I

Ikaheimo, S., 25
Ingersoll, J. E., 346, 347

Iori, G., 137n, 209, 241, 339n
Ippolito, R. A., 220
Ito, K., 254
Ivkovich, Z., 24n, 25, 32n

J

Jacklin, C., 28
Jaffe, J. F., 23
Jagannathan, R., 6
Jansen, D., 172
Jegadeesh, N., 23
Jennings, R., 443, 445, 446, 449, 450
Jensen, H., 178
Jin, H., 443, 464n
Jondeau, E., 172
Jones, E., 28
Joshi, S., 241
Jouini, E., 339
Joulin, A., 61, 74, 125, 126, 129
Joyeux, R., 79
Judd, K. L., 218, 347, 359, 374, 443

K

Kahn, C. M., 10
Kahneman, D., 33, 220, 258, 550
Kaldor, N., 188n, 483, 484
Kandel, E., 443, 455
Kaniel, R., 19, 26
Kantz, H., 232n13
Kaplan, J. M., 16
Karceski, J., 220
Kariv, S., 26n
Karpoff, J., 485
Katsaris, A., 242
Kedia, S., 37n, 38
Kehoe, P. J., 10, 14, 28
Keim, D. B., 86, 90, 280
Kelley, H., 255
Kelly, J. L., 434, 515, 527, 563
Kelly, M., 24n
Keloharju, M., 25, 26
Kempf, A., 86
Kertecz, J., 132, 133
Keynes, J. M., 441, 450, 483
Khanna, N., 15n, 20
Kim, C., 18
Kim, W., 23
Kindleberger, P., 163, 210
Kinoshita, Y., 37
Kirman, A. P., 164, 185, 188, 189, 190, 207, 241, 281, 315, 330, 349
Kirsch, D., 43
Kirsh, A., 28
Knight, F. H., 407n
Kodres, L. E., 26n25
Koedijk, K., 172
Kogan, I. I., 137n
Kogan, L., 431
Konno, H., 551
Kon-Ya, F., 39
Koopmans, T. C., 408
Korn, O., 86
Kovenock, D., 8n
Kremer, I., 19
Kreps, D., 443, 455, 484
Krugman, P. R., 30
Kubik, J., 22, 24, 24n22, 25
Kübler, F., 346, 347, 359, 374, 398
Kumar, A., 26
Kuran, T., 17, 33, 35
Kurz, M., 443, 456, 457, 459n, 461, 462, 464n, 465, 467, 468, 469, 470n, 472, 472n, 475, 477, 478, 479, 480, 481, 487, 488n, 498, 499
Kushner, H. J., 390
Kutsoati, E., 23
Kuznetsov, Y., 266
Kyle, A. S., 34, 34n, 70, 73, 76, 80, 111, 165, 179, 257n, 409, 409n

L

Laffont, J.-J., 512, 522
Laitenberger, J., 349, 353, 354, 355, 357
Lajeri, F., 354
Lakonishok, J., 22, 26n25, 26n26, 80, 88
Lancaster, P., 366
La Spada, G., 128
Leahy, J., 13, 16, 37
LeBaron, B., 80, 177, 184, 219, 241, 282, 496, 510
Lee, C. M., 26
Lee, I. H., 10n10, 32, 33
Lee, W., 164
Lemmon, M. L., 39
Lensberg, T., 513, 553, 554
Leombruni, R., 339n
Lerner, J., 36
LeRoy, S. F., 166, 349, 351n, 353, 368
Levy, H., 285, 302, 302n, 339n
Levy, M., 236n16, 285, 302, 302n, 339n
Lewis, K. K., 32n
Li, T., 339
Li, Y., 184, 314, 316, 319
Libby, R., 36n
Lieberman, M., 37
Lillo, F., 62, 75, 79, 80, 81, 82, 82n, 84, 86, 88, 94, 99, 101, 102, 107, 108, 109, 125, 126, 130, 133, 134, 146, 147
Lim, S. S., 38
Lindahl, E., 441
Lintner, J., 279, 346, 349, 353
List, J. A., 36
Liu, P., 187
Lo, A. W., 79, 86, 256, 510
Lobato, I. N., 79, 174
Löffler, A., 349, 353, 354, 355, 357
Longin, F., 172
Lovallo, D., 480
Lovo, S., 32n32
Lowry, M., 36
Lucas, R. E., 218, 346, 347, 443, 450, 452, 500, 522, 538
Luckock, H., 132
Luenberger, D. G., 357, 368
Lundholm, R., 10n
Lux, T., 17, 125, 171, 172, 177, 184, 190, 191, 194, 199n, 200, 200n, 203, 204, 205, 206, 207, 208, 219, 241, 281, 282, 302, 315, 315n, 331, 339, 339n, 511
Lynch, A., 17, 39
Lyons, R. K., 66, 73, 76, 79, 87, 129, 149n

M

MacKinlay, A. C., 86, 256
MacLean, L. C., 557
Macy, M., 36
Madhavan, A., 86, 94, 111, 287
Madrian, B., 25
Maggitti, P. G., 42n
Magill, M., 346, 374, 398
Mahoney, J. M., 23
Mailath, G., 407, 432
Malamud, S., 431
Malloy, C. J., 16n
Mandelbrot, B. B., 79, 125, 168, 170, 171, 177
Manski, C. F., 25, 224, 291
Mantegna, R. N., 125
Manzan, S., 248n
Marchesi, M., 200, 200n, 203, 204, 205, 206, 207, 241, 282, 315
Marimon, R., 254, 255
Markowitz, H. M., 279, 349, 352
Marsh, T., 280
Marshall, A., 407
Marsili, M., 69, 322
Martin, G. S., 24
Maslov, S., 137n
Mason, J. R., 28
Massa, M., 16n
McFadden, D., 224, 291
McGough, A., 236n

McGoun, E., 162
McKelvey, R., 17
McNichols, M., 443, 445, 449, 450
Mehra, R., 346, 464
Mei, J., 37n
Mendelson, H., 137n
Mendoza, E., 12
Merton, R. C., 33, 279, 280, 346, 353
Michaely, R., 24n
Mike, S., 101, 123, 124, 129, 132, 134, 137n, 141
Mikhail, M. B., 22
Mikkelsen, H., 174, 320
Milaković, M., 208
Milgrom, P. R., 30, 34, 68, 74, 111, 165, 280, 484
Miller, D. A., 43
Miller, E. M., 484
Modigliani, F., 43
Mody, A., 37
Morana, C., 497
Moreland, R. L., 33
Morone, A., 30
Morris, S., 443, 445, 452, 482, 483, 484, 500
Moscarini, M., 13
Moskowitz, T. J., 32n33
Mossin, J., 279, 346, 349
Motolese, M., 443, 444, 464n, 465, 467, 472, 475, 476, 477, 478, 479, 480, 487, 488n, 498, 499
Mukherji, A., 443, 452
Mullainathan, S., 17
Murthy, S., 443, 455
Muth, J. F., 218
Myrdal, G., 441

N
Nakata, H., 444, 472
Nanda, V., 6
Napp, C., 339
Neeman, Z., 28
Nelson, D., 496
Nelson, M. W., 36n
Newman, M. E. J., 178
Ng, L. K., 25
Nielsen, K. C., 443, 460, 463, 464n, 472, 480, 502
Nielsen, L. T., 349, 351, 353n, 354, 355, 357
Nikaido, H., 357
Noah, R., 17
Noeth, M., 36n
Nofsinger, J. R., 26n
Noreen, E., 22

O
O'Connell, P., 36
Odean, T., 39, 71, 218
Oechssler, J., 30
O'Grada, C. O., 24n22
Oh, P., 37n
O'Hara, M., 111, 165, 285
Orosel, G. O., 28, 347
Ottaviani, M., 6n, 7, 15n, 23
Owen, J., 349
Ozsoylev, H. N., 2n, 35

P
Packard, N. H., 75
Pagan, A., 162, 315
Palczewski, J., 560, 561, 562, 563
Palestrini, A., 339n
Palfrey, T., 17
Palmaon, D., 24
Pancotto, F., 306n
Pantzalis, C., 18
Pape, B., 207
Park, A., 18
Pattillo, C., 35
Paudyal, K., 22
Pavan, A., 443
Paye, B., 494
Pearson, N. D., 443, 455
Peck, J., 287
Peleg, B., 414
Pellizzari, P., 164
Peng, L., 18n
Pennacchi, G. G., 10
Penrose, E. T., 509
Percival, D., 154
Perron, P., 280
Persons, J. C., 13
Pesaran, H., 497
Pfleiderer, P., 6
Phelps, E., 443, 450
Pigou, A. C., 441
Pincus, M., 38
Place, C. M., 392
Platen, E., 171
Plerou, V., 84, 87, 123, 126
Pliska, S. R., 513, 522n
Plott, C., 26n
Polak, B., 6n7, 19
Polemarchakis, H., 347
Pollet, J., 18n
Pollock, T. G., 42n
Ponzi, A., 124, 126
Poterba, J. M., 494, 495, 496
Potters, M., 86
Pound, J., 24
Prat, A., 35
Prause, K., 177
Prendergast, C., 15n, 19, 20, 23
Prescott, E. C., 346, 464
Pritsker, M., 26n
Pruitt, S. W., 280
Puthenpurackal, J., 24
Pyle, D., 398

Q
Quinzii, M., 346, 374, 398

R
Rabinovitch, R., 349
Radner, R., 408, 546
Raffaelli, G., 322
Rajan, R. G., 6n, 18
Rajgopal, S., 37n, 38
Ramsey, J., 194
Rangel, J., 61n
Rao, H., 23
Rapson, R. L., 34
Rau, R., 3, 36
Raviv, A., 443, 455
Rees, W., 22
Reitz, S., 241
Rheinlaender, T., 331
Richardson, S., 22
Richerson, P. J., 17
Rindova, V., 42n
Ritter, J. R., 3, 24n, 36
Rochet, J.-C., 353, 353n, 398
Rockafeller, R. T., 351, 551
Rockinger, M., 172, 220
Rodman, L., 366
Rogers, B., 17
Roider, A., 30
Roll, R., 61
Romer, D., 33
Rose, A. K., 28, 443, 498
Rosenow, B., 91, 102, 129
Ross, S. A., 67, 346, 347
Rosu, I., 132
Rothe, C., 348
Ruelle, D., 184
Russo, J. E., 480

S
Saar, G., 26
Sacerdote, B., 26n
Saez, E., 24
Samuelson, L., 8n
Samuelson, P. A., 408n, 518
Samuelson, W., 25

Sandas, P., 136
Sanders, D., 22
Sandroni, A., 347, 348, 387, 406, 407, 421, 432
Sargent, T. J., 348
Saunders, A., 28
Savage, L. J., 410, 411, 416
Savin, N. E., 174
Schaller, H., 242
Scharfstein, D. S., 6n7, 18, 21, 23
Scheinkman, J., 24, 184, 349, 369
Schelling, T. C., 35
Schenk-Hoppé, K. R., 219, 331, 332, 333, 336, 348, 387, 434, 435, 512, 513, 524, 526, 528, 529, 533, 538, 548, 550, 553, 554, 560, 561, 562, 563
Schmalensee, R., 254
Schmalz, M., 209
Schmedders, K., 346, 398
Schmedders, K. H., 347, 359, 374
Schmeidler, D., 433
Schneider, M., 443, 461, 462, 472
Schoemaker, P. J. H., 480
Scholes, M., 279
Schornstein, S., 184, 206, 208, 331
Schreiber, T., 232n
Schumpeter, J. A., 407n
Schwartz, R. A., 75
Schweizer, M., 349
Schwert, G. W., 36, 497
Sciubba, E., 432
Seasholes, M. S., 26, 32n, 36
Sebenius, J. K., 68
Selden, L., 398
Sethi, R., 290
Sgroi, D., 18, 26n
Shapley, L., 513, 522
Sharma, S., 4n, 34, 35
Sharpe, W. F., 279, 346, 349
Shastri, K., 353
Shaw, W. H., 24n23
Shea, D., 25
Shiller, R. J., 15, 17, 24, 25, 35, 39, 40, 41, 42, 43, 61, 68, 220, 241, 244, 246, 246n, 247, 248n, 250, 254, 280, 307n, 491, 494, 495, 496
Shilling, G. A., 491, 491n
Shin, H. S., 443, 445, 452, 482, 500
Shive, S., 24n22, 25, 35, 41
Shleifer, A., 17, 26n, 70, 218, 235, 360, 387, 416, 428
Shubik, M., 513, 522
Sias, R. W., 26n, 35
Simon, H. A., 18, 280, 348, 369
Simonov, A., 16n
Simunic, D., 24n
Singleton, K., 443, 445, 452
Sirri, E. R., 220
Slanina, F., 132, 137n
Slezak, S. L., 15n, 20
Smallwood, D., 17
Smith, E., 137n
Smith, L., 9, 14
Smith, P. A., 24n22
Smith, R., 178
Smith, V., 220, 254
Solomon, A., 22
Solomon, S., 302, 302n, 339n
Sopranzetti, B., 10
Sørensen, P. N., 6n, 7, 9, 15n, 23
Spatt, C., 26n
Spear, S. E., 254
Sri Namachchivaya, N., 332
Stanley, M. H. R., 175
Stapleton, R., 374, 376
Starks, L. T., 27, 35
Starr, R., 501
Statman, M., 250
Stauffer, D., 209
Stein, J. C., 6n7, 21, 24, 24n, 25, 33, 218
Steinkamp, M., 331
Stickel, S. E., 22
Stiglitz, J. E., 34, 68, 432, 443, 445
Stinchcombe, R., 137n
Stock, H. J., 441, 465
Stokey, N., 68, 280, 484
Stole, L., 23
Stoll, H. R., 111, 117, 121, 287
Stouraitis, A., 3
Stracca, L., 280
Strang, D., 36
Strange, W. C., 37
Strobl, E., 37
Subrahmanyam, A., 17, 21, 87
Subrahmanyam, M., 374, 376
Suchanek, G. L., 254
Summers, L. H., 17, 247, 280, 360, 387, 416, 428, 494, 495, 496
Sunder, S., 254, 255
Sunstein, C., 33
Sutan, A., 255
Svenson, O., 480
Swary, I., 28
Szpiro, G. G., 184

T

Takens, F., 75
Taylor, H. M., 125
Taylor, M. P., 178, 254, 280
Taylor, S., 173
Teoh, S. H., 4n, 5, 18n, 22, 27, 28, 29, 38
Teräsvirta, T., 248
Tesar, L., 32n
Tesfatsion, L., 218
Tessone, C., 208
Teugels, J., 170
Teyssière, G., 189, 190, 241
Thakor, A., 27
Thaler, R. H., 218, 220
Theraulaz, G., 185
Thisse, J.-F., 224, 291, 348, 371
Thomas, A., 209
Thornton, H., 440
Thorp, E. O., 518
Timmerman, A., 494, 497
Tirole, J., 483
Titman, S., 17, 21, 24n, 26, 26n
Tkac, P. A., 220
Tobin, J., 349, 353
Toral, R., 208
Torre, N., 86
Townsend, R., 443, 452
Trubowitz, E., 431
Trueman, B., 6n, 19, 22, 24n
Tse, S. Y., 22
Tsutsui, Y., 39
Tsyvinski, A., 443, 452
Tufano, P., 220
Turnovsky, S. J., 254
Tversky, A., 33, 220, 258, 550

U

Uryasev, S., 551
Uttal, B., 28

V

Vaglica, G., 80, 82, 88, 90, 145, 145n, 146
Vanden, J., 162
van der Weide, R., 241
van de Velden, H., 255
van Ness, J. W., 79
van Norden, S., 242
van Wincoop, E., 445, 452, 482
Varian, H. R., 443, 455
Vayanos, D., 2n, 35
Velasco, C., 79, 174
Veldkamp, L. L., 16, 36n
Venables, A. J., 30
Verardo, M., 35

Verrecchia, R., 443, 445
Viceira, L., 302, 302n
Vishny, R. W., 26n, 235
Vissing-Jorgensen, A., 220, 254
Vives, X., 9, 10n, 11, 30
Vynckier, P., 170

W
Waelbroeck, H., 107
Wagener, F. O. O., 183, 227n, 236, 265, 330, 510, 522, 548
Wagner, F., 190, 208, 315, 339
Waldmann, M., 30
Waldmann, R. J., 18, 21, 360, 387, 416, 428
Walther, B. R., 22
Wang, D., 331n, 339n
Wang, J., 137n, 443, 445, 449, 450, 451, 452
Warner, J. B., 91
Warther, V. A., 13
Wasley, C., 38
Watson, W. M., 441, 465
Weber, P., 91, 102, 105
Wei, K. D., 27
Weidlich, W., 193
Weisbenner, S., 24n, 32n
Welch, I., 3, 4n, 6, 6n, 8, 9, 10n, 12, 13, 14, 17, 23, 26, 27, 28, 34, 35, 36, 37
Wenzelburger, J., 264, 282n, 287, 322, 331, 340, 348, 349, 353n, 359, 363, 364, 366, 367, 374, 375, 377, 378, 379, 381, 383, 386, 387, 392, 395, 397
Wermers, R., 26n
Wermers, R. R., 27
Werner, I. M., 32n
Werner, J., 349, 351n, 353, 368
Werning, I., 443, 452
Westerhoff, F. H., 241, 264, 265, 322
Weston, J. F., 353
Whitby, R. J., 39
White, R. E., 280
Wiesinger, J., 116
Williams, A. W., 254
Williams, G. C., 406
Willinger, M., 255
Willis, R. H., 22
Wilson, B., 28
Winker, P., 190, 241
Winter, S., 408
Woodford, M., 443, 452, 482
Wright, J., 178
Wu, F., 25
Wu, H. M., 443, 444, 457, 470n, 472, 484
Wyart, M., 75, 120, 121, 122, 129, 130, 131, 133
Wyplosz, C., 28
Wysocki, P. D., 22

X
Xiong, W., 18n, 236n, 257n

Y
Yaari, M. E., 414
Yamamoto, R., 80
Yamuzaki, H., 551
Yan, H., 431
Yeh, C.-H., 184
Yin, G. G., 390
Youssefmir, M., 241

Z
Zamani, N., 91, 150, 152
Zawadowski, A., 124, 125
Zeckhauser, R., 25
Zeeman, E. C., 182, 210, 281, 331
Zeira, J., 13
Zemsky, P., 31
Zhang, H. H., 32n
Zhang, J., 8n, 13, 19
Zheng, M., 331n
Zhou, W.-X., 123
Zhu, N., 32n
Ziegler, A., 340
Ziemba, W. T., 518
Zimmermann, M. G., 209
Zitzewitz, E., 23
Zovko, I., 82, 133, 134
Zschischang, E., 302
Zumbach, G., 131, 132, 133
Zwart, G. J., 235
Zwiebel, J., 2n, 6n, 19, 35, 38

Subject Index

A
Absolute risk aversion
constant. *See* Constant absolute risk aversion
decreasing, 302
Adaptive belief systems, 183–184, 219
behaviorally consistent, 235
examples of, 226–236
extensions of, 236
heterogeneous expectations, 219
market efficiency in, 230–232
short-run profits, 232
wealth accumulation, 232–235
Adaptive strategies
global dynamics with, 534–537
performance of, 549–552
Adverse selection, 111–112
Advertising, 16–17, 42
Agency-induced herding, 35
Agglomeration, 37
Aggregate behavior, 259
Aggregate consumption growth, 493
Aggregate generational portfolio, 386
Aggregate market impact, 86–89, 108
Aggregate risk, 433
Aggregates, 452
Aggregate transactions, 86–88
Aggregation, 85
Analyst earnings forecasts, 22
Anonymous set, 100
Arbitrary modeling construct, 449
Arbitrate efficiency, 67
ARMA models, 174
Arrow securities, 414–415
Artificial markets, 184, 219
Asset(s)
multiple risky, 322–330, 359
risk-free, 326
portfolio with, 349–351
short-lived, 519, 522, 563
evolutionary model with, 524–537
Asset payoffs, 522
Asset pricing, 514
under asymmetric information, 445–450
determination of, 521
equilibrium, 517
evolutionary, 557–560
with heterogeneous beliefs, 474–485
trading volume and, 484
Asset-pricing models, 178–179
capital, 29, 34, 346–347
asset-market equilibrium in, 355–358
certainty equivalent pricing formula of, 357
with heterogeneous agents, 397
nonergodic asset prices, 387–397
planning horizons, 374–387
as two-period equilibrium model, 349–359
empirical validation of, 241–253
evolutionary dynamics of, 224–225
with heterogeneous beliefs, 221–226, 253–264
laboratory experiments, 253–264
noisy rational expectations, 440
price-to-cash flows, 242–244
with social interactions, 197
summary of, 264–265
two-type example, 246–250
Asset returns
cubic law of, 172
distributional properties of, 172
fat tails of, 167–172, 491
Asymmetric information
asset pricing under, 445–450
market dynamics and, 453–454
secretive economy and, 454
Asymmetric liquidity, 99, 101–103, 105
Autocorrelation coefficients, 314–315, 318
Autoregressive dependence, 194

Availability cascades, 33–34, 42
Availability heuristic, 33
Average market belief, 465, 469, 481–482

B
Bank runs, 27–28
BARRA market impact model, 86
Bayesian inference, 467–469
Bayesian learners, 422
Bayes rule, 412
Beauty Contest metaphor, 483
Behavioral asset-pricing models, 178–179
Behavioral coarsening, 7–9
Behavioral convergence, 5–7
Behavioral economics, 455
Belief, 412–413, 440
 average market, 465, 469, 481–482
 correlation of, 486
 finite, 486, 488
 fixed distribution of, 444
 heterogeneous. *See* Heterogeneous beliefs
 higher-order, 481–482
 individual states of. *See* Individual beliefs
 learning rules based on, 413
 parametrized structure of, 486–491
 rational diverse. *See* Rational diverse beliefs
 rationality, 461–464
 state of, 465
Benchmark expectation rules, 257–259
Bias
 in analyst earnings forecasts, 22
 conversational, 41
 forward discount, 498–499
 persuasion, 35
 psychological, 17–18, 33, 40, 45, 243
Bid–ask spread
 determinants of, 111–125
 Glosten-Milgrom model for, 113–116
 liquidity crisis effects on, 123–125
 liquidity providing costs associated with, 111
 models for, 113–117
 MRR model with, 116–117, 155–156
Bifurcation(s)
 D-, 333–334
 Hopf, 230, 241, 269–270
 P-, 336–338
 period-doubling, 268–269
 pitchfork, 239–240, 270–271
 saddle-node, 268
 stochastic, 332–338
Bifurcation theory, 266–267
Biology, 405–407
Boom-and-bust scenario, 394–398
Bounded actions, 10–11
Boundly rational heterogeneous agent, 279, 281
 empirical behavior of, 314–321
 framework of, 339
 models, 282
Bubbles, 34, 252–253

C
Cancellations, 65
Capital asset-pricing model, 29, 34
 asset-market equilibrium in, 355–358
 certainty equivalent pricing formula of, 357
 with heterogeneous agents, 397
 nonergodic asset prices, 387–397
 overview of, 346–347
 planning horizons, 374–387
 as two-period equilibrium model, 349–359
Capital markets, 2–4
Capital market trading, 2
 models of, 34
CARA. *See* Constant absolute risk aversion utility
Center manifold theorem, 267
Central Limit Law, 169–170
Certainty equivalent pricing formula, 357
Chaotic price fluctuations, 182, 230
Chartists
 Fundamentalists and, interactions between, 179–185, 323–325
 stochastic model with, 331–332
Clearing prices, 30
Clustering, 10
 volatility, 173–174, 183, 320
Cohorts of investors, 375–378
Competitive equilibrium, 414–415
Concave function, 90–92
Conditional stability theorem, 461–464
Conformism, 17
Constant absolute risk aversion utility, 282–284
 price dynamics implied by, 288–302
 summary of, 338
Constant relative risk aversion, 282–283, 285, 338
 heterogeneous agents, 312–313
 optimal portfolio allocation under, 302
 price behavior and wealth dynamics implied by, 302–314
Consumption growth, 493
Consumption plan, 411
Contagion
 of bank runs, 27–28
 of financial memes, 39–44
Continuous belief systems, 241
Continuous signal values, 14–15
Continuous state models, 472
Continuous-time evolutionary finance, 560–563
Continuous time reformulation, 502
Continuous unbounded actions, 10–11
Contracts, 18–20
Convergence, 259, 444

Conversation, 16–17
Conversational bias, 41
Correlation, 128
Covariance matrix, 324, 350
CRRA. *See* Constant relative risk aversion
Cubic law of asset returns, 172
Cum-dividend prices, 363

D
D-bifurcation, 333–334
Decision making
  about payoffs, 17
  conversation effects on, 16
  financing, 36–37
  information cascades' effect on, 7–8
  investment, 36–37
Decreasing absolute risk aversion, 302
Demand elasticity of price, 60
Derivatives, 398
Dirac measure, 369
Direct trading cascades, 31
Disclosures, 38–39
Discrete actions, 10–11
Discrete signal values, 14–15
Dispersing, 5
  market interactions as cause of, 30
Dispersion, reputational, 6–7
Dividend yield, 495–496
Dow Jones Industrial Average, 548
Dynamic infinite horizon models, 450–453
Dynamic stability with rational expectations, 371–374
Dynamic systems theory, 528

E
Economic law, 378
Effective market order, 65
Efficiency
  arbitrate, 67
  informational, 68–69
  market, 67–68, 230–232
Efficient market hypothesis, 165, 173, 175
Emotional contagion, 17
Endogenous amplification, 444, 453, 469–470, 472–473
Endogenous processes, 348
Endogenous uncertainty, 457, 476–480, 501
Endorsements, herding on, 24–25, 37
Endowment stream, 411
Environmental variables, 13–14
Equilibrium
  in capital asset-pricing model, 355–358
  competitive, 414–415
  with heterogeneous beliefs, 358
  Kelly Rule in, 546–547
  models of, 61, 68
  temporary, 359–362
Equilibrium allocations
  competitive equilibrium, 414–415
  Pareto optimality, 413–414, 501–502
Equilibrium asset pricing, 517
Equilibrium dynamics, 497
Equity Premium Puzzle, 493–494
Ergodicity, 459
Ergodic Theorem, 457
Evolutionary asset pricing, 557–560
Evolutionary finance
  applications of, 547–560
  asset pricing, 514
  background for, 509–511
  characteristics of, 510
  continuous-time, 560–563
  dynamic interaction, 513–514
  dynamics of, 512–515
  genetic programming, 552–553
  heterogeneity, 512–513
  selection and stability, 514–515
  strategies, 513
  summary of, 563–564
Evolutionary fitness, 224
Evolutionary ideas, 509–510
Evolutionary model, 518–519
  assumptions, 521–524
  components of, 519–521
  dynamics of, 523–524
  with short-lived assets, 524–537
Evolutionary modeling, 510
Evolutionary stability, 523
Evolutionary stock market model, 537–547
Excess demand, 178
Exchange membership codes, 82–83
Ex-dividend prices, 363
Exogenous processes, 348
Expectations feedback, 287
Extreme value distributions, 170

F
FARIMA process, 107
Fat tails of asset returns, 167–172, 491
Financial ideas, 41
Financial ideologies, 4, 42
Financial markets
  herding in, 25–27
  runs in, 27–28
Financial memes, 4
  assemblies, 41–42
  contagion of, 39–44
  market conditions' effect on, 43
  reproduction of, 41
  spread of, 46

Financial prices, martingale property of, 164–165
Financing decisions, 36–37
Finite belief state, 486, 488
Finite state models, 472
First-jump moment, 194
First moments, perfect forecasting rules for, 363, 379–381
Fitness, evolutionary, 224
Fitness index, 405
Fitness measure, 370
Fokker-Planck equation, 193, 337
Forecast(s)
  cross-sectional variance of, 441
  historical studies of, 441
  stationary, 441
Forecasters, herd behavior by, 21–24
Forecasting
  experiments to learn, 255–256
  nearest neighbor, 232
Forecasting rules, 225–226, 347
  perfect. *See* Perfect forecasting rules
  technical trading rule, 372
  unbiased, 381, 388
Foreign exchange market, 183
Foreign exchange rates, 498–499
Forward discount bias, 498–499
Fractional integrated GARCH model, 320
Fragility, 9, 44
Fundamentalists
  asset model of, 292
  characteristics of, 219–220
  Chartists and, interactions between, 179–185, 323–325
  estimated fraction of, 250
  heterogeneous beliefs, 307–308
  opposite biases vs., 229–230
  price deviation, 316
  stochastic model with, 331–332
  technical analysts vs., 219–220
  trend and bias vs., 230
  trend followers vs., 227–229, 246, 307–308, 323
Fundamental price, 222–223, 288–289, 306–307
Fundamental shock, 251
Fundamental steady state, 298–300, 310
Fundamental value of stock, 66–67
Fund managers, 27
Futures, 398, 502

G
Gapped actions, 10–11
GARCH model/effects, 184, 320, 496–497
Gaussian hypothesis, 125
Gaussian model, 470–472
Gaussian random variable, 153
Generalized central limit law, 170
Generational portfolio, 397
Genetic programming, 552–553
Geographical proximity, reporting practices and, 38
Geometric decay process
  limiting, 290
  nonlinear dynamics under, 300–301
Gibbs probabilities, 224
Global dynamics
  with adaptive strategies, 534–537
  with constant strategies, 543–546
Global random attractor, 335
Glosten-Milgrom model, 113–116
Glosten-Sandas model, 135–137
Gordon dividend growth model, 307
Gordon growth model, 246

H
"Half cubic" law, 83
Harsanyii doctrine, 443
Heavy tails, limit order placement affected by, 133–135
Herd behavior, 5
  in corporate investment decisions, 37
  by forecasters, 21–24
  by stock analysts, 21–24
Herding, 5
  agency-induced, 35
  causes of, 19
  on endorsements, 24–25, 37
  exploiting of, 28
  in financial markets, 25–27
  forecasts vs., 22
  information aggregation affected by, 29
  investigative, 20–21
  in investment newsletters, 23
  in mutual funds, 27
  payoff and network externalities, 6
  psychological bias and, 17
  reputational, 6–7
  in securities trades, 24–27
  on trades, 25–27
Heterogeneity, 145, 220
Heterogeneous agents
  capital asset-pricing model with, 397
  constant relative risk aversion, 312–313
  models, 236, 241
Heterogeneous beliefs, 223–224, 244, 289–290, 307–308, 359–374, 444
  asset pricing with, 474–485
  equilibria with, 358
  perfect forecasting rules, 362–366
  stochastic models with, 330–331
Hidden orders, 80, 85, 88, 90
  identifying of, 145–146
  large, 106–108
  size variables of, 146
Higher moments, volatility clustering and dependency in, 173–174

Higher-order beliefs, 481–482
Hopf bifurcation, 230, 241, 269–270
Horse races, 515–518
H trader types, 237
Hurst exponent, 79

I
Idiosyncrasy, 9, 44
Idiosyncratic risk, 368
IID markets, 419
I-investors, 450–451, 520
Imitation, 6
Incentive contracts, 20
Incomplete markets, 432
Independent identically distributed order flows, 109–110
Index-herding behavior, 35
Indirect approach, 330
Individual beliefs, 464–465
  Bayesian inference used to deduce dynamics of, 467–469
Individual transactions, 86, 91–94
Infinite sequences, 455
Information
  liquidity effects on, 72–73
  measuring of, 70–71
Information aggregation, 29, 45
Informational efficiency, 68–69
Informational hierarchy, 5
Information blockages, 6
  information cascades as cause of, 7
  in social learning settings, 13
Information cascades, 3, 6, 11
  advantages of inducing, 28
  availability cascades, 33–34
  behavioral coarsening, 7–9
  causes of, 9
  decision making affected by, 7–8
  direct trading, 31
  exploiting of, 28
  in firm behavior, 36–39
  implications of, 9
  information aggregation affected by, 29
  information blockages caused by, 7
  investigative cascades as, 11
  quasi-, 31–32
  in rational setting, 33
  requirements for, 45
  in trading decisions, 31–32
Informed trades, 113–114
Informed trading, 70–71
Intensity of choice, 224
Inventory risk, 111
Investigative cascade, 11
Investigative herding, 20–21
Investment decisions, 36–37
Investment newsletters, 23
Investment strategy, 527–529, 532, 534, 543
Investors
  cohorts of, 375–378
  wealth dynamics of, 517

J
Joint empirical distribution, 469

K
Kaldor speculation, 483
Kalman Filtering, 451
Kelly Rule, 434, 511, 515–518, 525, 532–535, 544
  in general equilibrium, 546–547
  in genetic programming, 553
Keynes Beauty Contest metaphor, 483
Kindlebeger's theory, 163
Kirman's model, 185–190
Kullback-Leibler distance, 418
Kurtosis, 169
Kyle model, 76–77, 113, 148

L
Large type limit, 236–241
Latent liquidity, 73
Learning, in capital markets, 2–4
Leptokurtosis, 169
Leveraged buyouts, 13
Levy-stable distributions, 171
Limiting geometric decay process, 290
Limit order, 65
  heavy tails in placement of, 133–135
  limit price and, 133
  market order vs., 117–123
Limit order books
  shape of, 133–135
  statistical model of, 139–142
Limit price, 133
Linear forecasting rules, 226
Liquidity, 72–73
  asymmetric, 99, 101–103, 105
  bid–ask spread and, 111
  fluctuations in
    agent-based model for, 141–142
    volatility caused by, 127–129
  price changes and, 125–126
  self-organization of, 111
  spread dynamics after temporary crisis in, 123–125
Local dynamics
  in evolutionary model with short-lived assets, 528–529
  in evolutionary stock market model, 540–543
Log-optimum investment, 518
Log-returns, 494–495
London Stock Exchange, 65, 82, 86
Long-lived assets, 522–523

Long memory of order flows, 77–84
  causes of, 79–80
  empirical evidence for, 77–79
  exchange membership codes, 82–83
  market clearing delay as cause of, 80
  origin of, 79–80
  strategic order splitting, 80–82
  trading volume effects, 83
Long-memory processes, 78
Long-run equilibria
  characterization of, 389–390
  convergence to, 390–391
Long-term resilience, 96–98
Lucas tree economy, 432
Lyapunov exponents, 333–335, 528

M

Macrodynamic models, 340
Macroeconomics, 499–500
Mandelbrot/Fama hypothesis, 170
Many risk assets, 322–326
Market(s), 29–36, 61
  artificial, 184
  bubbles in, 34
  evolutionary models of, 510
  memes affected by, 43
  socioeconomic group dynamics in, 191–207
  stability of, 530
  structure of, 64–65
  two-asset, 328
  volatility in, 29, 75–76
Market belief. *See* Belief(s)
Market caps, 132–133
Market-clearing, 285–287, 304, 325–326, 513, 521
Market dynamics, 440
  asymmetric information and, 453–454
  rational beliefs' effect on, 476
Market ecology, 73–75
  empirical characterization of, 144–148
Market efficiency, 67–68, 230–232
Market equilibrium, 308–310
  selection and, 421–422
Market fraction model, 316–319
Market impact, 59–60
  aggregate, 86–89, 108
  diversity of, 84–90
  empirical results, 103–105
  execution strategies, 142–144
  explanations for, 69–71
  hidden order, 88, 90
  individual transactions, 86, 91–94
  large hidden order effect on, 106–108
  noise trader explanation of, 71–72
  permanent impact model, 93–94
  theory of, 90–111
  transient, 95–96
    equivalence with, 100
    mean reversion and, 96
  in upstairs market, 90
Market makers, 102, 111–112, 122
  characteristics of, 301
  market clearing, 285–287
  market clearing under, 325–326
  price and wealth behavior with, 306–314
Market making, 74–75, 119–121
  price behavior under, 298–301
Market order, 65
  informed, 113
  limit order vs., 117–123
  strategies for, 117–118
Market portfolio, 350
  aggregate generational, 386
  efficient, performance of, 391
  holdings, 385–387
  modified, 366–367
  with one risk-free asset, 349–351
Market price of risk, 352
Market selection hypothesis, 219, 405
  history of, 407–409
Market structure, 426–428
Markov process, 464, 524, 533, 541
Markov state variables, 464–474
Markov-switching multifractal model, 177
Martingales, 164–167, 173
Mean reversion, 96
Mean-variance efficiency, 352
Mean-variance investment strategy, 262
Mean-variance optimization, 510, 548–549
Mechanical impact, 150–153
Media, 16–17
Mediators, 369–371, 387
Membership code, 82–83
Meme, 4
  assemblies, 41–42
  contagion of, 39–44
  market conditions' effect on, 43
  reproduction of, 41
  spread of, 46
Memetic approach, 40–41
Mere exposure effect, 33
Model building, 63–64
  phenomenological approach to, 64
Modified market portfolio, 366–367, 397
Monotonic convergence, 259
Moves
  exogenous, 12–13
  timing of, 12–13
Moving average process, nonlinear dynamics under, 300
MRR model, 94–95, 99, 112
  with bid–ask spread, 116–117, 155–156

Multifund separation theorem, 377
Multinomial logit model, 224
Multiperiod planning horizons, 374–387
Multiple risky assets, 322–330, 359
Multiple survivors, 423–426
Mutual funds
  herding in, 27
  theorem regarding, 351–355
Myopic-investor economies, 447

N
Naive expectations, 257–258
Natural selection, 409
Nearest neighbor forecasting, 232
Necessary conditions for survival, 421
Network externalities, 6
News media, 16
Noise, 287
Noise traders, 70–71, 360, 364, 409
  laws of large numbers for, 428–431
  life and death of, 426–431
  market structure effects, 426–428
  survival of, 426
Noisy rational expectations asset-pricing theory, 440
Nominal GNP, 452
Nonergodic asset prices, 387–397
Non-EU traders, 432–433
Nonfundamental steady state, 310
Nonlinear deterministic dynamics, 313
Nonlinear dynamics
  under geometric decay process, 300–301
  under moving average process, 300
Nonmechanical impact, 150–153
Non-Normality, 169
Nonstationary economy, 457–461
Nonsystematic risk, 366–369
Nonvanishing effects, 444
NYSE, 64–65, 82, 121

O
Observational influence, 5
Off-book market, 65
Opinion formation model, 185–190
Optimal demand, 288–289
Optimal portfolio, constant relative risk aversion and, 302
Order(s), 64
  effective market, 65
  limit. *See* Limit order
  market. *See* Market order
Order books
  mechanical impact for, 150–152
  shape of, 133–135
  statistical model of, 139–142
Order flows, 99–101
  independent identically distributed, 109–110
  long memory of. *See* Long memory of order flows
  predictable, 99
  statistical models of, 137–142
Overconfidence, 480–481

P
Paradoxicality, 9, 45
Pareto optimality, 413–414, 501–502
Pareto optimal market, 67
Past actions, 10–15
  consequences of, 15–16
  with noise, 12
Path dependence, 9
Payoff(s)
  decision making about, 17
  externalities, 6, 20
  homogeneous vs. heterogeneous, 14
  stochastic nature of, 15
Payoff interaction hierarchy, 6
P-bifurcation, 336–338
Perfect forecasting rules, 346, 348, 397
  for first moments, 363, 379–381
  foresight, 226
  heterogeneous beliefs, 362–366
  multiperiod planning horizons, 379
  for second moments, 363–364, 381–385
Performance measures, 290–291, 370
  empirical returns as, 392–393
  empirical Sharpe ratios as, 393–394
Period-doubling bifurcation, 268–269
Persuasion bias, 35
Phenomenological approach, 64
Pitchfork bifurcation, 239–240, 270–271
Planning horizons, 374–387
Portfolio, 350
  aggregate generational, 386
  efficient, performance of, 391
  with one risk-free asset, 349–351
  wealth and, 524
Portfolio consumption, 560
Portfolio holdings, 385–387
Portfolio optimization, 283–284
  of many risky assets, 322–326
Positive payoff externalities, 20
Power-law behavior, 319–321
Power laws, 178
Predecessors payoffs, 11
Predictable order flow, 99
Prediction strategies, 259–262
Price behavior
  constant relative risk aversion, 302–314
  under market-maker mechanism, 298–301
  with market makers, 306–314
  under Walrasian auctioneer mechanism, 291–298

Price changes, liquidity affected by, 125–126
Price-dependent strategies, 556
Price–dividend ratios, 494–497
Price dynamics, constant absolute risk aversion utility function, 288–302
Price generation, 256–257
Price impact, 84
Price-to-cash flows, 242–244
Profitability, 262–263
Pro forma earnings, 38, 44
Psychological bias, 17–18, 33, 40, 45, 243
Public information disclosure, 8
Publicly observable state variable, 13
Pullback process, 335

Q
Quasi-cascades, 31–32

R
Random attractors, 334–336
Random dynamical system, 332, 512
Random dynamic systems theory, 528
Randomness, 390, 519
Random walk hypothesis, 128
Rational agent, 460
Rational belief equilibrium, 499
Rational bubble solutions, 223
Rational diverse beliefs, 457–458, 487
  aggregate dynamics and, 456
  general theory of, 454–485
  model of, 493–494
  overview of, 440–441
  volatility and, 455–457
Rational expectations, 149, 243, 257, 287, 347–348, 379, 454–455, 501
  convergence to, 444
  dynamic stability with, 371–374
Rational expectations equilibrium, 407, 440, 446, 450, 485, 498
Rational expectations traders, 406
Rationality, for Gaussian model, 470–472
Rational learning, 3
  implications of, 9
  principles of, 7–9
  psychological bias effects, 33
Rational observational learning, 5–6, 44
Rational overconfidence, 480–481
Rational route to randonmess, 229
Realized orders, 85
Realized spread, 117
Reference portfolio, 366
Reinforcement learning, 220, 224
Relative asset payoff, 525, 542
Relative asset prices, 558–559
Relative performance, 38
Repeated trading, 449
Replicators, 4
Reporting practices, 38–39
Representation bias, 188
Reputation, 18–20, 22, 45
Reputational herding, 6–7
Resale value, 524
Residential housing, 39–40
Resilience, long-term, 96–98
Return behavior, 305–306
Returns
  distributional properties of, 167
  fat tails of, 167–172
  leptokurtosis of, 169
  non-Normality, 169
  predictability of, 494–496
Revealed liquidity, 73
Risk-adjusted profit, 225
Risk-free assets, 326
  portfolio with, 349–351
Risky assets, 322–330, 359, 509
Robustness, 431–435
Runs, financial market, 27–28

S
Saddle-node bifurcation, 268
Santa Fe artificial stock market, 184
Scaling laws, 162
  in natural science, 163
  stylized facts as, 175–178
Second moments, perfect forecasting rules for, 363–364, 381–385
Securities, 398
Security analysis, investigative herding, 20–21
Selection, 416
  in complete IID markets, 419
  equations, 420
  evolutionary, 514
  example of, 416–417
  literature regarding, 416
  market equilibrium and, 421–422
  necessary condition for survival, 421
  over non-EU traders, 432–433
  over rules, 434–435
Sentiment factors, 164
Sharpe ratio, 369
  as performance measures, 393–394
Short-lived assets, 519, 522, 563
  evolutionary model with, 524–537
Short-run profits, 232
Short-term market impact, 69–70
Signals, 11
Simple rule, 435

Simulations, 440, 485–486, 548–552
Simultaneity, 9, 45
Single actions, 14
Smith, Adam, 61
Social interactions
  asset-pricing model with, 197
  framework of, 191–197
Social learning, 2, 13, 406–407
S&P 500
  autocorrelations of returns, 319
  stylized factors in, 314–316
Speculation, 180, 483–485, 502
  Kaldor, 483
  Kirman's model of, 185–190
Spread
  bid–ask. *See* Bid–ask spread
  economics of, 111–114
  liquidity crisis effects on, 123–125
  realized, 117
  volatility vs., 129–132
Spurious agglomeration, 37
Stability, 530
Stabilization policy, 502
State of belief, 465
State prices, 414
Stationary forecasts, 441
Statistical efficiency, 99
Statistical physics, 175
Statistical stability, 457–458
Steady state
  fundamental, 298–300, 310
  market-making strategy, 119–120
  nonfundamental, 310
Stealth trading, 90
Stimulated refill, 102
Stochastic bifurcations, 332–338
Stochastic experiments, 306
Stochastic models
  with fundamentalists, 331–332
  with heterogeneous beliefs, 330–331
Stochastic processes, 348, 422–423
Stock, 486–487
  fundamental value of, 66–67
Stock analysts
  dispersing by, 23
  herd behavior by, 21–24
  reputation concerns of, 22
Stock market "gurus," 25
Stock market model, evolutionary, 537–547
Stock returns, 494–496
Strategic order splitting, 80–82
Stylized factors
  fat tails of asset returns, 167–172
  in S&P 500, 314–316
  volatility clustering and dependency in higher moments, 173–174
Stylized facts, 162
  Martingales, 164–167
  as scaling laws, 175–178
Subjective expected utility, 405, 410
Supply and demand, fluctuations in, 69
Survival index, 405, 423
Survivors, 423–426
Switching, 290–291
Systematic risk, 366–369

T

Tail index, 170–171
Takeover markets, 37
Tâtonnement, 59, 61, 66, 96
Technical analysts, 219–220
Technical trading rule, 372, 388
Temporary equilibrium
  heterogeneous beliefs, 359–362
  multiperiod planning horizons, 378–379
Temporary equilibrium map, 347–348, 359
Time-series switching model, 242
Traders, 411–412
  noise. *See* Noise traders
  types of, 236–241
Trades
  herding on, 25–27
  informed, 113–114
Trading packages, 88
Trading volume, 502
  long memory of order flow affected by, 83
Transient impact, 95–96
  equivalence with, 100
  mean reversion and, 96
  model of, 110–111
Transversality condition, 222
Trend extrapolation rule, 258
Trend followers, 227–229, 246, 249
  asset model of, 292
  heterogeneous beliefs, 307–308
Trueman model, 19
Two-asset market, 328
Two-fund separation theorem, 351–355
Two-period equilibrium model, 349–359
Type I stocks, 142

U

Unbiased forecasting rule, 381, 388
Unbounded economies, 431
Unconverted interest parity, 188
Uninformed trading, 70–71

Universal preasymptotic behavior, 172
Upstairs market, 65, 90

V
Value investment, 511
Variance ratio test, 494
Vicarious learning, 15
Volatility, 75–76
  anatomy of, 486–498
  foreign exchange rates, 498–499
  liquidity fluctuations as cause of, 127–129
  rational diversity and, 455–457
  short-time, 96
  spread vs., 129–132
  volume fluctuations as cause of, 127–129
Volatility clustering, 173–174, 183, 320
Volatility moments, 491–492
Volatility tests, 280
Volume
  at best prices, 135–137
  fluctuations
    accounting for, 153–155
    volatility caused by, 127–129

W
Walrasian auctioneer, 285–286, 289
  characteristics of, 301
  price behavior under, 291–298, 303–306
  wealth behavior with, 303–306
Walras's Law, 522, 539
Wealth
  accumulation of, 232–235
  portfolio and, 524
Wealth behavior with market makers, 306–314
Wealth dynamics, implied by constant relative risk aversion, 302–314
Wealth share weighted average, 417
Wiener process, 332

Z
Zero-intelligence models, 137–139, 141