Microscopic Simulation of Financial Markets

From Investor Behavior to Market Phenomena

MOSHE LEVY
HAIM LEVY
SORIN SOLOMON

PREFACE

The classical models in finance are analytical models. These models make assumptions regarding the market and the behavior of individuals operating in the market and derive their results by mathematical analysis. For example, the Capital Asset Pricing Model (CAPM) assumes that investors maximize their von Neuman and Morgenstern (1944) expected utility, in addition, it makes specific assumptions of a perfect market with no taxes and no transaction costs, normal rate of return distributions, a one-period investment horizon, and homogeneous expectations.

Many of the assumptions made in the CAPM, as well as in most other models in finance, are admittedly false. First, there are experimental findings that cast doubt on the expected utility paradigm, which is the foundation of most models in finance. Second, there are also many model-specific assumptions that have been criticized. For instance, the assumption of no taxes and no transaction costs does not conform to the actual facts. It is also clear that, in contrast to the homogeneous expectation assumption, investors differ in their expectations, holding periods, decision-making processes, and so on. Although false, these assumptions are required to obtain analytical tractability. Thus, most of the cornerstone models in finance, such as the CAPM, the Arbitrage Pricing Theory, the

Black and Scholes Option Pricing Model, and the Modigliani and Miller Capital Structure and Valuation Model, are based on underlying assumptions that are, at best, problematic.

There are several arguments in defense of models based on unrealistic assumptions. The first is that a model with unrealistic assumptions is better than no model at all, and one should not reject a model unless a better one is found (see Stigler, 1966). In addition, although some of a model's assumptions may not hold in reality, it is possible that the model still provides realistic results. According to Friedman (1953a), a model's quality should be measured by the model's explanatory power, not by its (possibly unrealistic) assumptions. For example, even though investors do not sit down to calculate their expected utilities, the market may behave "as if" investors were expected utility maximizers. Unfortunately, the empirical results that test the various theoretical models in finance are, at best, controversial. Moreover, there are several anomalies (e.g., the January effect, the small-firm effect) that contradict the prediction of theoretical models. Thus, in the case of most finance models, the empirical findings do not support the "as if" argument.

The expected utility models mentioned previously are *normative*. Given a set of assumptions or axioms, the model tells us how investors *should behave* and how the end result (e.g., the pricing of an asset) is determined. In contrast, experimental studies analyze how investors, or more precisely laboratory subjects, actually *do behave* and make no normative claims regarding how investors should behave. The leading candidate to compete with the expected utility paradigm is Prospect Theory, advocated by Kahneman and Tversky (1979). Kahneman and Tversky conducted a series of experiments revealing that decision-making behavior is in sharp contradiction to the predictions of expected utility theory. In particular, the subjects make choices according to *change* in wealth rather than *total* wealth and employ *decision weights* (which do not obey the probability rules) rather than objective probabilities.

Thus, expected utility models in finance have been attacked on the following fronts:

1. Some of the model-specific assumptions that are commonly made (no taxes, homogeneous expectations, etc.) are unrealistic.
2. Even if the model-specific assumptions are intact (or if we justify these assumptions by the "as if" argument), experimental studies, particularly Prospect Theory, cast doubt on the validity of the expected utility paradigm, which is the foundation of these models.
3. Empirical studies either do not support or only weakly support the various models in finance.

Prospect Theory and other descriptive models of investor behavior attempt to capture some systematic elements of investor behavior. Unfortunately, there does not seem to be a single theory that makes sense of the mixed empirical and experimental data (see Davis and Holt, 1993). In addition, the mathematics of these descriptive models is typically cumbersome, and therefore the implications of these models for pricing are difficult to analyze.

It seems that theoretical research in finance may have a problem: on the one hand, unrealistic simplifying assumptions are needed to ensure analytic tractability; on the other hand, these assumptions lead to results that cannot be convincingly supported by the empirical evidence. Perhaps we have been searching for the lost coin only under the lamppost.

In this book we suggest a research methodology that may expand the realm of investigation in finance. This methodology is called Microscopic Simulation (MS). MS is a methodology that was developed in the physical sciences as a tool for the study of complex systems with many interacting "microscopic" elements. Such complex systems generally do not yield to analytical treatment. The main idea of the MS methodology is to study complex systems by representing each of the microscopic elements individually on a computer and simulating the behavior of the entire system, keeping track of all of the elements and their interactions in each time period. Throughout the simulation, global, or "macroscopic," variables that are of interest can be recorded, and their dynamics can be investigated. For example, in physics, in order to study and predict the behavior of a nuclear reactor, the multitude of atomic particles interacting in the reactor can be simulated, and the dynamics of the system's temperature, pressure, and so on, can be investigated. The main advantage of the MS method is that, unlike analytical methods, it does not force one to make simplifying assumptions for the sake of tractability. Thus, virtually any system with heterogeneous elements and complicated interactions can be investigated.

The idea of studying complex systems by simulating all of the systems' microscopic elements is very natural, and as such it has been reinvented in various fields of science. As a consequence, this methodology is known by several different names, such as microscopic simulation, microsimulation, and agent-based simulation. Although there may be nuances differentiating these various methods, the basic idea behind all of them is the same.[1]

[1] Microscopic simulations are sometimes confused with Monte Carlo simulations. Although these methods are related, they are not the same. Monte Carlo simulations refer to any computer simulations involving random number generation, whether a microscopic simulation of a system with many elements or only a computer program designed to investigate the statistical properties of a probability distribution or a stochastic equation. Microscopic simulation, in contrast, always refers to simulating a system with many interacting "microscopic" elements. Such a simulation may involve randomness (in which case it is a special case of the Monte Carlo method), or it may be deterministic (see, for example, Chapter 11).

We believe that MS has a great potential as a research tool in finance and in economics. There is no doubt that financial markets, in which a multitude of heterogeneous quasi-rational investors operate, are very complex systems. MS allows the modeling and investigation of such systems without unrealistic simplifying assumptions. Indeed, the unrealistic assumptions can be relaxed one by one, and the effect of each simplifying assumption on the results can be investigated.

Although MS is a standard tool in most natural sciences, it is not yet commonly employed in the social sciences. Indeed, physicists who realize that the powerful MS machinery they possess can solve problems in social science are starting to conduct microscopic simulations of social science systems. The purpose of this book is to present the MS methodology to the finance and economics community and to suggest ways in which this methodology can be implemented in these areas.

We believe that MS holds promise in two main research avenues:

1. *Extension of existing models*. With MS, one can extend the existing analytical cornerstone models in finance and investigate the robustness of these models. Namely, one can study the effect of relaxing each of the models' assumptions on the results. Do the results of the model *approximately* hold when unrealistic simplifying assumptions are replaced with more realistic assumptions? Which of the important models are robust, and which break down with a slight change of the underlying assumptions?

2. *New models*. MS allows the researcher to explore new models without having to make restricting assumptions. In particular, one can model markets as realistically as desired. For instance, if the experimental and empirical evidence suggests that the behavior of 60% of the investors is best described by expected utility maximization, whereas for 20% of the investors Prospect Theory provides a better description, and for 20% of the investors noise or liquidity trading seems to be the best explanation, these findings can easily be incorporated into an MS model. Various fundamentalist and technical trading strategies which are observed in experiments and in actual markets can also all be modeled with MS in a realistic setting where transaction costs and taxes prevail. Finally, while classical models typically deal with static equilibrium in a market of homogeneous rational agents, with MS one can investigate dynamic models, models in which investors are heterogeneous in many respects, and models in which investors are constantly learning and changing strategies as time goes by.

This book is organized as follows. Chapter 1 reviews the main models in finance and some of their problematic underlying assumptions. Chapters 2 through 4 review the main experimental findings regarding investors' behavior. These behavioral elements are later incorporated into MS models, and their effects on pricing and on market dynamics are investigated.

In Chapters 5 and 6 we present the MS method and review some of its applications in various fields of science. Chapters 7 through 9 discuss several MS models in finance. Chapters 10 and 11 analyze the robustness of the CAPM and the Black and Scholes Option Pricing Model when some of the models' problematic underlying assumptions are relaxed.

ACKNOWLEDGMENTS

We thank Harry Markowitz for illuminating discussions on the subject of this book. We are grateful to Dietrich Stauffer for his extensive comments on the manuscript and for his encouragement. We thank Thomas Lux for sharing his results of detailed and comprehensive comparisons of various market simulations. We are grateful to Phil Anderson for his insights. We also thank Golan Benita, Allon Cohen, and Aaron and Chava Rothenberg for their technical help, and Hyla Berkowitz for her excellent typing and numerous advice on how to improve the manuscript. The encouragement of the editor, Scott Bentley, is greatly appreciated.

Moshe Levy
Haim Levy
Sorin Solomon

1

CLASSIC MODELS IN FINANCE

Solved and Unsolved Issues

1.1. INTRODUCTION

The classic models in finance, and in particular the studies that constitute the pillars of this relatively young field of research, are mainly analytical. These analytical models deal with investment decision making under uncertainty and, in particular, with the prevailed complex capital market composed of many risky assets. To achieve analytical results, it is necessary to assume a decision framework (e.g., expected utility framework) and to make many specific assumptions, some of which are very unrealistic. The fact that some of the assumptions are very unrealistic was not concealed from the researchers who made them; however, these assumptions were made for the sake of tractability in order to obtain analytical results.

Let us illustrate here with the Sharpe-Lintner Capital Asset Pricing Model (CAPM). The CAPM deals with rational investors who maximize expected utility. To obtain the CAPM, one needs to make many assumptions such as: no taxes, homogeneous expectations regarding future distribution of returns (which are assumed to be normal), no transaction costs, that all investors will have the same holding period, and so on. Obviously, these assumptions do not conform with the actual facts—in particular, taxes and transaction costs do exist. Thus, one may claim that the CAPM is unrealistic because the assumptions are unrealistic. It is also possible that these assumptions are intact, yet investors are irrational.

How crucial are these assumptions and what effect do they have on the derived model? What will be the effect on the theoretical results if one relaxes one or more of these assumptions? To neutralize the effect of these assumptions, and to test expected utility theory or portfolio theory even in the ideal case where such assumptions *are* intact, one can conduct experimental studies. Thus, one can test the investor's rationality when all the assumptions are intact. Investment laboratory experiments can be conducted such that all the crucial assumptions mentioned hold. For example, there have been experiments in which the subjects participating had to choose from several securities and were told that there were no transaction costs, no taxes, and that the returns were drawn randomly from normal distributions with known parameters. By conducting the test in such a way, one can test the other ingredients of portfolio theory, in particular the expected utility theory (EUT) models. For example, by Markowitz's (1952a) portfolio theory, which is the foundation of the CAPM, we expect that when one changes the correlation between two stocks, say, from $+0.8$ to -0.8 (with no change in the parameters of the other stocks), generally a higher proportion of the subject's wealth will be allocated to these two stocks. Also, when returns are drawn *randomly* from a given distribution with parameters that are known to the subjects, the historical rates of return are irrelevant for future decision making. The behavior found in laboratory experiments contradicts EUT: it was found that investors did not increase their investment proportion in stocks when the correlation between these two stocks was changed from $+0.8$ to -0.8, and that investors insist on asking for information on historical rates of return even though they are irrelevant (see Kroll, Levy, and Rappoport, 1988a, 1988b). There are many more experimental findings rejecting EUT even in artificially simplified settings.[1] The deviation of investors from rationality induces deviations in market prices from what theoretical models predict. Indeed, Arrow (1982) asserts: "I hope to have made a case for the proposition that an important class of international markets shows systematic deviation from individual rational behavior" (see Arrow, 1982, p. 8).[2]

Most experimental studies do not involve securities. The subjects were asked to make some hypothetical choices from which it was concluded that a substantial group of them do *not* behave as EUT asserts. They behave

[1] For a collection of experimental studies dealing with choices, bargaining, and valuation of preference, see Allvin E. Roth (1987).

[2] See also Summers (1986).

irrationally, distort probabilities subjectively (but in some systematic way), and make decisions based on *changes* in wealth rather than *total* wealth. Kahneman and Tversky (1979), who conducted many of these experiments, suggest prospect theory (PT), which describes how investors behave in a laboratory environment, pinpointing the various contradictions to EUT.

Many experiments follow Kahnemann and Tversky's 1979 PT paper. Most studies support PT, confirming the basic observations of Kahneman and Tversky.[3] Yet a few studies reject the PT or some elements of it.[4] As mentioned previously, most PT studies are experimental studies. Yet, in a few exceptions the studies rely on empirical market data: for example, Fiegenbaum and Thomas (1988) and Fiegenbaum (1990) employ the COMPUSTAT data to confirm that risk-seeking behavior exists below a certain level of rate of return. D'Aveni (1989) uses data of bankrupt firms, finding that managers behave according to the predictions of PT, and Benartzi and Thaler (1995) use PT to explain the observed risk-premium in the market. Thus, there are attempts to use market data or to explain market phenomena like the risk premium by using PT.

Prospect theory casts doubt on the EUT, which is the foundation of most theoretical models in finance and economics. In addition, people do not react to changes in the various parameters as EUT predicts. Thus, the assumption that individuals act according to EUT is very problematic. On top of this, many of the specific assumptions underlying classical models are also unrealistic. According to Friedman (1953a), models should not be judged by their assumptions, but by their predictive power. Unfortunately, the empirical evidence either rejects the various theoretical models in finance or supports them only weakly. Thus, the classic models in finance and economics are attacked on several fronts.

There are no analytical tools to measure the effect of various deviations from the EUT assumption and from model-specific assumptions on the results and, in particular, on asset price determination and on the risk-return relationship. It is possible that the fact that some of the assumptions regarding investors' behavior that underlie a given model are

[3] See Schoemaker and Kunreuther (1979), Payne, Laughhunn, and Crum 1984, Arkes and Blumer (1985), Uecker, Schepanski, and Shin (1985), Gregory (1986), Haka, Friedman, and Jones (1986), Tversky and Kahneman (1986), Chang, Nichols, and Schltz (1987), Budescu and Weiss (1987), Loewenstein (1988), Diamond (1988), Fiegenbaum and Thomas (1988), Qualls and Puto (1989), Bromley (1989), D'Aveni (1989), Fiegenbaum (1990), Meyer and Assuncao (1990), Kameda and Davis (1990), Garland and Newport (1991), Jegers (1991), Kanto, Rosenqvist, and Survas (1992), Shefrin and Statmen (1993), Whyte (1993), Hardie, Johnson, and Fader (1993), Drake and Freedman (1993), Salminen (1994), Hirst, Joyce, and Schadewald (1994), Schaubroeck and Davis (1994), Levy (1994), Benartzi and Thaler (1995), Mayer (1995), Leclerc, Schmitt, and Dube (1995), and Sebora and Cornwall (1995).

[4] The list of papers rejecting some of the PT elements is much shorter: Hershey and Shoemaker (1980), Martinez-Vazquez, Harwood, and Larkins (1992), Brockner (1992) and Casey (1994).

not so crucial, and the fact that they do not hold has only a minor impact on the model's results. Alternatively, it is possible that even a small deviation from the assumptions may completely reverse the theoretical results. Since it is hard, if not impossible, to *analytically* test the effects of various deviations from the model's assumptions on asset prices, in this book we suggest a different approach to investigate these effects.

We suggest employing the microscopic simulation (MS) methodology to analyze the impact of various deviations from EUT and other unrealistic model-specific assumptions on the theoretical results. With the MS methodology one can allow for heterogeneous expectations, various heterogeneous investment holding periods, and even a violation of EUT in favor of prospect theory or other experimentally observed behavior patterns. With MS one can model investors as being quasi-rational or bounded-rational (see Russell and Thaler, 1994, and Black, 1986): they maximize some expected utility or value function but they may make errors in their decision making. The MS methodology allows us to relax one or more of the model's assumptions and to examine the effects of this relaxation on the results. MS allows us to obtain results that are impossible to obtain analytically.

1.2. EUT, ALTERNATIVE MODELS, AND NOISE TRADERS

One of the most famous violations of EUT is the Allais paradox (1953) (see Chapter 2). There are a variety of modifications of EUT that may explain this paradox. Machina (1982) proposes a generalized expected utility theory, Kahneman and Tversky suggest prospect theory, Chew and MacCrimmon (1979) develop weighted utility theory, Bell (1982), Fishburn (1982), and Loomes and Sugden (1982) suggest the regret theory, and Quiggin (1982) proposes the rank dependent utility theory. The states of all these theories is best summarized by Davis and Holt (1993):

> For a variety of reasons, none of these alternatives has become widely accepted. First, the pattern of violations of expected utility theory is not as clear as was once believed, and no rival theory persuasively organizes the somewhat mixed data. Second, their application involves complicated specifications of utility functions, which theorists find rather cumbersome. (p. 448)

Thus, it is not obvious that any of the other proposed theories are better than EUT, in particular because they do not provoke simple testable hypotheses or even do not suggest an equilibrium price determination model. We do not reject a theory unless there is a better one. Therefore we do not reject the EUT. This idea was summarized as early as 1966 by George Stigler:

> When we assume that consumers, acting with mathematical consistency, maximize utility, . . . it is not proper to complain that men are much more complicated and

> diverse than that. So they are, but, if this assumption yields a theory of behavior which agrees tolerably well with the facts, it must be used until a better theory comes along. (Stigler, 1966, p. 6)

Thus, it seems that the correct approach is not to reject EUT but to improve it, either by relaxing some of the assumptions or by adding some components to the decision-making process (e.g. some sort of irrationality). Indeed, this is the approach adopted in this book.

Generally, the investment decision process can be classified into two extreme regimes: one asserting that all investors maximize some utility function and act exactly as implied by EUT. The other is the complete opposite—a chaos in which investors buy or sell stock randomly, completely unrelated to the asset's fundamental value. Nowadays, more and more economists believe in *quasi-rationality* or *bounded-rationality*, which lies between these two extremes scenarios. It is believed that quasi-rationality best describes how individuals in financial markets operate.[5] In the quasi-rationality framework, investors still maximize some utility or value function (act according to some model) but may have some degree of irrationality. For example, they can act on wrong signals or incorrect information, rely on technical analysis that is not related to the fundamental value of the asset, and so on. Such investors, who do not conform with EUT, are sometimes referred to as "noise traders." Friedman (1953b) argues against the importance of noise traders. According to his view, irrational investors are met in the market with rational arbitrageurs who trade against them, and in the process the rational investors drive the asset price close to its fundamental value. Moreover, the irrational investors, according to Friedman's view, lose money to the rational investors and so eventually disappear from the market.

DeLong, Shleifer, Summers, and Waldmann (1990) show that this may not be the case. Arbitrage does not eliminate the effect of noise traders. Moreover, noise traders who are bullish earn a higher expected return than the sophisticated rational investors who trade with them. The noise traders also affect the volatility of prices and certainly do not disappear from the market. DeLong *et al.* assert that noise traders select their portfolio on the basis of incorrect information, such as information from technical analysis, economic consultants, or stockbrokers. The irrational investors believe that these signals carry valuable information when in practice they do not. To achieve analytical results, Delong *et al.* suggest a very simple model in which there are only two assets, one risky and one safe. All investors have a constant absolute risk aversion utility function. The noise traders misperceive the expected return of the risky asset. Specifically, investor *i*'s error is denoted by ρ_i (which can be either positive

[5] For the importance of noise traders and quasi rationality to explain market phenomena, see Black (1986) and Russell and Thaler (1994).

or negative). DeLong *et al.* assume that

$$\rho_i \sim N\left(\rho^*, \sigma_\rho^2\right)$$

where ρ^* measures the average "bullishness" of the noise traders and σ_ρ^2 is the variance of the noise traders' misperceptions of the expected return of the risky asset. DeLong *et al.* show that when $\rho^* > 0$, noise traders may receive a higher average return than sophisticated investors. However, when $\rho^* < 0$, noise traders receive both lower realized return and lower utility levels. The sophisticated investors' utility level increases with the existence of noise traders: it allows them to trade in both the safe and unsafe asset, otherwise they would trade only in the safe asset. Thus, the increase in opportunities increases the utility level.

The study of DeLong *et al.* is one of the few attempts to analytically discuss the classic models in finance and economics when some of the assumptions are relaxed, making the models more realistic.[6] In this specific study, some degree of irrationality, or acting based on wrong information, is introduced.

Several other attempts have been made to examine the effects of relaxing some of the typical problematic underlying assumptions. For example, Lintner (1965) relaxes the homogeneity of expectations assumption. Levy (1978), Mayshar (1979), Merton (1987), Markowitz (1990), and Sharpe (1991) relax the perfect market assumption, allowing investors to hold only a few assets in their portfolios. Brennan (1970) derives the risk-return relationship when personal taxes prevail. Even if one of the classic model's assumptions is relaxed, the analytical results become very complicated, and sometimes more specific assumptions are needed to obtain meaningful and clear-cut results. For example, in the DeLong *et al.* study some assumptions, such as constant absolute risk aversion and normal distributions, are needed to derive the analytical results. One may ask whether the same results are obtained if ρ_t is not normally distributed (e.g. negatively or positively skewed)? What if the utility function would be with constant *relative* risk aversion (as suggested by the experimental evidence) rather than constant *absolute* risk aversion as assumed by DeLong *et al.*? Each of these changes needs another analytical model, which may be unsolvable, and if solvable, would probably yield other results that may differ from the results obtained by DeLong *et al.*

The dilemma researchers in finance face is how to adjust the EUT models in order to make them as realistic as possible and yet to keep the

[6] A recent stream of "behavioral finance" papers follows this research avenue. See, for example, Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyan (1999), and Hong and Stein (1999).

modified models analytically solvable. The more the models are adjusted to real-life situations, the less solvable they are. This dilemma is not a simple one, and indeed only a few meaningful analytical models relax the severe assumptions of perfect rationality, a perfect market, homogeneous expectations, and no transaction costs. It is very difficult to obtain analytical results with a relaxation of one of these assumptions, let alone with a relaxation of several or all of them simultaneously.

In this book we employ a research tool known as microscopic simulation (MS) in order to investigate financial markets. MS is based on computer simulations, and it is therefore not confined by the typical restrictive assumptions needed to ensure analytic tractability. For example, in the Levy, Levy, and Solomon (1994, 1995) microscopic simulation model, which is described in Chapters 7 and 9, we assume that investors generally maximize expected utility; however, some degree of irrationality is allowed. For example, the investor may deviate to some extent from the optimal investment policy, consider changes in wealth rather than total wealth, distort probability distributions, and so forth. We also allow noise trading. We allow investors with heterogeneous expectations—that is, one investor may form his estimate of the future rate of return distribution by looking at the last 10 rates of return, while another investor may base her estimation by looking at the last 20 observations. We also allow diverse holding periods (e.g., one investor may adjust the portfolio every month and one every two months). All these changes taken separately or simultaneously are impossible to solve analytically. However, they can be analyzed by MS. The MS methodology allows us to analyze which of all the deviations from EUT models and specific assumptions underlying each model are important in price determination and which have only a marginal effect. In addition, MS not only determines equilibrium prices of assets under conditions of uncertainty, it also provides a dynamic price behavior with possible trends, booms, and crashes. Indeed, in Chapter 7 we employ MS to study stock price dynamics in a market with one risky asset (e.g., S & P index) and one riskless asset. In particular, we analyze the effect of heterogeneity of investors (in their expectations as well as in their preferences) on price dynamics. We find that even if only a small minority of investors employ the *ex post* return distribution in order to estimate the *ex ante* distribution (as found experimentally), the model generates many of the empirically documented "anomalies" such as excess volatility, short-term momentum, longer-term mean reversion, and the positive correlation between volume and contemporaneous and lagged returns. We also analyze the conditions under which booms and crashes in the stock market are unavoidable. In Chapter 9 we employ MS to analyze the effect of the experimental findings of PT on asset pricing and on market dynamics.

1.3. CLASSICAL ANALYSIS IN MODERN FINANCE

The investment decision and the related models can be classified into four main categories: arbitrage pricing models, risk-return expected utility models, mean-variance and stochastic dominance investment efficiency analysis, and decision making by the firm (dividend policy, investment decision, capital structure, etc.). In this section we elaborate on each category, emphasizing the confining assumptions needed to conduct the classical analytical treatment. Then, we discuss the role of MS when some of the assumptions are relaxed. We briefly review the various assumptions needed in some of the main models in finance, but we focus on the Black and Scholes option pricing model and Sharpe-Lintner CAPM, because these two models, no doubt, constitute cornerstones in modern finance.

1.3.1. Arbitrage Pricing Models

The main arbitrage models in finance are the Modigliani and Miller (1958, 1969) capital structure and valuation proposition, Ross's arbitrage pricing theory (APT) and the Black and Scholes's option pricing model. These models do not rely on EUT. The common assumption to all these no-arbitrage models is that investors are nonsatiable: they prefer more to less wealth, an assumption that even those who question EUT would not argue with. In particular, using prospect theory rather than EUT does not change the no-arbitrage models' results, which are intact also for a value function that depends on change of wealth rather than on total wealth, as long as the value function is nondecreasing in the change of wealth.

It is interesting to note that experimental studies reveal results of the following sort: "80% of the subjects choose A and 20% choose B" (see, for example, Kahneman and Tversky, 1979). Thus, even if 80% choose irrationally or ignore arbitrage opportunities available in the market, it is enough to have 20% or for that matter even one investor who chooses rationally in order to derive the arbitrage equation. For example if $V_U > V_L$ (when V_U and V_L are the values of the unlevered and levered firms, respectively; see Modigliani and Miller, 1958) this one rational investor would short an infinite amount of the stock of the unlevered firm and buy the stock of the levered firm and by doing so would create an arbitrage *money machine*. This process will continue until $V_U = V_L$. The crucial point is that to achieve the arbitrage results there is no need for all investors to be rational; it is enough that there is one rational investor who sees the available arbitrage opportunity in order to derive the equilibrium results. Thus, the existence of irrational investors or noise traders does not affect the equilibrium no-arbitrage results.

This discussion also holds with respect to Ross's APT model and Black and Scholes's option pricing model. Ross shows that it is possible to create a zero-beta, zero-investment portfolio. If this portfolio yields a positive return, we have a money machine and it is enough to have one investor who realizes this opportunity to ensure no arbitrage. From a no-arbitrage argument, Ross derives the risk-return equilibrium linear relationship.

Black and Scholes derive the equilibrium price of an option. Once again, if the market price of the option is different from the model value, a zero-variance portfolio that yields a return higher than the riskless interest rate can be created, implying that a money machine exists.

The experimental findings and, in particular, the violations of EUT as advocated by PT have no effect on the no-arbitrage models. Even the experimentalists do not argue with the well-accepted hypothesis that if an investor has more wealth he or she must be better off. Moreover, they even require that first-degree stochastic dominance (FSD) (see Levy, 1992) will not be violated, which is much stronger assumption than assuming nonsatiable investors (see Tversky and Kahneman, 1992). Thus, the fact that the value function $V(x)$ suggested by PT is nondecreasing ($V^1(x) > 0$ for all $x \neq 0$, where x is change in wealth) implies that all the arbitrage models are intact also in the PT framework.

Though experimental results do not contradict the no-arbitrage profit models, these models may have no explanatory power for different reasons: the specific assumptions required to derive these models are unrealistic. For example, to derive Modigliani and Miller's results one has to assume that there is a perfect market, that firms and individuals can borrow and lend at the same interest rate, that there are no transactions costs, and that bankruptcy is costless. To derive Ross's APT one needs to assume a specific return generating process, no transactions costs, that unlimited short sales are allowed when the proceeds of the short sales are received by the short seller, and a given holding period common to all investors. To derive Black and Scholes' option pricing one has to assume a specific return generating process for the underlying asset (Brownian motion), no transactions costs, that short sales are allowed, a perfect market with a borrowing interest rate that is equal to the lending rate, and, in particular, that the standard deviation of the underlying asset's returns is known and agreed on by all investors.

A relaxation of each of these unrealistic assumptions may either change the model equilibrium results only slightly or change them very drastically. It is hard, if not impossible, to figure out analytically the effect of the relaxation of each of these assumptions on the model's equilibrium results. Moreover, the arbitrage models, as they stand, generally imply no trade in the assets, an unacceptable result. The MS methodology may be the answer for such difficult situations. Let us demonstrate with a relaxation of one of the assumptions of the Black and Scholes option pricing model.

Black and Scholes assume that the standard deviation of the stock's returns, σ, is known to all investors, and all agree on this value. Since all other parameters of the Black and Scholes model are observed (t, S, E, and r), the call option price, C, is determined and in equilibrium there will be no trade in the option. If, for some reason, the call option were underpriced, *all investors* would want to buy it, there would be no sellers, there would be no transactions, and hence the price would not change. This is, of course, very unrealistic. Thus, to allow trade in the option one must introduce some "noise" into the system. We may introduce noise traders (with no information feedback), liquidity traders, or heterogeneous beliefs regarding the value of σ. For example, suppose that the investor denoted by i_{30} uses the past 30 days to estimate σ and investor i_{60} uses the past 60 days to estimate σ. Each may obtain a different estimate of σ and hence a different subjective equilibrium call price. Moreover, at the observed market price of the option it is possible that one investor may be willing to sell and one investor may be willing to buy the option. Hence there will be trade. With many investors with different subjective beliefs, it is hard to predict what will be the equilibrium price of the option and its price dynamics. One can employ MS to analyze this issue. Each investor will sell or buy some quantity of the option as a function of her or his subjective estimate σ, the possible error in this estimate, and the observed option price. From the demand-supply analysis, one can derive the option price and compare it to the Black and Scholes option value. Moreover, one can complicate the analysis and introduce noise traders (whose buy and sell orders are determined by some random process) in order to obtain a market price with heterogeneous beliefs as well as noise. When using MS, one can also determine the option price when transaction costs, constraints on the amount of short sales, and so on are introduced, pricing that is impossible to obtain with analytical tools. In Chapter 11 we employ MS to investigate option pricing with uncertainty and disagreement regarding the value of σ.

1.3.2. Expected Utility Pricing Models

The capital asset pricing model (CAPM) and its extensions rely on EUT. Decision making by the firm (e.g., dividend policy, project selection) are also generally analyzed in the EUT framework; the goal is to maximize the value of the firm and hence maximize investor utility. Similarly, efficiency analysis—that is, finding the efficient set of portfolios (in the mean-variance or stochastic dominance frameworks)—also relies on EUT. In addition to the EUT assumption, each of these models relies on a set of other model-specific assumptions. A violation of EUT or each of the specific model's assumptions may have a drastic effect on the theoretical results. It is possible that a relaxation of some factors can be treated analytically, but some cases cannot be treated analytically, and MS is very helpful in such cases. Let us demonstrate this issue with the CAPM.

To derive the CAPM one has to assume (among other things) normal return distributions and that investors are risk-averse expected utility maximizers. Also, homogeneous beliefs are required. Under these assumptions, all investors hold the same market portfolio and there will be no trade. For example, suppose that IBM announces the development of a new efficient computer. Then by the CAPM's homogeneity assumption, all investors will revise (upward) the expected return of IBM: all would like to buy more stock of IBM and no one would wish to sell it. There will be no trade, and hence the price of IBM will not change in spite of the new development—an absurd and unacceptable result! Using PT rather than the EUT framework, with subjective changes in the various portfolios' cumulative distributions, would not change the CAPM results. Levy (1999) shows that with cumulative prospect theory (CPT), which is a modification of PT as suggested by Tversky and Kahneman (1992), the CAPM results are still intact. Thus, also according to CPT there will be no trade in stocks.

In homogeneous expectation models like the classic CAPM, there will be no trade. In heterogeneous expectation models, one would be tempted to believe that trade would take place, but this is not necessarily so. It depends on what an investor believes regarding other investors' information, and possible information feedback. Let us quote from Black (1986) to understand why heterogeneous beliefs are not sufficient to guarantee trade:

> A person with information or insight about individual firms will want to trade, but will realize that only another person with information or insights will take the other side of the trade. Taking the other side's information into account, is it still worth trading? From the point of view of someone who knows what both the traders know, one side or the other must be making a mistake. If the one who is making the mistake declines to trade, there will be no trading on information. In other words, I do not believe it makes sense to create a model with information trading but no noise trading where traders have different beliefs and one trader's beliefs are as good as any other trader's beliefs. Differences in beliefs must derive ultimately from differences in information. A trader with a special piece of information will know that other traders have their own special information, and will therefore not automatically rush out to trade. ...
>
> Noise trading provides the essential missing ingredient. Noise trading is trading on noise as if it were information. People who trade on noise are willing to trade even though from an objective point of view they would be better off not trading. Perhaps they think the noise they are trading on is information. Or perhaps they just like to trade. (See Black's article 1986, in Thaler, 1993, pp. 5–6. On the role of noise traders, see also Kyle, 1984, 1985a, and 1985b).

In this book we adopt Black's approach: when we employ heterogeneous beliefs, we assume that traders act on their beliefs. Thus, the fact that there is another side to the trade with different beliefs does not affect the trader's decision to trade. For example, an investor who analyzes a firm's data believes he or she possesses the ability to understand the available data better than the other traders, and hence is willing to trade on this information even though it is also available to all other investors (this is similar to Roll's "hubris hypothesis," 1986). In this setting, there will be trade in the market, and prices may deviate from the homogeneous expectations of the CAPM equilibrium. MS allows finding the equilibrium pricing and the risk-return relationship when investors have heterogeneous beliefs regarding the parameters of the return distributions.

Another assumption of the CAPM is that all investors have the same (1-period) holding period. Suppose that all investors agree on, say, the 1-year parameters, but some invest for periods shorter than one year and some for periods longer than one year. What will be the equilibrium prices with the diverse holding period assumption? And how are prices determined if on top of the diverse holding period noise traders also exist? Such issues are hard to solve analytically and MS comes to the rescue. It is impossible to analyze all these cases in this book. However, we demonstrate the application of the MS approach to some of these issues in Chapter 10.

1.4. SUMMARY

Most models in finance and economics assume the expected utility framework, which implies that investors are rational, and in addition, each model relies on some specific simplifying assumptions. With these assumptions, analytical results are obtained. However, these models have two main drawbacks: they typically imply no trade, and at best they have only weak empirical support. Experimental studies allow us to examine investors' rationality and their systematic deviations from the expected utility framework. However, they do not tell us what are the effects of the systematic deviations from rational choice on pricing. It is also usually impossible to study the effects of relaxing the many other unrealistic model-specific assumptions on the model's results.

Because studying these effects analytically is difficult, if not impossible, we suggest in this book the microscopic simulation (MS) methodology. With MS, one can analyze the effects of noise trading, systematic deviations from expected utility, as well as the relaxation of many of the problematic assumptions underlying various models in finance and economics. This is the purpose of this book.

2

DECISION WEIGHTS, CHANGE OF WEALTH, AND VALUE FUNCTION

The Experimental Evidence

2.1. INTRODUCTION

Efficient Market Humbug: Economist Disproves Theory by Beating Market[1]

Remember the old joke about the ardent believer in the efficient market theory who refused to pick up a $100 bill lying on the sidewalk? According to the theory, he reasoned, the money couldn't be there.

The efficient market theory quaintly holds that the price of a stock or bond accurately reflects all currently available information about its fundamental value, and therefore it's impossible for the average investor to get an edge on the market. University of Chicago economist Richard H. Thaler, no friend of the theory, points out that one way to test it is to apply it to closed-end funds.

Unlike open-end funds, closed-end funds issue a fixed number of shares, which they invest in underlying securities. So the efficient market theory imposes the simple requirement that the fund's share price equal its net-asset value [NAV]. But since closed-end funds frequently trade at a substantial discount to NAV, the theory proves invalid in this case. And if the market can't perform a no-brainer like pricing securities for which market prices already exist, how can it be expected to accurately process information like positive earnings surprises?

[1] Gene Epstein, *Barron's*, February 1, 1999, p. 44, Copyright Dow Jones & Co., Inc., reprinted with permission.

In last week's column, I nominated Professor Thaler for a Nobel Prize based on his pioneering work in the economic theory of semi-rational man. This week, as I promised, I want to discuss the three stock mutual funds he helped inspire that seek to exploit the foibles of the semi-rational investor. Since the funds' trading strategies have pretty consistently beaten the market, they also tend to undermine the validity of the efficient market theory.

Start with the fund that's most striking in its approach, called **Behavioral Value**. (The others are **Behavioral Growth** and **Behavioral Long / Short**). This fund follows the classic contrarian approach of buying stocks that have been big losers over a long period of time, but with a special edge on that well-worn strategy. As Thaler's partner, former finance professor Russell Fuller, explains, such stocks tend to be undervalued because the market will often extrapolate bad news. Battered by what must have been a lot of negatives in the past, it now places an undue weight on more of the same. On the other hand, the contrarian makes the more reasonable bet that good, bad or average news is about equally likely to happen.

But to get an additional edge on those odds, Behavioral Value buys one of these companies only when it announces either a stock repurchase or when SEC data show insider buying. Normally, the market will also get excited about these events, but in this case its negative conditioning makes it hesitate.

While Behavioral Value began trading this past December, Fuller can provide an audited record for the past three years. Over that period, it's returned 18.3% per annum, net of fees, compared to 14.4% for the relevant market index, the Russell 2000 Value.

The performance of the Behavioral Growth Fund has been even more impressive. This fund began trading in December 1997, and for last year, it returned 33.2%, which compared with 28.6% for the S&P 500. Prior to that, the six-year record of the same trading strategy returned 25.5% per annum, also better than the 17.8% for the S&P.

Behavioral Growth buys stocks of companies posting quarterly earnings that exceed analysts' expectations. In this case, the behavioral assumption is that analysts will be slow to recognize that they've been wrong. Academic research already has shown these companies will tend to do better than the market. But the fund tries to increase these odds by doing some analysis of its own, in order to determine whether quarterly earnings will continue to perform at a high level.

You might imagine it's a bit presumptuous for anyone to think he can out-analyze professionals who watch a company on a continuous basis. But Fuller contends that under these circumstances, there can be a definite advantage in a fresh approach. "We have no stake in defending our mistaken forecast," he comments, "so we stand a better chance of getting things right."

Finally, the Behavioral Long/Short Fund uses a combination of various strategies. On the long side, it buys stocks based on the approach of each of the first two funds. On the short side, it sells stocks by applying the mirror image of the contrarian approach. It sells the highflyers, and then only after a company has issued shares or there has been insider selling of its stock.

As Fuller explains, Behavioral Long/Short will underperform a bull market. It's best suited to a market that's either gently rising or in a trading range. (For more on these funds, see www.fullerthaler.com and www.undiscoveredmanagers.com)

Most models in economics and finance assume that investors are completely rational and that the market is efficient. With these two assumptions, one can develop models that derive equilibrium prices of risky assets, the price of a unit of risk, the risk premium, the risk-return relationship, and so on. The *Barron's* article suggests that the market is inefficient and that investors are only "semirational," where semirational means that investors make systematic mistakes in their decision making. The article claims that stocks that have been big losers over long periods of time tend to be undervalued because the investors (who are not perfectly rational) will often wrongly extrapolate bad news, and they do not correct this mistake. To be more specific, the market places an undue weight on "more of the same" negative performance. This means that investors use the past performance to estimate the future performance and believe that the past will repeat itself. Thus, investors undervalue stocks with bad *ex post* performance. This implies that investors tend to err systematically in their investment behavior, and a sophisticated entrepreneur who understands this phenomenon can gain from these errors by establishing a mutual fund that invests in these undervalued stocks. These sophisticated investors make an abnormal profit due to the market inefficiency. As the article explains, such funds already exist, exploiting market inefficiency and investors' semirationality. In the LLS (Levy, Levy, and Solomon) Microscopic Simulation model, which is described in Chapter 7, we employ a similar approach by assuming that some of the investors look at the past rates of return and form their expectation regarding the future rate of return distribution based on these historical observations. Other more sophisticated investors (like the managers of Behavioral Value fund in the *Barron's* article), can use this knowledge about investors' behavior to predict trends and to reap abnormal profits by exploiting these trends.

In this book we make a distinction between analytical models that are generally very restricted in their assumptions regarding investors' behavior and microscopic simulation models which can incorporate almost any type of investor behavior, including the experimentally documented deviations from rationality. Analytical models in finance and in economics generally postulate a set of assumptions (in most cases including the assumption that the investors are rational) that lead to analytically derived results, as, for example, equilibrium prices of risky assets. Assumptions such as no-transaction costs and no taxes are often made despite the fact that it is well known that such assumptions do not hold in reality. Nevertheless, these assumptions are needed to derive the analytical results, which otherwise become difficult and in some cases even impossible to obtain. In a microscopic simulation model, one can relax these assumptions and solve (though not analytically) the problem under consideration. The analytical models that are based on unrealistic assumptions can be justified, however, on two grounds:

a. These models are the best that we have. Therefore, we accept a given model as long as we do not have a better model to replace it.
b. We accept a theoretical model despite its unrealistic assumptions as long as it is able to explain the empirical facts.

Milton Friedman (1953a) asserts that if a theoretical model fits the corresponding empirical data well, one can accept it even if it makes unrealistic assumptions. This is a *positive economic* approach as opposed to a *normative economic* approach. Friedman claims that investors behave "as if" the assumptions underlying the model are intact even though it is well known that they do not hold in reality. The "as if" argument is a justification for using these unrealistic models. The famous "pool table player" is probably one of the best examples describing this positive economic approach. The brilliant pool table players need to know high-level physics and mathematics to be able to know exactly where and with what force to hit the ball in order to succeed. The successful pool table players probably do not have this sophisticated knowledge. Yet they are successful players, which implies that they hit the balls "as if" they had this knowledge. The same is true with theoretical models that make unrealistic assumptions. If these models fit the empirical data well, particularly prices or return patterns, we conclude (as in the pool table player's success) that the price of risky assets is "as if" the assumptions underlying the model hold.

The empirical argument for justifying a given theoretical model has some pros and cons. We will not discuss them here but rather say that for many theoretical models, such as, the capital asset pricing model (CAPM) and the option pricing model (OPM), there are some empirical studies that support the theoretical models and some that reject it. Thus, the "as if" argument does not provide an ultimate justification for most theoretical models in finance.

An alternative method to the empirical approach is to evaluate a theoretical model by examining investors' behavior in a laboratory. This approach is known as *experimental economics*. We focus in this book on financial assets and, in particular, on stock market issues; hence the name "experimental finance" is more appropriate. Yet, in the rest of the book we call this approach "experimental economics," as this is the common name given in the literature to such experiments. The advantage of laboratory tests in comparison to empirical tests is that in laboratory tests a given issue can be analyzed without interference of other variables. Namely, the researchers can design an experiment that tests a given hypothesis when all other factors are well controlled. Thus, the experimental framework not only tests the theoretical models, it also tests the subjects' behavior and examines which of the assumptions of the model are violated by the subjects' behavior.

Another advantage of the experimental approach over the empirical approach in testing a theoretical model is that investors' behavior, as found in the experiment, can be incorporated in microscopic simulation models. Also, the subjects can be interviewed or explain in a questionnaire their investment decisions. For example, suppose that in line with the article that opens this chapter, one finds that investors typically examine the latest three annual rates of return in forming their expectations, even though they are told that future returns follow a random walk. Such behavior can easily be incorporated into a microscopic simulation model. In this and the next two chapters, we discuss various experimental results, some of which are consistent and others of which are in contradiction with the theoretical model. In Chapters 7 through 9 we show how the experimental findings regarding investors' behavior can be incorporated in a microscopic simulation model. In these chapters we also show that the experimental findings regarding investment behavior can explain many market anomalies or "puzzles."

In this book we distinguish between experiments in which the subjects have to make a choice between artificial bets (e.g., how much one is ready to pay for a bet of $\$ -50$ or \$350 with an equal probability) and experiments in which financial assets (e.g., stocks) rather than artificial bets are considered. Moreover, when financial assets are involved, in some of the experiments, the subjects trade in securities and collectively determine assets prices or assets returns, and each subject's decision affects the reward of the other subjects. Thus, the first type of experiments are more hypothetical, while the latter type tries to mimic market conditions in order to allow us to test directly some theoretical models. This chapter is devoted to the first type of experiments and the latter type will be discussed in Chapter 4.

The purpose of this chapter is to discuss investors' behavior as obtained in experimental studies. As we will see, the findings strongly suggest behavior that is very different from the one assumed in the classic models. A large portion of the rest of this book is devoted to the question of whether, and under what conditions, investors' deviations from rationality are important in determining asset prices and market dynamics.

In this chapter and in chapter 4 we focus on the following experimental findings regarding investors' behavior which violate the standard assumptions of models in economics and finance and, in particular, violate expected utility theory:[2]

a. Decision weights, $w(p)$, are used rather than the true probabilities, p.
b. Subjects make decisions based on *change* of wealth, x, rather than total wealth, $w + x$, where w is the initial wealth and x is the change in wealth.
c. Subjects have "mental departments" (i.e., make several accounts in their minds, which implies that the aggregate wealth maximization

[2] Despite these contradictions to expected utility theory, the CAPM has been shown experimentally to be intact (see Chapter 4).

is not the goal of their investment). This will also be validated with capital market experiments in Chapter 4.

d. Correlations of rates of return on various assets are ignored in the investors' decision-making process (this will be discussed in Chapter 4).
e. Subjects assign weights to past rates of return even though such rates of return may be irrelevant (as discussed in the opening article of this chapter). Thus, even when the subjects are told that the market is efficient and rates of return are drawn randomly from a given distribution known to the subjects, they form their expectations based on past returns.

This is a short list, but we believe it covers the most important systematic deviations from the rationality assumption in investment decision making.

2.2. DECISION WEIGHTS AND OBJECTIVE PROBABILITIES

Under a given set of axioms (e.g., monotonicity, transitivity, etc.),[3] von-Neuman and Morgenstern (1947) and others have proved that investors who face various options should select the optimum investment by the maximum expected utility criterion (MEUC), where the utility function reflects the investors' preferences (i.e., their attitude toward risk).

Experiments show that this is not the case. This may lead to the conclusion that investors either do not maximize expected utility (i.e., one or more of the axioms do not hold—for example, investors prefer having less money to having more money, which violates the monotonicity axiom) or that investors employ decision weights (subjective probability) rather than the objective probability, which leads to decisions that seem to contradict the MEUC. Of course, it is possible that both occur—that investors do not maximize expected utility and, on top of that, use decision weights rather than objective probabilities. In this section we focus on decision weights and probabilities. We would like to emphasize at the outset that if some patterns are discovered (e.g., investors overstate low probabilities of winning a large amount of money), we can incorporate this property in a microscopic simulation model and analyze the effect of such behavior on asset pricing.

Let us start with what is probably the most well-known case of contradiction to the MEUC—the Allais paradox (Allais, 1953). The violation to the MEUC is revealed in a two-part experiment. In Part I, a choice is offered between prospects A and B, and in Part II, a choice is offered

[3] For another set of axioms where FSD serves as one of the axioms, see Fishburn (1982).

between prospects C and D, as follows:

PART I:	*Prospect A*	
	Income (in million \$)	*Probability*
	1	1
	Prospect B	
	Income (in million \$)	*Probability*
	0	0.01
	1	0.89
	5	0.10
PART II:	*Prospect C*	
	Income (in million \$)	*Probability*
	0	0.89
	1	0.11
	Prospect D	
	Income (in million \$)	*Probability*
	0	0.90
	5	0.10

Results show that in Part I, most investors choose prospect A, and in Part II, most investors choose prospect D. In the following comparisons we show that these decisions are inconsistent and contradict expected utility theory. Denoting by U the utility function, the preference of A over B implies the following (all figures in million of dollars):

$$1U(1) > 0.01U(0) + 0.89U(1) + 0.1U(5) \tag{2.1}$$

and the preference of D over C implies:

$$0.9U(0) + 0.1U(5) > 0.89U(0) + 0.11U(1) \tag{2.2}$$

Inequality (2.1) can be rewritten as:

$$0.01U(0) + 0.1U(5) < 0.11U(1) \tag{2.3}$$

and inequality (2.2) can be rewritten as:

$$0.01U(0) + 0.1U(5) > 0.11U(1) \tag{2.4}$$

Because these results are inconsistent it is suggested that either the subjects do not maximize expected utility or that the expected utility model needs to be modified in order to accommodate and explain paradoxical results such as these. The common explanation offered in the literature for this behavior is that subjects overweigh the 0.01 probability of receiving nothing in option B, which explains the preference of A over B in Part I of the experiment.

For example, suppose that an investor who knows that the objective probability of receiving zero is $p = 0.01$ assigns to this specific outcome a subjective probability (or a decision weight) of $w(p) = 0.02$ and also reduces the probability of winning \$5 million from $p = 0.1$ to $w(p) = 0.09$

(such that the total probability remains 1). Choosing A over B with these decision weights (which are also called subjective probabilities) implies that

$$1U(1) > 0.02U(0) + 0.89U(1) + 0.09U(5)$$

which implies that

$$0.02U(0) + 0.09U(5) < 0.11U(1) \tag{2.5}$$

and Eq. (2.5) does not contradict inequality Eq. (2.4) anymore. Thus, one can reconcile the Allais paradox with the expected utility paradigm simply by replacing the objective probabilities, p, by decision weights, $w(p)$, which are nothing but subjective probabilities. An alternative explanation for the preference of option A is that there is a "certainty effect," where options with certain positive income are overvalued.

Of course, this is a possible explanation for the Allais paradox, but the hypothesis that investors assign subjective decision weights instead of the objective probabilities needs to be proven empirically or experimentally. We turn next to the experimental findings regarding this issue.

Preston and Baratta (1948) report results of an experiment in which the subjects "preferred" certain probabilities—namely, they used decision weights that overvalued these probabilities. The subjects participating in the experiment consistently overvalued low probabilities of high winnings called "long shots" and undervalued high probabilities of low winnings called "short shots." The indifference point, where probabilities were neither overvalued nor undervalued, was about a 0.20 probability of winning. Preston and Baratta interpret their results in terms of a scale of subjective probability (i.e., decision weights) different from the scale of objective probability. Griffith (1949) reports similar results in the analysis of actual betting on horse races.

In two papers published in 1953 and 1954, Edwards ran experiments where the subjects played a pinball machine and the reward was a function of the results in this game. Some of the experiments were for bets with equal expected monetary value (EV) and some for unequal monetary values. Edwards found that there is a general tendency to prefer some probabilities over others. Edwards is aware of the possibility that the results may be explained by the fact that the subjects maximize expected utility rather than expected monetary value; hence, the results cannot be unequivocally attributed to probability preferences. For example, the preference of \$1.05 with 0.4 probability over \$4.20 with 0.10 probability (both with the same expected value) may be explained by diminishing marginal utility rather than by a preference of probability 0.4 over probability 0.1 (i.e., $w(0.4) > 0.4$, and $w(0.1) < 0.1$). Edwards concluded, however, that probability preferences rather than diminishing marginal utility account for his experimental results:

> The use of probability rather than amount of win or loss as the independent variable to which the results are related in this experiment is based on the belief that utility curves are not the best way of accounting for these results. A later experiment will provide evidence to support the view that these results should be attributed to preferences among probabilities rather than to properties of utility curves of money. (Edwards, (1953), p. 362–363)

Edwards's main conclusion from the experimental results is that there is a general tendency to prefer a bet with a low probability of a large amount of gain and a tendency to avoid a bet with a small probability for a big loss (i.e., in both cases small objective probabilities seem to have been given larger subjective weights).

In his 1954 study, Edwards conducted an experiment where some of the bets had the same expected monetary values and some of the bets had an unequal expected value. With the equal expected value bets, his previous results of a 1953 study indicating a preference for certain probabilities were confirmed. When the subjects had to choose from a pair of bets of unequal expected value, there was a preference for the high expected value bets as well as a preference for certain probabilities, as obtained in the first experiment reported previously. The expected value of the bet was the dominant factor for the choices in those bets in which the subject could only win or break even. The expected value and probability preference were both important factors for the bets in which the subjects could only lose or break even, and the probability preference factor was the dominant one for those bets in which the subjects could either win or lose. Because in actual investment decisions in an uncertain world investors generally can win or lose money, it seems from Edwards's experiment that the most relevant result is that probability preference is the important factor in explaining investors' behavior.

Edwards's pioneering experiment paved the way to more experiments that will be discussed later on in the chapter. Yet Edwards's experiments have one important drawback: they do not rule out that other factors, in particular risk, account for the experimental results. In the 1954 experiment, Edwards added the expected monetary value (EV) as a possible factor explaining the subjects' choice (and not only the probability preference factor), but he did not incorporate risk. To emphasize how different expected values and probability preferences are intertwined and cannot be separated, let us cite from the summary of Edwards's 1954 article:

> It is concluded that probability preferences are important in determining decisions among bets even when there are objective reasons for preferring one bet over another, although as the difference in expected values increases it seems likely that the importance of probability–preference will decrease. (Edwards, 1954, p. 67).

Edwards wrote his papers in the early 1950s, when Harry Markowitz (1952a) published his breakthrough article on the investment portfolio selection by the mean-variance criterion. Thus, Edwards analyzed two factors: differences in expected monetary values (EV) and probability preferences. Nowadays it is obvious that another factor (risk) should be incorporated into such an analysis. To be more precise, there is no need to compare bets with different EV values. On the contrary, the expected return can be neutralized by selecting equal expected value bets. However, risk cannot be ignored in such a situation because it is a very major factor in financial decision making, and different probabilities imply different risk. The critical question is whether there is a probability preference after the differences in risk (as well as expected values, if they differ) are accounted for. This question is not answered by Edwards because risk is not incorporated into his analysis. Despite this drawback, Edwards paved the way for more research in this area. Probably the most well-known and influential of this research was conducted by Kahneman and Tversky, who in a series of papers analyze the subjects' behavior in a framework in which the entire distribution of returns (which accounts for the risk involved) is taken into account. We next review the main results of Kahneman and Tversky in detail.

In a breakthrough article, Kahneman and Tversky (1979) conducted several experiments showing that subjects tend to behave in contradiction to expected utility theory. The main findings are the following:

a. The subjects employed decision weights, $w(p)$, rather than the objective probabilities, p, (similar to Edwards's results).
b. The subjects made decisions based on change of wealth, x, rather than total terminal wealth, $w + x$, as implied by expected utility theory.
c. The subjects maximize the expected value function, $V(x)$, rather than the expected utility function, $U(w + x)$, where the value function is convex for $x \leq 0$ and concave for $x > 0$, and where x is the change of wealth and w is the investor's initial wealth. The $V(x)$ function passes through the origin ($V(0) = 0$). The value function $V(x)$ is S-shaped, as illustrated in Figure 2.1. It is worth mentioning that items b and c are hypothesized by Markowitz (1952b) (though not tested experimentally).[4]
d. Investors have "mental departments"; for example, they make optimization of their stock portfolio (one department), another optimization of their investment in their saving accounts in the

[4] See also Mosteller and Nogee (1951).

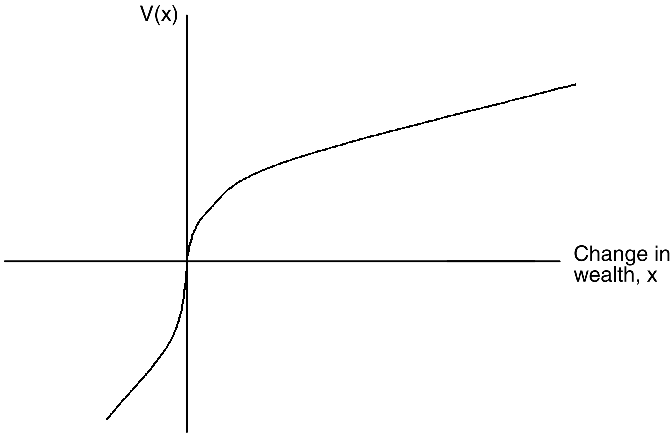


FIGURE 2.1 The value function $V(x)$ with $x = 0$ as a reference point. (From Kahneman & Tversky, Copyright 1979 The Econometric Society, reproduced with permission).

bank (another department), and so on, rather than a global optimization of all their assets as implied by portfolio theory.

Kahneman and Tversky first ran experiments that are similar to the Allais paradox. The subjects had to choose from prospects A and B (task I) and then from prospects C and D (task II). The prospects A, B, C, and D are given in Table 2.1. The percentage figure is the percent of the subjects preferring each prospect. Out of $N = 72$ subjects, 82% preferred

TABLE 2.1 The Certainty Effect

Task I: Select between A and B		
A	2500	With probability 0.33
	2400	With probability 0.66
	0	With probability 0.01
	$N = 72$	[18%]
B	2400	With certainty
		[82%]
Task II: Select between C and D		
C	2500	With probability 0.33
	0	With probability 0.67
	$N = 72$	[83%]
D	2400	With probability 0.34
	0	With probability 0.66
		[17%]

Note: Table taken from Kahneman & Tversky, Copyright 1979 The Econometric Society, with permission.

B over *A* implying that

$$0.33U(2{,}500) + 0.66U(2400) < U(2400) \Rightarrow 0.33U(2500) < 0.34U(2400)$$

Note that $U(0)$ is assumed to be equal to zero everywhere. This is possible because one can always make a linear transformation to the utility function such that $U(0) = 0$.

When the pair *C* and *D* are compared, 83% selected *C*, implying that

$$0.33U(2{,}500) > 0.34U(2400)$$

Thus, like in Allais's paradox, Kahneman and Tversky obtained a contradiction, but this time this inconsistent choice was verified in an experiment. Like in the Allais paradox, one can explain this result by a change in probability or what Kahneman and Tversky called the "certainty effect," meaning that subjects overweight certain outcomes relative to uncertain outcomes.

The next test is also related to small probabilities. One of the axioms of expected utility theory is that if option *B* is preferred to option *A*, then the option (B, p) must be also preferred to option (A, p) where *B* and *A* are obtained with a probability *p*. The experimental results reveal that this is not the case. Table 2.2 provides the outcomes and probabilities of options *A* and *B* and the experimental results. Thus, under *A*, there is a probability of 45% to obtain \$6000 and 55% to get zero, and under *B* there is a probability of 90% to earn \$3000 and 10% to obtain zero. As we can see, 86% of the subjects preferred option *B* over option *A*. Now let us compare the choice between *C* and *D*.

Note first that *C* and *D* are exactly as *A* and *B* but are obtained with a probability

$$p = \frac{1}{450}$$

TABLE 2.2 Overstating Low Probabilities

Task I: Select between A and B			
A $N = 66$	(6000, 0.45) [14%]	B	(3000, 0.90) [86%]
Task II: Select between C and D			
C $N = 66$	(6000, 0.001) [73%]	D	(3000, 0.002) [27%]

Note: Table taken from Kahneman & Tversky, Copyright 1979 The Econometric Society, with permission.

that is,

$$C = (A, p) \text{ and } D = (B, p), \text{ where } p = \frac{1}{450}$$

The preference of B over A implies that

$$0.45U(6{,}000) < 0.90U(3{,}000) \Rightarrow 0.001U(6{,}000) < 0.002U(3{,}000)$$

while the results of C and D reveal that the opposite holds:

$$0.001U(6{,}000) > 0.002U(3{,}000)$$

(Note that again $U(0) = 0$ is assumed.) Once again we obtain a contradiction. This contradiction can be explained by the fact that when there is a low probability to obtain a large prize (see option C) the subject tends to overstate this probability, exactly as Edwards claimed. Thus, the preference of C over D by 73% of the subjects is probably due to the fact that the subjects employ a decision weight $w(0.001) > 0.001$ when they select between C and D. It is possible that also $w(0.002) > 0.002$, but the bias is probably bigger with the lower probability of 0.001. Note that in this simple experiment, as in the Allais paradox, there is no need to neutralize the expected return or risk, because a contradiction to expected utility is obtained. Thus, the conclusion from these results is that either expected utility is not a valid paradigm or that probability is distorted—that is, subjective decision weights, $w(p)$, rather than the objective probabilities, p, are employed.

From Kahneman and Tversky's results, from Allais preliminary results, as well as from Edwards' results, we can conclude that indeed investors employ decision weights rather than the objective probabilities, at least in certain extreme cases.

While we will return later to some other results of Kahneman and Tversky's 1979 study, we would like to discuss now some of the results they published in 1992, which correspond to probability distortions. In another experiment with 25 graduate students, Tversky and Kahneman (1992) studied the subjects' tendency to distort probabilities and discovered some patterns. Table 2.3 presents the bets and the subjects' choices. The left column presents the outcomes of alternative bets and the first row provides alternative probabilities, p, of the second (more extreme) outcome, where $1 - p$ (not given in the table) is the probability of the first outcome. For example, the first line asserts that there was a bet of outcome 0 with a probability of $(1 - p) = 0.9$ and of \$50 with a probability of $p = 0.1$ (only $p = 0.1$ is given in the table). In this bet the median certainty equivalent (determined by the subjects) is \$9. Similarly, (0, 50) under $p = 0.50$ means that there is a probability of 0.50 to obtain 50 and 0.50 to obtain zero. The median certainty equivalent is 21 (see first line in the table). Having all

TABLE 2.3 Median Cash Equivalents (in Dollars) for all Nonmixed Prospects

	Probability								
Outcomes	**.01**	**.05**	**.10**	**.25**	**.50**	**.75**	**.90**	**.95**	**.99**
(0, 50)			9		21		37		
(0, −50)			−8		−21		−39		
(0, 100)		14		25	36	52		78	
(0, −100)		−8		−23.5	−42	−63		−84	
(0, 200)	10		20		76		131		188
(0, −200)	−3		−23		−89		−155		−190
(0, 400)	12								377
(0, −400)	−14								−380
(50, 100)			59		71		83		
(−50, −100)			−59		−71		−85		
(50, 150)		64		72.5	86	102		128	
(−50, −150)		−60		−71	−92	−113		−132	
(100, 200)		118		130	141	162		178	
(−100, −200)		−112		−121	−142	−158		−179	

Note: The two outcomes of each prospect are given in the left-hand side of each row; the probability of the second (i.e., more extreme) outcome is given by the corresponding column. For example, the value of \$9 in the upper left corner is the median cash equivalent of the prospect (0.9; \$50, 0.1). Table taken from Tversky & Kahneman, Copyright 1992 Kluwer Academic Publishers, with permission.

these choices, one cannot learn much about the relationship between p and $w(p)$ because the certainty equivalent is a function of both $w(p)$ and the utility (or value) function. However, Tversky and Kahneman assume (an assumption which is supported by some evidence provided by the experiment itself) that the preference that best fits the subjects' behavior is a two-part power value function of the form

$$V(x) = \begin{cases} x^{\alpha} & \text{if } x > 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases}$$

Having these preferences, and estimating the parameters α, β, and λ (by a nonlinear regression), one can use the results of Table 2.3 to analyze the relationship between p and $w(p)$. Thus, preferences and decision weights can simultaneously determine the results of Table 2.3. But as a value function is selected (with some justification), one can analyze the remaining factor—the decision weights. The results of Tversky and Kahneman corresponding to Table 2.3 are given in Figure 2.2. The straight 45° line reflects the no-bias case, namely, $w(p) = p$. As we can see from this figure, for both positive and negative prospects (w^+ and w^-, respectively), curves which deviate from the 45° line are obtained, implying that $w(p) \neq p$—that is, the subject "prefers" certain probabilities over others.

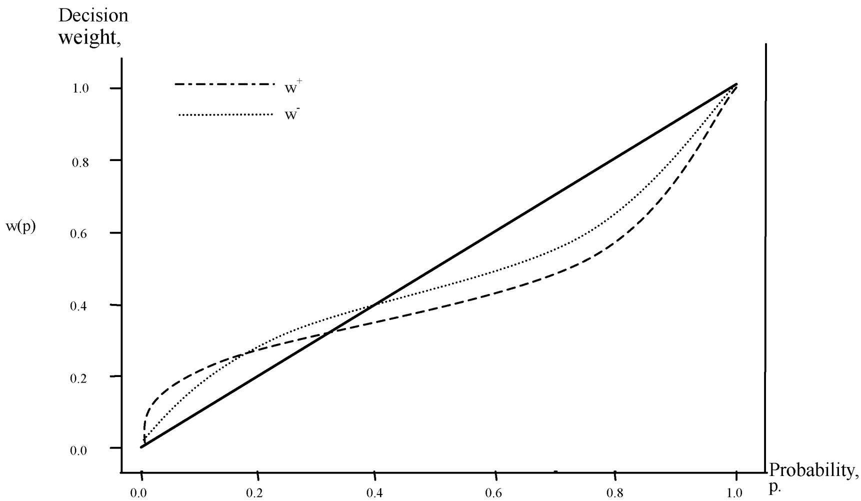


FIGURE 2.2 Probability p and decision weight $w(p)$ for positive (w^+) and negative (w^-) prospects. (From Tversky & Kahneman, Copyright 1992 Kluwer Academic Publishers, reproduced with permission).

The main conclusion as summarized in Figure 2.2 is that small probabilities are overweighted and high probabilities are underweighted for both positive and negative bets. Figure 2.3 presents the objective density function f, the distorted density function f_1, and the corresponding cumulative distributions F and F_1. As we can see, neither F nor F_1 dominates the other by first degree stochastic dominance (FSD), which means that we cannot claim that investors are subjectively better or worse off with the distorted probability distribution. Even if risk aversion is assumed, investors are not (subjectively) better off because F_1 cannot dominate F by second degree stochastic dominance (SSD).[5] Yet the distortion of probabilities as found in the experimental studies allows us to explain the existence of lotteries as well as insurance:

> Overweighting of small probabilities contributes to the popularity of both lotteries and insurance. (Tversky and Kahneman, 1992, p. 316)

To see this, consider the positive prospect (a lottery). Risk averters will not participate in a lottery as long as they perceive the objective probabilities.

[5] For the relationship between FSD, SSD, and attitude toward risk, see Levy, H., *Stochastic Dominance: Investment Decision Making under Uncertainty*, Kluwer Academic Publishers, 1998.

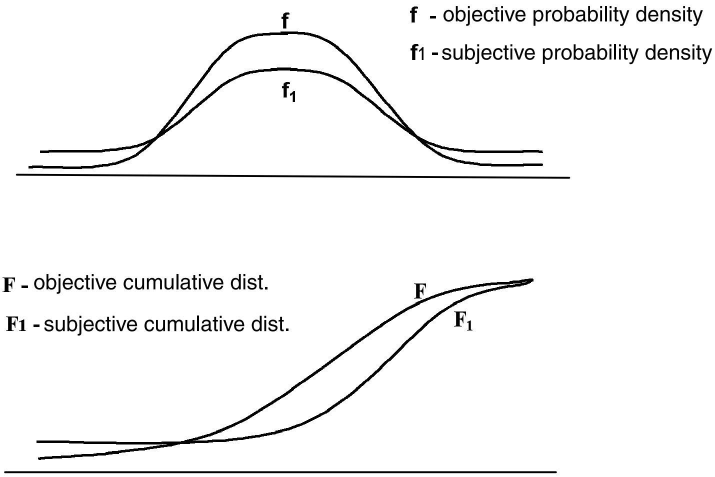


FIGURE 2.3 The effect of probability distortions. (From Tversky & Kahneman, Copyright 1992 Kluwer Academic Publishers, reproduced with permission).

In a lottery there is a small probability of winning the big prize; however, one can explain a participation in a lottery even with risk aversion. Simply, the subjective probability, $w(p)$, of winning the big prize is larger than the objective probability, p. To explain buying an insurance policy, there is no need to assume that the subject employs decision weights because risk aversion is sufficient to explain it. However, Tversky and Kahneman claim that even with negative prospects (losing the house if a fire breaks out), the low probabilities are overweighted (see curve w^- in Figure 2.2); hence, this factor enhances the tendency to buy insurance policies. These results are consistent with Edwards's findings.

2.3. CHANGE IN WEALTH VERSUS TOTAL WEALTH

Expected utility is defined in terms of terminal wealth, $w + x$. However, it is claimed that investors may contradict expected utility by making decisions based on the change of wealth, x, rather than on the terminal wealth, $w + x$. Like Markowitz, Kahneman and Tversky claim that the subjects tend to choose from a set of investments according to changes in wealth rather than the total wealth. Kahneman and Tversky support their view by experimental results. For example, when the subjects are given \$1000 and have to choose between prospects A and B given in Table 2.4, 84% of them selected B, when $N = 70$ is the number of subjects participating in the experiment (see Task I in Table 2.4). In Task II the subjects are given \$2000, and they have to choose either prospect C or prospect D. It is

TABLE 2.4 The Effect of the Initial Wealth

Task I: Select between A and B when additional $1000 is given to each subject	
A: (1000, 0.50) N = 70 [16%]	B: (500) [84%]
Task II: Select between C and D where additional $2000 is given to each subject	
C: (−$1000, 0.50) N = 68 [69%]	D: (−$500) [31%]

Note: All figures are in dollars. Table taken from Tversky & Kahneman, Copyright 1992 Kluwer Academic Publishers, with permission.

found that there is a preference for C by 69% of the subjects. Expected utility theory would predict that if B were selected by most subjects in Task I, D should also be selected by most subjects in Task II, because in terms of total wealth, both problems are identically given by

$$A = C = (2{,}000, 0.50; 1{,}000, 0.50)$$

and

$$B = D = 1500$$

(Recall that $1,000 is added in Task I and $2000 in Task II.) The fact that B was preferred over A while C was preferred over D by most subjects indicates that subjects base their decision making on *changes* in wealth rather than on *total* terminal wealth.

In an experiment conducted by Thaler and Johnson (1990), they found that investors behave according to a *path-dependent utility* function with an emphasis on change of wealth rather than total wealth, which conforms with Kahneman and Tversky's 1979 prospect theory. In other words, the subjects are not indifferent to obtaining $200 or, say, $100 today and $100 in two weeks (when the discount factor is not relevant, e.g., it is assumed to be zero). To see this result, let us cite from Thaler and Johnson's experiment, which is called Experiment 2 in their study (see Thaler and Johnson, 1990, p. 649):

> Experiment 2: Hedonic Editing: Temporal Spacing
>
> Instructions: Below you will find three pairs of events. In each case the same events occur, either on the same day (for A) or two weeks apart (for B). You are asked to judge whether A or B is happier, or in the event of two negative events, who is more unhappy. Would most people rather be A or B? If you think the alternatives are emotionally equivalent check "no difference." In all cases the events are intended to be financially equivalent. (Note: Having the events occur together does not imply that they occur sooner or later than if they were apart. That is not the question. You are only asked to judge whether it is better to have the events separately or together).

The following are Thaler and Johnson's results:

Task I: The subjects win $\$x$ in an office lottery (A) or win $\$\frac{1}{2}x$ and after two weeks win again $\$\frac{1}{2}x(B)$. Who is happier?

A 25% B 63% No difference 12% $N = 65$

Task II: The subjects receive a letter from the federal income tax authority saying that due to an arithmetical mistake $\$x$ must be paid on a given day (A), or alternatively $\$\frac{1}{2}x$ should be paid, and two weeks later another letter is received asserting that $\$\frac{1}{2}x$ should be paid again (B). Who is more unhappy?

A 57% B 34% No difference 9% $N = 65$

Task III: The subjects receive a $\$x$ parking ticket (A), or alternatively $\$\frac{1}{2}x$ tickets twice when a period of two weeks separates between the two tickets (B). Who is more unhappy?

A 75% B 17% No difference 7% $N = 65$

Summarizing this experiment, Thaler and Johnson state:

> The responses to question I reveal that for pairs of gains subjects did respond in the way suggested by the hedonic editing hypothesis. Subjects preferred to spread out the arrival of pleasant events, presumably to help segregate the pleasures experienced. Using the same logic, subjects should prefer to have pairs of losses occur on the same day, to facilitate their integration. However, subjects did not express this preference. Rather, in questions 2 and 3 subjects indicated that they prefer to experience the losses separately. We have obtained this result repeatedly, for small or large losses, for nonmonetary as well as monetary losses, and for unrelated and related pairs of events. This result is a severe blow to the hedonic editing hypothesis. (Thaler and Johnson, 1990, p. 649. Reprinted by permission, The Institute for Management Sciences (currently INFORMS), 901 Elkridge Landing Road, Suite 400, Linthicum, Maryland 21090-2909 USA.)

Thaler and Johnson consider the results of Tasks II and III as a severe blow to the hedonic editing hypothesis. Indeed, the persistence of these results in many studies seems to be quite puzzling. However, in a recent paper by Levy and Wiener (1999), these surprising results are explained by a combination of a path-dependent utility function, $U(w, x)$ (where w is wealth and x is change of wealth), and by the von-Newman and Morgenstern utility function, $U(w + x)$. The main result of Levy and Wiener that we would like to emphasize here is that investors are myopic: they make decisions according to change of wealth, x, but after a period of time elapses, they realize that their wealth is $w + x$, and their utility may change even though no cash flows occur. Levy and Wiener also show that the path-dependent utility function is consistent with the observed stock market overreaction as well as the experimental results of Thaler and Johnson.

To summarize, the experimental findings of Kahneman and Tversky and others regarding change in wealth and value function are as follows:

a. Change of wealth rather than total wealth determine the subject's choices.
b. Risk seeking prevails for negative outcomes and risk aversion for positive outcomes.
c. Ignoring the discount rate, the subjects are not indifferent between receiving \$$x$ (or paying \$$x$) on one day or receiving \$$\frac{1}{2}x$ (or paying \$$\frac{1}{2}x$) when there is a 2-week separation between payments.

From these three properties, Kahneman and Tversky conclude that the investors maximize the expectation of a value function, rather than the expectation of the common utility function, U. The value function is denoted by $V(x)$, where x is the change of wealth. This is in contrast to the maximization of the expected utility function, $U(w + x)$, where w is the initial wealth and x is the change in wealth. The value function has the property that $V'(x) > 0$ for $V \lesseqgtr 0$ and $V''(x) < 0$ for $x > 0$ and $V''(x) > 0$ for $x < 0$ (see Fig. 2.1).

2.4. RISK AVERSION AND RISK SEEKING

In most experiments the subjects are university students who are asked some hypothetical questions of choices when uncertainty prevails. Jacob Marchak (1964) asserts that experiments with students have limited relevance in a world of business. More specifically, he states:

> Tentative explorations performed . . . on graduate students or by these students on their wives do supply some preliminary evidence that deserves to be treated in a more rigorous way. It would be worthwhile to perform such experiments on mature executives rather than on students. (p. 104)

Indeed, this is exactly what Ralph O. Swalm (1966) did. In his experiment he interviewed about 100 businessmen and investigated their choices under conditions of uncertainty. They all answered his questionnaire from the executive perspective rather than from an individual perspective. Thus, all cash flows presented to them were the cash flows faced by the corporations they ran and they were supposed to find the certainty equivalent of uncertain prospects. Let us present one of these bets to illustrate the emphasis on the corporation rather than the individual involved:

> Suppose you are faced with choosing between one of two alternative courses of action. The first involves undertaking to bid on a new project. If the bid is successful, your company will make a net gain of, say, \$100,000. If unsuccessful, you will be reimbursed for the costs of making the bid, making your net gain zero. Your best available information leads you to assign a 50–50 chance to these possible events.

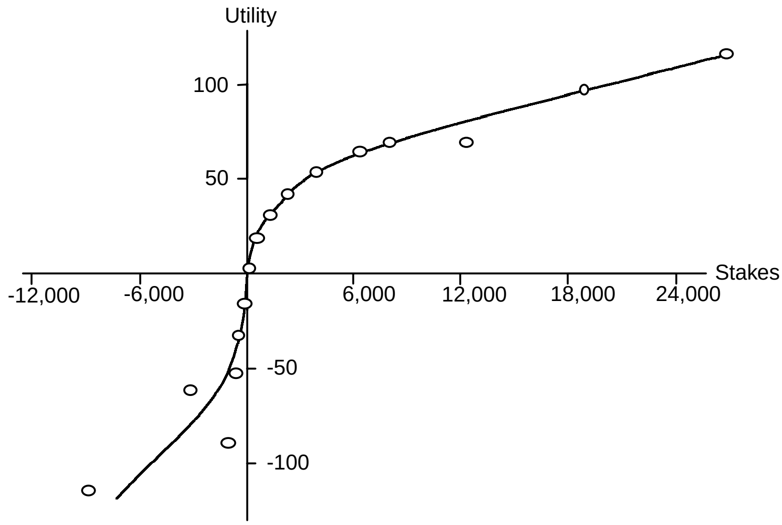


FIGURE 2.4 A typical utility function of an executive as obtained by Swalm. (Reprinted by permission of *Harvard Business Review*. [Exhibit VI] From "Utility Theory—Insights into Risk Taking" by Ralph O. Swalm, Nov.–Dec./1966. Copyright © 1966 by the President and Fellows of Harvard College; all rights reserved.)

> Your second possible course of action is to put the human resources you might spend in making the bid into cost-reduction efforts. Based on past experience, you are certain that this would result in a net gain. How large would this certain gain have to be to make you indifferent as to which choice to make? In other words, at what certain income would you be indifferent to your company's getting that income or getting a 50–50 chance of making $100,000 or nothing?

And when negative cash flows are involved, questions like the following have been asked:

> Suppose your company is being sued for patent infringement. Your lawyer's best judgment is that your chances of winning the suit are 50–50; if you win, you will lose nothing, but if you lose, it will cost the company $1,000,000. Your opponent has offered to settle out of court for $200,000. Would you fight or settle?

By repeating many questions of this sort with negative and positive cash flows, Swalm investigated the utility function of each executive. He obtained utilities of a wide range. However, the typical utility obtained is shown in Figure 2.4. This reveals risk seeking for negative outcomes and risk aversion for positive outcomes. Several remarks are called for:

a. The utility curve is very similar to the value function introduced by Kahneman and Tversky (1979) in their famous prospect theory (see Figure 2.1).

b. Obviously, the corporation for which the executive works has a lot of wealth, hence even the negative bets should be located in the positive domain because the firm's wealth is $w + x$, where w stands for initial

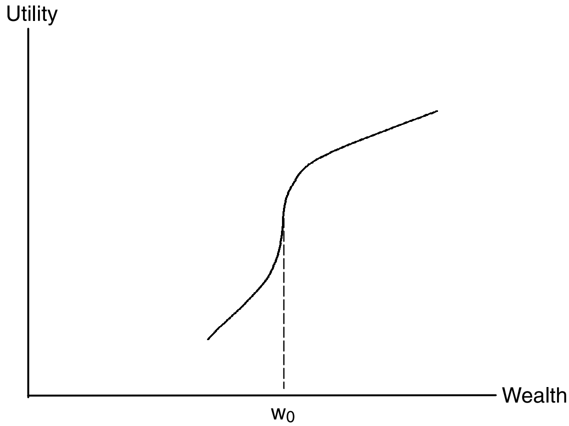


FIGURE 2.5 The S-shape utility function with a reference point at w_0. (From Kahneman & Tversky, Copyright 1979 The Econometric Society, reproduced with permission).

wealth and $x < 0$ stands for the possible losses. Thus, one expects to obtain the same attitude toward risk for $x < 0$ and for $x > 0$, because in both cases $w + x > 0$. This is not the case, because for most executives, there are two segments to the utility function with a clear-cut distinction between the range $x < 0$ and the range $x > 0$.

This implies that the utility function is a function of change of wealth rather than total wealth, exactly as claimed by Markowitz (1952b) and by Kahneman and Tversky (1979). Alternatively, one can add the initial wealth, w, such that $w + x > 0$ for both positive and negative bets. In this case the utility function would be of the form described in Figure 2.5, where w_0 is the initial wealth without the bet. Namely, w_0 rather than zero is the reference point of the utility function.

c. Maybe the most important conclusion with implication to our book is that the executives are not always consistent (see the dots that deviate from the utility function in Figure 2.4). They are not always rational and do make mistakes. This is well summarized in a compact way by Raiffa as well as Swalm:

> People do not always behave in a manner consistent with maximizing their utility, according to Raiffa, and this ... clearly demonstrates how important it is to have a theory which can be used to aid in the making of decisions under uncertainty. If most people behaved in a manner roughly consistent with [the] theory, then the theory would gain stature as a descriptive theory but would lose a good deal of its normative importance. We do not have to teach people what comes naturally. But as it is, we need to do a lot of teaching. (Swalm, 1966, p. 127. Reprinted by permission of *Harvard Business Review*. [Exhibit IV] From "Utility Theory—Insight into Risk Taking" by Ralph O. Swalm, Nov.–Dec./1966. Copyright © 1966 by the President and Fellows of Harvard College; all rights reserved.)

Swalm adds:

> The utility model is not a good descriptive theory for risk-taking situations. When people do what comes naturally, they are as inconsistent as can be. The real question is whether they want to act inconsistently or whether they wish to employ a methodology which would allow them to resolve their inconsistencies and allow them to analyze complex problems.
>
> But, alas, most businessmen are not perfectly rational men; rather, they are as inconsistent as all mortals are: there is no need to have a prescriptive theory of utility for the perfectly rational man. Just tell him to do what comes naturally. The raison d'être for utility theory is that most of us are not supermen—we make errors of judgment and are inconsistent in our choices. And knowing this, some of us, when faced with an important and complex problem, might wish to employ, in a conscious manner, decision aids that will help police our inconsistencies and help guide us to an appropriate course of action. Utility theory purports to be such a decision aid. (Swalm, 1966, p. 127. Reprinted by permission of *Harvard Business Review*. [Exhibit IV] From "Utility Theory—Insight into Risk Taking" by Ralph O. Swalm, Nov.–Dec./1966. Copyright © 1966 by the President and Fellows of Harvard College; all rights reserved.)

From these citations and from the results of the experiment conducted by Swalm, we can conclude that investors are not "supermen," they make mistakes and they are often inconsistent. However, this does not mean that utility theory and other theoretical models are worthless. On the contrary, one can learn from them to improve the decision-making process and to police the existing inconsistencies in our decisions. In the LLS microscopic simulation model, which is described in Chapters 7 and 9, we assume that investors have a goal in mind, such as maximizing expected utility or some other value function, yet they make mistakes and hence deviations from rationality may take place.

Tversky and Kahneman (1979), who conducted their experiment with students rather than executives, found, like Swalm (1966), that the subjects reveal risk aversion in the positive domain and risk seeking in the negative domain. Let us elaborate on their experiment.

The pair of choices given in Table 2.5 were presented to the subjects. "Positive prospects" refer to prospects with positive monetary value and "negative prospects" refer to prospects with outcomes falling in the negative domain. As can be seen, 80% of the subjects preferred \$3000 with certainty, over \$4000 with a probability of 80% and \$0 with a probability of 20%. As

$$\$4000 \cdot .80 = \$3{,}200 > \$3000$$

the expected return of the preferred prospect is smaller than the expected return of the inferior prospect; hence the expected value cannot account for the observed choices. Therefore, Kahneman and Tversky conclude that risk aversion prevails in the positive domain. When we examine the prospect with losses, the opposite holds; the subjects prefer a loss of \$4000 with an 80% probability over a loss of \$3000 with certainty. As the expected loss is bigger with the uncertain prospect, Kahneman and Tversky

TABLE 2.5 Risk Seeking and Risk Aversion

	Positive prospects		Negative prospects
Task 1	(4000, 0.80) < (3000) N = 95 [20%] [80%]	I	(−4000, 0.80) > (−3000) N = 95 [92%] [8%]
Task II	(4000, 0.20) > (3000, 0.25) N = 95 [65%] [35%]	II	(−4000, 0.20) < (−3000, 0.25) N = 95 [42%] [58%]
Task III	(3000, 0.90) > (6000, 0.45) N = 66 [86%] [14%]	III	(−3000, 0.90) < (−6000, 0.45) N = 66 [8%] [92%]
Task IV	(3000, 0.002) < (6000, 0.001) N = 66 [27%] [73%]	IV	(−3000, 0.002) > (−6000, 0.001) N = 66 [70%] [30%]

Note: Table taken from Kahneman & Tversky, Copyright 1979 The Econometric Society, with permission.

conclude that risk seeking prevails in the negative range of outcomes. Of course, these conclusions are based on the assumption that the probability of 0.80 is a relatively large probability; hence, it is not subjectively distorted. Also, even if the probability of 0.20 for a zero outcome is distorted, as $U(0) = 0$ by construction, this does not affect the results.

Task II (see the positive prospects) reveals that when probabilities are multiplied by one-fourth (i.e., the chance of winning $4000 is reduced from 80% to 20%), the choices are reversed (compared to Task I), exactly as obtained before. This can be explained by the "certainty effect" by which certain options are overvalued. Alternatively, we can assert that the relatively low probability of 0.20 of winning $4000 is overvalued with a decision weight $w(0.20) > 0.20$. Comparisons of Tasks III and IV reveal basically the same results of Tasks I and II showing that by moving from, say, prospects A and B to, say, prospects (A, p) and (B, p), the choice may be reversed in contradiction to expected utility theory. Similar results are obtained with the negative prospects given in Table 2.5, and the preference of the prospects is reversed once the options are obtained with a probability, p (compare Tasks III and IV). One can explain these results by the possibility that the low probability of 0.001 to get −6000 (see Task IV) is overstated, hence the preference is reversed.

Another conclusion from Table 2.5 is that in the negative domain, risk seeking prevails. This conclusion is based on the finding that (−4000, 0.80) is preferred to −3000 with certainty (see Task I with negative prospects). From these results, Kahneman and Tversky conclude the following:

a. Subjects overvalue options with certain positive income and undervalue options with certain negative income.
b. Subjects tend to overvalue relatively small probabilities of a relatively big gain and of a relatively big loss.
c. Risk aversion prevails for $x > 0$ and risk seeking prevails for $x < 0$ (see Figures 2.1 and 2.4).

2.5. CUMULATIVE PROSPECT THEORY: DECISION WEIGHTS AND STOCHASTIC DOMINANCE

First degree stochastic dominance (FSD) developed by Hadar and Russell (1969) and Hanoch and Levy (1969) states that prospect F dominates prospect G for any preference (in the expected utility framework) if and only if $F(x) \leq G(x)$ for all x and $F(x_0) < G(x_0)$ for at least one value x_0, where F and G are the cumulative distributions of the rates of returns of these two prospects, respectively. Thus, by FSD we have:

$$\begin{array}{lcl} F(x) \leq G(x) & & E_F U(x) \geq E_G U(x) \text{ for all utility } U \in \mathbf{U}_1 \\ & \Leftrightarrow & \\ \text{and } F(x_0) < G(x_0) & & \text{and } E_F U_0(x) > E_G U_0(x) \text{ for some } U_0 \in \mathbf{U}_1 \end{array}$$

where $U \in \mathbf{U}_1$ if $U' \geq 0$. Actually, FSD is another way to formulate the monotonicity axiom asserting that investors are always better off by having more rather than less wealth.

FSD is very fundamental (it assumes nothing about the shape of the distributions and about the preferences), hence researchers are not willing to accept a theory that violates it. Unfortunately, prospect theory as formulated in 1979 by Kahneman and Tversky violates FSD. Namely, according to prospect theory, prospect B may be preferred to prospect A by some subjects, even if A dominates B by FSD. The violation of FSD in the prospect theory framework stems from the fact that the decision weights, $w(p)$, are not a linear function of the objective probabilities, p, and that the sum of the decision weights are not necessarily equal to 1. To see the drawback of prospect theory in this respect consider two values, x_1 and x_2, where $x_2 > x_1$. For example, take $x_1 = 10$ and $x_2 = 11$. Now compare the two bets A and B given in Table 2.6. We may get that

$$E_A U(x) = U(11) < 0.75 \cdot U(10) + 0.35 \cdot U(11) = E_B U(x)$$

Thus, the inferior prospect B may be preferred because we may obtain $EU_B(x) > EU_A(x)$ for some utility function such that $U \in \mathbf{U}_1$. Note that in this case a relatively small probability is heavily overweighted but a

TABLE 2.6 The Violation of FSD by Prospect Theory

Return	Probability	Decision weight
	Prospect A	
$x_2 = 11$	1	1
	Prospect B	
$x_1 = 10$	0.8	$\pi(0.8) = 0.75$
$x_2 = 11$	0.2	$\pi(0.2) = 0.35$

TABLE 2.7 The Violation of FSD with $\Sigma w(p) = 1$

Return x	Probability p	Decision weights $w(p)$
	Prospect A	
5	1/4	1/4
8	1/4	1/4
10	1/2	1/2
	Prospect B	
5	1/8	2/8
6	1/8	2/8
8	1/8	2/8
20	5/8	2/8

relatively high probability is only slightly underweighted. Also, when $p = 1$, $w(p) = 1$ (see Prospect A).

One may attempt to correct this drawback of prospect theory by normalizing the decision weights such that their sum will be equal to 1. However, even with such a normalization, FSD may be violated. To see this, consider the two prospects with three possible outcomes shown in Table 2.7. Note that $\Sigma p_i = \Sigma w(p_i) = 1$. Thus, the decision weights are selected such that there is no need for a normalization. Yet it is obvious that $F_B(x) \leq F_A(x)$ with objective probabilities, but with decision weights the two distributions cross. Figure 2.6 illustrates that $F_B(x) \leq F_A(x)$ but $F_A^*(x)$ is not smaller than $F_B^*(x)$ everywhere (where the superscript denotes cumulative distributions with decision weights). Thus, even when $\Sigma w(p) = 1$, FSD may be violated, which is a severe drawback of prospect theory.

Another drawback of the decision weights as suggested by prospect theory is that one cannot apply them to continuous distributions (e.g., normal distributions).

Quiggin (1982, 1993) suggests that probability distortions can be represented by a subjective transformation of the cumulative probability distribution. This transformation, on the one hand, helps to explain some of the observed paradoxes found in experimental findings and, on the other hand, avoids the drawbacks of the decision weights, $w(p)$, and, in particular, avoids the FSD violation. Quiggin suggests performing a transformation, T, on the cumulative distribution, F, such that a subjective cumulative distribution, F^*, is obtained. To be more specific, he suggests the following transformation:

$$F^*(x) \equiv T(F(x))$$

where $T' \geq 0$, $T(0) = 0$ and $T(1) = 1$. This type of transformation has the following advantages over probability weights:

a. It applies to both continuous and discrete random variables.

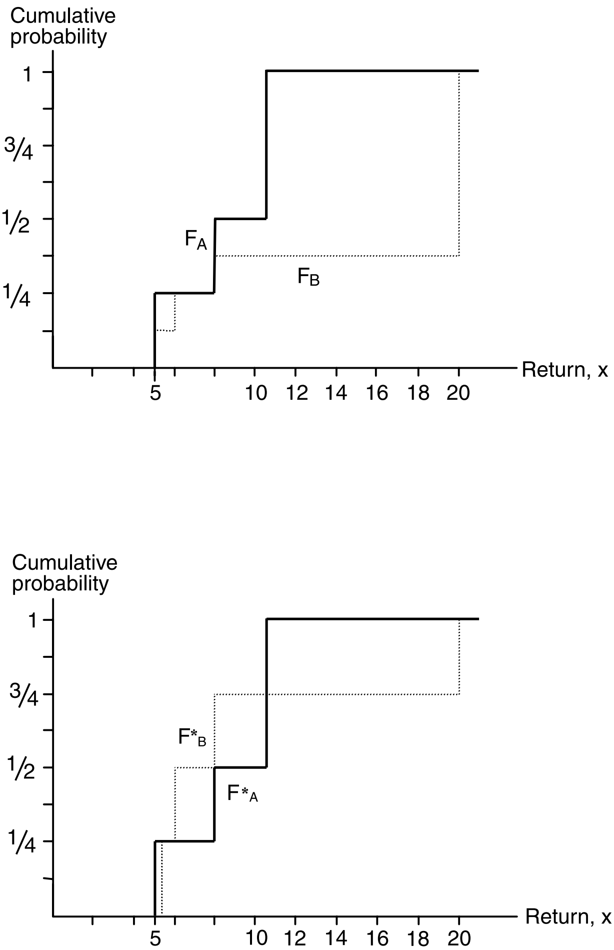


FIGURE 2.6 The cumulative distributions F_A and F_B with objective probabilities, and F_A^* and F_B^* with decision weights.

b. The sum of probabilities is, by construction, always equal to 1, because $F(1) = 1$ and $T(1) = 1$.
c. It allows the employment of subjective probabilities (overweighting and underweighting, as found in the experimental studies).
d. It does not violate FSD because

$$F(x) \leq G(x) \Leftrightarrow T(F(x)) \leq T(G(x))$$
$$\text{where } T'(\cdot) \geq 0.^6$$

[6] For a proof, see Levy and Wiener (1998).

Using this framework, Tversky and Kahneman (1992) suggest modifying prospect theory, where a transformation of the cumulative distribution substitutes for the decision weights as suggested in 1979.

Yet Tversky and Kahneman claim that if the prospect has only two outcomes of the form $(x, p; y, 1 - p)$, where $x > y > 0$ (or $x < y < 0$), then the original 1979 prospect theory is similar in its prediction to the new cumulative prospect theory, CPT (which is also called rank dependent utility function). Moreover, they see some advantages to prospect theory over cumulative prospect theory:

> Despite its greater generality, the cumulative functional is unlikely to be accurate in detail. We suspect that decision weights may be sensitive to the formulation of the prospect as well as to the number, the spacing and the level of outcomes. In particular, there is some evidence to suggest that the curvature of the weighting function is more pronounced when the outcomes are widely spaced. The present theory can be generalized to accommodate such effects but it is questionable whether the gain in descriptive validity, achieved by giving up the separability of values and weights, would justify the loss of predictive power and the cost of increased complexity. (Tversky and Kahneman, 1992, p. 317)

Thus, Tversky and Kahneman seem to prefer the original prospect theory with the decision weights rather than the more complex and less intuitive cumulative prospect theory even though the weights function may imply choosing a prospect that is inferior by FSD. This does not bother Tversky and Kahneman, because they believe that agents may not be completely rational. The best description of their view is manifested in the last paragraph of their 1992 paper:

> Prospect theory departs from the tradition that assumes the rationality of economic agents; it is proposed as a descriptive, not normative theory. The idealized assumption of rationality in economic theory is commonly justified on two grounds: the conviction that only rational behavior can survive in a competitive environment, and the fear that any treatment that abandons rationality will be chaotic and intracable. Both arguments are questionable. First, the evidence indicates that people can spend a lifetime in a competitive environment without acquiring a general ability to avoid framing effects or to apply linear decision weights. Second, and perhaps more important, evidence indicates that human choices are orderly, although not always rational in the traditional sense of the word. (Tversky and Kahneman, 1992, p. 317)

It is interesting that the conclusion of Tversky and Kahneman is in line with the article opening the chapter: investors are "semirational" but not completely rational. Also, the market may be inefficient and remain so for the "life time" of the investor. The LLS microscopic simulation model, which is described in Chapters 7 and 9, is constructed in line with the conclusion of Tversky and Kahneman's paper. In this model, investors are assumed to maximize expected utility, yet some of them have "noisy" decision making, which constitutes a departure from perfect rationality.

For investment optimization the *ex ante* rate of return distribution has to be estimated. In the LLS model some investors employ the *ex post* return distribution to estimate the *ex ante* distribution. Given a set of *ex post* returns and in absence of additional information, an unbiased investor forms his or her estimation of the *ex ante* distribution by assigning an equal weight to each of the observed past rates of return. An investor who is biased, as Kahneman and Tversky find most people to be, assigns decision weights with particular emphasis to low probabilities of extreme outcomes. One can examine how these weights affect asset pricing and the market dynamics. Alternatively, one can draw the historical cumulative distribution, employ a transformation on the distribution, and once again incorporate the transformed distribution rather than the original distribution as a factor in the microscopic model. Using microscopic simulations one can also examine the effects of basing investment decisions on change in wealth rather than on terminal wealth on asset allocation, asset pricing, and market dynamics. These issues are discussed in Chapter 9.

2.6. SUMMARY

Theoretical models like the expected utility paradigm tell us how investors should behave. Empirical and experimental studies tell us how people do behave in practice. The fact that there is a gap between the normative and positive approach does not mean that the theoretical models are worthless. On the contrary, they should be taught in order to guide investors in their investment decision making so that they will behave consistently and reduce the number of errors made.

Yet investors may make the same type of mistake over and over again. The mind simply collects and digests data subjectively and not as predicted by the theory. If indeed this is the case, this systematic deviation in practice of the investment behavior from the normative behavior is here to stay. It is important to understand the deviation from the normative behavior for two reasons: first, it allows us to understand price behavior and many of the observed market "anomalies" and, second, one may take advantage of the semirationality and even make money from other investors' errors, as explained in the *Barron's* article opening the chapter.

The three main systematic deviations from normative behavior and contradictions to expected utility as found in the laboratory tests are as follows:

a. The subjects apply decision weights, $w(p)$, rather than the objective probabilities, p, with some systematic bias.
b. As hypothesized by Markowitz and tested by Tversky and Kahneman, the subjects make decisions based on change of wealth rather

than total wealth, which is in sharp contradiction to the expected utility paradigm.

c. The subjects maximize the expectation of a value function, $V(x)$, rather than a utility function, $U(w + x)$, with $V(0) = 0$, $V'(x) > 0$ for $x \gtrless 0$, $V''(x) > 0$ for $x < 0$, and $V''(x) < 0$ for $x > 0$. This type of value function was found by Swalm (1966), who tested businessmen, and by Kahneman and Tversky (1979) and Tversky and Kahneman (1992), who tested students.

None of the studies mentioned in this chapter attempt to analyze the effect of these experimental findings on assets' prices. Indeed, modeling these experimental findings analytically usually raises serious tractability problems.[7]

Microscopic simulation allows the investigator to incorporate these and other experimental findings (which will be described in the next chapters) into asset pricing models and to test their effects on asset prices and on market dynamics. In Chapters 7 and 9, we study the effects of the experimental findings regarding investors' behavior on asset pricing in the context of the LLS market model.

Finally, note that even when subjects commit an error (e.g., overestimate a certain probability), the experiments reveal that only a certain proportion of the subjects commit this error, a fact that can be incorporated in microscopic simulation models. Thus, a market with some proportion of rational investors and some proportion of irrational investors can be constructed, and the market dynamics can be analyzed, with an emphasis on the sensitivity of prices as determined in such markets to the proportion of irrational investors.

[7] Levy (1998) shows that the CAPM is still intact even under cumulative prospect theory (CPT).

3

EMPIRICAL AND EXPERIMENTAL EVIDENCE REGARDING PREFERENCES

Absolute and Relative Risk Aversion

3.1. INTRODUCTION

The modern economic theory of decision making under uncertainty is based on the expected utility framework developed by von-Neuman and Morgenstern (1947). The expected utility framework makes several assumptions regarding preferences (for example, that people prefer more rather than less), and from these assumptions a decision-making rule is derived—namely, each individual can be characterized by a utility function $U(W)$, where W denotes wealth, such that given any uncertain situation, maximizing the individual's welfare is aligned with maximizing the expected value of the utility function $EU(W)$. All the information about the preferences of an individual is given by the form of his utility function.

The common functional forms that are assumed for the utility functions in the economic and financial literature include quadratic, logarithmic, power, and negative exponential utility. The negative exponential utility function $U(W) = -e^{-\alpha W}$ is an especially popular assumption because it is very convenient for analytical treatment. Quadratic utility is also an attractive modeling assumption because it allows mean-variance analysis that disregards higher moments of the return distribution. Since the microscopic simulation (MS) method is not as restricted by tractability constraints as analytical methods, one has the luxury of being able to work with any form of utility function. In this framework the choice of utility function is motivated by the realism of the preference assumed, rather than by implications for analytical tractability. Obviously, if one wishes to obtain reliable and even quantitative results from the model, one needs to be as realistic as possible in the assumptions regarding preferences.

In this chapter we review empirical and experimental evidence regarding the form of individuals' utility functions. The experimental studies, and with some qualifications the empirical studies, covered in this chapter shed light on the typical shape of subjects' preferences as well as the degree of the risk aversion that prevails among the subjects. We employ these findings later in the MS framework. Obviously, we do not assert that all subjects have a certain preference, and heterogeneous preferences are possible. Moreover, if we find that the majority of the subjects behave according to some preference and some small proportion of the subjects behave differently (e.g., some are risk lovers), this preference composition, in principle, can be incorporated into an MS model.

The shape of the investors' utility functions is a subject of many theoretical, empirical, and experimental studies. Friedman and Savage (1948) postulate that investors typically are characterized by a utility function with a risk-averse segment, followed by a risk-seeking segment, and then, once again, a risk-averse segment. This utility function advocated by Friedman and Savage is derived by a positive approach. Namely, it is deduced from the actual investment behavior. To be more specific, Friedman and Savage observe that investors buy insurance policies, buy lottery tickets, and buy both insurance and lottery tickets simultaneously. Markowitz (1952b) refines Friedman and Savage's analysis, making more suggestions regarding the shape of the utility function. Markowitz also postulates that investors make decisions based on *change* of wealth rather than on total wealth.

Experimental studies confirm Markowitz's hypothesis regarding the role played by change of wealth rather than total wealth in the investment decision-making process. These studies also show that the preference is typically an S-shaped function with risk seeking in the negative domain of returns and risk aversion in the positive domain of returns (for more details, see Chapter 2). In the experiments discussed in the next chapter, as well as in virtually all empirical studies, it is observed that a positive risk premium prevails in the market, implying that risk aversion dominates the market. Therefore, in this chapter, we focus on risk-averse utility functions or the risk-averse segment of S-shape functions.[1]

[1] Levy and Wiener analyze the risk premium and the change in risk premium as wealth changes for all S-shape value function. For more details, see Levy and Wiener (1998).

We show that various experiments strongly suggest that there is decreasing absolute risk aversion (DARA) and that there is also support, albeit not as strong as for the DARA, for a constant relative risk aversion (CRRA). These two properties lead to a myopic (power) utility function. We also use experimental results to estimate the range of the risk aversion coefficient of the myopic utility function. We find that the risk aversion parameter is typically in the range 0.6 to 2.0. This information regarding individuals' preferences is employed later on in the LLS microscopic simulation model.

3.2. ARROW AND PRATT RISK PREMIUM AND THE SUBJECT'S WEALTH

It is commonly accepted that virtually all investors are risk averters. The risk-averse set of utility functions $\mathbf{U}_2$ (when $U \in \mathbf{U}_2$ if $U' > 0$ and $U'' \leq 0$) is very wide and includes many irrelevant utility functions. Indeed, Arrow (1965) and Pratt (1964) suggest restricting the risk-averse utility function set $\mathbf{U}_2$ even further. They suggest measuring the risk premium in two ways: an "absolute measure" and a "relative measure." The absolute risk-aversion measure, $R_A(W)$ and the relative risk-aversion measure $R_R(W)$ suggested by Arrow and Pratt are defined, respectively, as[2]

$$R_A(W) = -U''(W)/U'(W)$$

and

$$R_R(W) = -WU''(W)/U'(W)$$

where U is the utility function ($U \in \mathbf{U}_2$) and W denotes the investor's wealth. If we have $\partial R_A(W)/\partial W < 0$, we assert that there is a decreasing *absolute* risk aversion (DARA). Similarly, if $\partial R_A(W)/\partial W = 0$, we have constant absolute risk aversion (CARA); and if $\partial R_A(W)/\partial W > 0$, we have an increasing absolute risk aversion (IARA). By a similar way, taking the derivative of $R_R(W)$ with respect to W defines decreasing, constant, and increasing *relative* risk aversion, denoted by DRRA, CRRA, and IRRA, respectively. DARA implies that investors are willing to invest more of their money (in absolute dollar amounts) in risky prospects as they get wealthier. CRRA implies that investors maintain a constant *proportion* of their wealth invested in the risky prospect as their wealth changes. Thus, CRRA is a special case of the more general DARA.

Arrow (1971) hypothesizes that most investors are risk averters characterized by DARA and IRRA. This class is, of course, a subset of $\mathbf{U}_2$. He

[2] Actually, Arrow and Pratt defined these two measures in two different ways, but the common and relevant factors in both measures are $-U''/U'$ and the initial wealth W.

claims that the DARA "seems supported by everyday observation" (see Arrow, 1971, p. 96). However, regarding the IRRA, Arrow asserts that

> the hypothesis of increasing relative risk aversion is not easily confrontable with intuitive evidence. The assertion is that if both wealth and size of bet are increased in the same proportion, the willingness to accept the bet (as measured by the odds demanded) should decrease. The hypothesis will be defended partly by its consistency with general theoretical principles and partly by its success in explaining economic behavior. (Arrow, 1971, p. 97)

Arrow relies on the ratio of money held in cash to total wealth as supportive evidence of IRRA:

> It should be noted, moreover, that it is hard to imagine any alternative explanation for the observed secular constancy or rise in the ratio of money held to income or wealth, ... thus the notion that security in the form of cash balances had a wealth elasticity of at least one seems to be the only explanation of the historical course of money holding. (Arrow, 1971, pp. 103–104)

While decreasing absolute risk aversion is widely accepted, less agreement exists regarding relative risk aversion. While Arrow (1971) and others tend to believe that IRRA is reasonable to assume, other researchers think that it is more reasonable to assume CRRA. The main proponents of CRRA are Latané (1959), Hakansson (1971), and Markowitz (1976).

Latané (1959) suggests a theoretical argument supporting the logarithmic utility function, which is a special case of CRRA, with relative risk aversion equal to 1. The argument for the log utility function relates to the investment horizon. In particular, it is theoretically claimed that when investing for the long run (i.e., the investment holding period is very long), the investment strategy that maximizes the geometric mean *almost certainly* outperforms any other investment policy. Maximizing the geometric mean implies a log utility function. Thus, it is claimed that long-run investors should be characterized by the log utility function. This conclusion is challenged by Samuelson and Merton (1975), who claim that for a power utility function ($U(W) = \frac{W^{1-\alpha}}{1-\alpha}$), the investment policy in each period is independent of the number of periods employed. Hence, they theoretically show that the conclusion regarding the maximization of the geometric mean (and hence of the log preference) is wrong. Yet they themselves employ a power utility function, which also implies CRRA (but with relative risk aversion not necessarily equal to unity, as implied by the log utility function). Regardless of who wins in the geometric mean debate, both sides employ CRRA functions.

Friend and Blume (1975), using cross-sectional survey data regarding assets held by households, conclude

> First, regardless of their wealth level, the coefficients of proportional risk aversion for households are well in excess of one and probably in excess of two...

> Second, the paper concludes that the assumption of constant proportional risk aversion for households is as a first approximation a fairly accurate description of the market place. These first two findings imply that the utility function of a representative investor is quite different from those which have often been assumed in the literature. However, it should be pointed out that our conclusion of constant proportional risk aversion follows from our treatment of investment in housing. Other plausible treatments would imply either moderately increasing or moderately decreasing proportional risk aversion. (Friend and Blume, 1975, pp. 900–901)

Though cross-sectional survey data should not be employed to test the existence of absolute and relative risk aversion, the study of Friend and Blume provides a rough estimate regarding these two measures. DARA is observed everywhere, and CRRA is supported, but moderately increasing or moderately decreasing relative risk aversion are not ruled out.

Drawing conclusions from cross-sectional data or from the aggregate holding of assets across time is subject to criticism. Let us elaborate on this issue and explain the importance of experimental studies in such tests. If DARA and IRRA indeed exist, and the investor has to diversify between one risky asset and the riskless asset, it can be proven that as wealth increases, the *amount* of money invested in the risky asset increases, and the *proportion* of wealth invested in the risky assets decreases. Testing these two properties as implied by DARA and IRRA seems to be easy but it is not. Actually testing empirically the properties of DARA and IRRA is very difficult because the investment behavior of the *same person* should be examined at various levels of wealth. Thus, in principle, to conduct such a study one has to wait until the wealth of the subject changes substantially and then examine the investment strategy employed at different levels of wealth. Therefore, the analysis of the absolute and relative risk aversion measures should be done with time-series data and not with cross-sectional survey data. Recall that we wish to analyze the effect of change of wealth on the investment behavior for a given preference and not to analyze these changes for all individuals together. Thus, Arrow's observation regarding the proportion of wealth held as cash is an aggregate observation and not an observation regarding a single person's preference. Therefore, it cannot be used to test the absolute and relative risk aversion. Moreover, even if one waits for years and follows the investment behavior of a given investor as his or her wealth changes, one cannot isolate the wealth effect alone, because with growing age preference may also change, regardless of the level of wealth. Since such time-series data are rarely available, cross-sectional empirical studies are employed (as done by Friend and Blume, 1975). The implicit assumption of such cross-section studies is that there is one representative preference and that the various levels of wealth represent various points on this preference function.

Experimental studies come to the rescue in the investigation of absolute and relative risk aversion. By giving the same subject different levels of wealth or by playing a sequential series of games where the wealth changes from one game to another (due to the returns obtained from the investment), absolute and relative risk measures can be tested. Actually, the analysis of such issues provides an excellent example of the importance of experimental studies. The empirical tests of these issues are difficult if not impossible and experimental studies can be easily constructed and employed, with no need to wait years in order to test the change of the same investor's investment behavior for various levels of wealth. Moreover, in the experimental studies one can change the wealth in a relatively short period of time with no risk that preference will change with aging. Nevertheless, experimental studies are not yet widely employed in finance. In particular, only a few experiments have directly tested the DARA and IRRA hypotheses, and we will discuss each of them separately.

3.3. THE GORDON, PARADIS, AND RORKE EXPERIMENT

Gordon, Paradis, and Rorke (1972) were the pioneering researchers to conduct laboratory tests to analyze various issues in modern finance. They tested the investment allocation between a risky asset and a riskless asset for various levels of wealth, as well as the choice among various assets. They employed the mean-variance rule and the capital market line (CML) to analyze investment efficiency or inefficiency as selected by the subjects participating in the experiment. In this section we discuss their findings regarding DARA and IRRA, and their other findings will be discussed in the next chapter. Gordon *et al*. analyze the dollar amount and the fraction of the investment allocated to the risky asset and to the riskless asset. Following Arrow (1971), their hypothesis is that the *amount* invested in the risky asset increases with wealth, hence the investor has decreasing absolute risk aversion, and that the *fraction* of wealth invested in the risky asset decreases with wealth, implying increasing relative risk aversion (IRRA).

Thirty-four second-year MBA students at the University of Toronto participated in the experiment. Each subject had 11 trials (i.e., 11 investment decisions); altogether $34 \times 11 = 374$ investment decisions were made. Each subject was given an initial wealth that ranged from \$110,000 to \$190,000 (paper money). The subjects could borrow or lend at a zero interest rate. There were restrictions on the amount of borrowing to prevent bankruptcy (i.e., that the net wealth would never be negative even if the worst-case scenario occurred). The subjects could choose only *one* game (one bet) on each trial and mix it with the riskless asset; however, diversification among bets (on a given trial) was not allowed.

Table 3.1 provides the five gambles from which the subjects could select. The payoff given in Table 3.1 is per \$1 of investment. To illustrate,

TABLE 3.1 Investment Alternatives

Gamble number	Payoffs*		Probability	
	Red	Black	Red	Black
1	$1.30	$.80	.5	.5
2	1.50	.70	.5	.5
3	1.90	.40	.5	.5
4	2.50	.00	.5	.5
5	100.00	.00	.005	.955

* Amount investor receives per dollar played.
Note: Table taken from Paradis *et al.*, 1972, with permission.

assume that a subject had invested $100,000 in gamble number 1 (i.e., selected to bet on this game); he or she would then receive $130,000 or $80,000 with an equal probability.

The subjects made choices and their wealth was recorded. Then, the diversification policy between the risky and the riskless assets adopted by each subject was recorded. The amount of dollars as well as the proportion of wealth invested in the risky asset and the riskless asset were investigated for various wealth levels. By observing the subject's wealth, the amount of money invested in the risky asset, and the proportion of wealth invested in the risky asset, one can check whether DARA and IRRA prevail as hypothesized by Arrow.

In analyzing DARA and IRRA, the risky asset must be the same for all levels of wealth. Most choices of the subjects were for game number 2. Hence, Gordon *et al.* analyzed DARA and IRRA corresponding to one specific risky asset, which is game number 2. Table 3.2 provides the results regarding absolute and relative risk aversion of those subjects who selected game number 2. As the table shows, the higher the amount of wealth W, the larger the mean amount invested in the risky asset, which strongly supports the DARA hypothesis. Regarding the IRRA, CRRA, or DRRA, there are no clear-cut results. Initially, the fraction of wealth invested in the risky asset tends to decrease as wealth increases, from which one tends to conclude that IRRA prevails. However, even Gordon *et al.* are hesitant regarding the IRRA conclusion, because the mean fraction of the investment in the risky asset is not monotonically declining. It strongly declines at relatively low levels of wealth, but from a wealth of $100,000 and above, it tends to stabilize and even increases in the $225,000 to $250,000 wealth interval (from 0.56 to 0.76), as well as in the wealth intervals of $275,000 to $300,000 and $350,000 and over. Gordon *et al.* conclude regarding IRRA, "There is also a clear tendency for the fraction of wealth played to decline as wealth increases at least up to a wealth of $200,000" (Gordon *et al.*, 1972, p. 110). Gordon *et al.* did not provide any significant test for the mean fractions given in Table 3.2. From the data available, it seems to us

TABLE 3.2 The Dollar Amount and the Fraction of Wealth Invested in the Risky Asset

Wealth class, W	Frequency of decisions	Mean wealth	Mean amount invested in the risky asset	Mean fraction of wealth invested in the risky asset
Less than 25,000	16	$15,750	$38,594	2.51
25,000–50,000	20	38,180	71,975	1.92
50,000–75,000	20	62,360	81,850	1.31
75,000–100,000	31	88,503	93,332	1.07
100,000–125,000	39	112,077	96,428	.86
125,000–150,000	29	134,779	114,341	.85
150,000–175,000	33	162,852	126,827	.78
175,000–200,000	10	184,530	128,400	.70
200,000–225,000	9	212,067	121,778	.58
225,000–250,000	14	263,083	149,167	.57
250,000–275,000	6	263,083	149,167	.57
275,000–300,000	5	287,100	177,200	.61
300,000–325,000	6	309,333	159,767	.52
325,000–350,000	5	336,520	156,640	.47
350,000 and greater	10	431,440	251,980	.58

Note: Table taken from Paradis *et al*., 1972, with permission.

that for a wide range of wealth levels, CRRA describes the subjects' behavior better than IRRA. Moreover, it seems that the sharp drop in the proportion of wealth invested in the risky asset at low levels of wealth is an artifact of the experimental design. The reason for this interpretation of Gordon *et al*.'s results is that in the Gordon *et al*. experiment there was no financial penalty for losses. Let us elaborate.

The subjects did not get any monetary payoff and could not win or lose real money in this experiment. The only thing that was probably important to the subjects was their prestige and the comparison of their achievement with the other subjects in class. Thus, at a very low level of wealth, the subjects had nothing to lose, therefore they tended to borrow and to invest more than 100% of their initial wealth in the risky asset. This explains the fractions of investment in the risky asset that are greater than 1 given in Table 3.2. If the subjects could also lose out-of-pocket money, they would probably be more careful with borrowing, hence the investment proportion in the risky asset at the lower level of wealth would be much smaller, and the observed IRRA at relatively low wealth levels would probably disappear.

Finally, note that absolute and relative risk aversion should be measured with a given utility function (a given person) for various levels of wealth. Table 3.2 reports the mean values for all 34 subjects, which presumably are characterized by different preferences. Thus, these results are accurate only on the assumption that all subjects have the same preference. Gordon *et al.* are aware of this deficiency,

> but it cannot be ascertained from the data whether this is due to covariation between play and wealth for given participants or due to covariation between play and wealth among participants. (Gordon *et al.*, 1972, p. 111)

3.4. THE KROLL, LEVY, AND RAPOPORT EXPERIMENT

Kroll, Levy, and Rapoport (1988a) conducted an experimental study where the subjects selected one out of two available risky assets and the riskless asset. The various portfolio games were identical, apart from the fact that the initial wealth varied in the range of \$500 to \$2500; hence, DARA and IRRA, CRRA, or DRRA could be tested. There were two sessions and 15 subjects. In this study DARA and IRRA were tested more directly—namely, the behavior of *each subject* at various levels of wealth was analyzed separately. Hence, covariation among the subjects did not affect the results of this experiment (as occurs in the Gordon *et al.* study).

Table 3.3 presents the correlations between the *amount* of investment capital and the *amount* of capital invested in the riskless asset. Capital borrowed is given a negative sign. There are 15 subjects and two sessions,

TABLE 3.3 Correlations between Amount of Investment Capital Wealth and Amount of Capital Investment in the Riskless Asset

Subject	Session 1	Session 2
1	−0.46**	−0.46**
2	−0.26*	−0.96**
3	0.09	−0.84**
4	−0.42**	−0.56**
5	−0.54**	−0.72**
6	−0.75**	−0.88**
7	−0.98**	−1.00**
8	−0.13	−0.21
9	−0.75**	−0.84**
10	−0.42**	−0.07
11	0.25	−0.06
12	−0.96**	−0.97**
13	−0.67**	−1.00**
14	−0.34**	−0.15
15	−0.62**	−0.25

* Significant at $p < 0.01$.

** Significant at $p < 0.001$.

Note: Table taken from Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission.

hence the relationship between the investment in risky assets and wealth can be tested 30 times.

As shown in the table, 28 of the 30 correlations are negative, indicating that the larger the wealth, the smaller the amount invested in the riskless asset (hence the larger the dollar amount invested in the risky asset), strongly supporting DARA. Twenty-two of the 28 negative correlations are significant at the 0.01 level. The remaining positive correlations are small and insignificant.

Table 3.4 displays the correlations between the amount of investment capital and the *proportion* of capital invested in the riskless asset by each subject. The correlations are positive in 26 of 30 possible cases. Fifteen of the 30 correlations are significant. Thus, we have 15 correlations supporting the IRRA hypothesis and 15 correlations that are not significantly different from zero, supporting the CRRA. Recall that if CRRA prevails, the proportion of wealth invested in the riskless asset should be constant, and the correlation is not expected to be significantly different from zero. Taking these results at face value, one would conclude that 50% of the population are characterized by CARA and 50% by IRRA. However, a deeper analysis shows that there is no solid ground for the IRRA results, and CRRA is probably the dominating result. In other words, we claim that the results reported in Table 3.4 are biased in favor of IRRA for exactly the same reason that explains the bias in the Gordon *et al*. study; in the Kroll *et al*. experiment the subject could win money (according to his or her performance) but could not lose out-of-pocket money. Hence, at a

TABLE 3.4 Correlations between Amount of Investment Capital and Proportion of Capital Invested in the Riskless Asset

Subject	Session 1	Session 2
1	0.45**	0.41**
2	0.53**	0.00
3	0.05	0.48**
4	0.71**	0.61**
5	0.57**	0.45**
6	0.70**	0.34**
7	−0.11	−0.22
8	0.03	0.14
9	0.24	0.24
10	0.12	0.18
11	0.59**	0.59**
12	−0.03	0.41**
13	0.71**	−0.12
14	0.40**	0.14
15	0.08	0.25

** Significant at $p < 0.001$.

Note: Table taken from Kroll *et al*., 1988a, Copyright Academic Press, Inc., with permission.

low wealth level subjects have nothing to lose, and the optimal policy is to borrow as much as one can and to invest in the risky asset. This is not the case at a higher level of wealth because the subjects may lose the financial reward already accumulated. Thus, investment of a higher proportion of wealth in the risky assets at lower wealth levels artificially contributes to the IRRA. This deficiency in the Kroll *et al*. study is identical to the deficiency in the study by Gordon *et al*. A better experimental design would probably reduce or completely eliminate the IRRA result. This is indeed the case in the experiment discussed in the following section, in which financial penalty exists and the IRRA result completely disappears.

3.5. DARA AND IRRA WHEN FINANCIAL REWARDS AND PENALTIES ARE POSSIBLE

3.5.1. The Game's Wealth

In an experiment with 62 subjects and 20 risky assets, Levy (1994) examined the subjects' investment behavior in a realistic market atmosphere. He examined the absolute and relative risk aversion measures of each of the subjects. In this experiment, each subject was allocated $30,000 in paper money, and this wealth changed as a result of changes in the asset prices and the investment policy of the investor. Unlike most other experiments, in this experiment the market price of the risky asset was determined collectively by the demand-supply function for stocks provided by the subjects. There were 10 trading rounds, hence one can analyze the relationship between wealth level and risk taking for each subject, which is essential for the examination of risk-aversion properties. A unique feature of this experiment (unlike the two experiments discussed previously) is that the subjects could lose out-of-pocket money or make money, depending on their performance. Hence, there is no incentive to artificially borrow at a low level of wealth, as in the Gordon *et al*. and the Kroll *et al*. experiments. The subjects were told at the beginning of the experiment to check and report their savings and other income resources to the experiment manager and to make sure that they had enough money to cover possible losses in case they chose to borrow intensively and to adopt an aggressive investment policy in stocks. Most of the subjects were about 25 years old and most were employed either part time or full time by the private sector or by the government.

At every period of the experiment, information was recorded on the k^{th} investor's net wealth W_k, on his or her portfolio composition, and, in particular, on the proportion of the investor's wealth allocated to the riskless asset. To test whether DARA holds, we run for each subject k the following time-series regression:

$$S_{k,t} = \gamma_0 + \gamma_1 W_{k,t-1} + e_{k,t}$$

where $S_{k,t}$ is the *amount* of dollars invested in *risky* assets (stocks) by the k^{th} investor in period t, $W_{k,t-1}$ is the k^{th} investor's net wealth in period $t-1$, namely just before the next investment decision is made, $e_{k,t}$ is the residual term, $t = 1, 2, \ldots 10$ is the trading round, and $k = 1, 2, \ldots 62$ is the subject number. If DARA holds, we expect γ_1 to be significantly positive—that is, the more wealth the investor has, the less risk-averse he or she is in absolute terms and, hence, the larger the *amount* of money that is invested in the risky assets.

Table 3.5 provides the DARA results. Out of 62 subjects, 49 behaved according to DARA and only 13 contradicted this property. However, at a 5% significance level, only 38 out of the 62 values of the γ_1 coefficients are significant, 34 are significantly positive, and only 4 are significantly negative. Thus, one can safely conclude from Table 3.5 that DARA characterized the behavior of most subjects. Only four subjects out of 62 significantly contradicted this property.

Each investor initially received \$30,000 in paper money, which represents \$30 in real money. An increase or a decrease of the net wealth even by 30% in the first trading round translates to a mere increase of \$9. *A priori*, one would expect to get very poor results with DARA, since for such small changes in wealth the utility function is almost linear. Hence we expected to have most γ_1 coefficients not significantly different from zero, even if the DARA property indeed holds. Therefore, it is quite surprising that 38 out of the 62 values of the γ_1 coefficients are significant, and that 34 of those are positively significant, supporting the DARA hypothesis. The interpretation of these results is that either the preferences are indeed sensitive to relatively small changes or that the subjects created their own frame and in their decision making treated the paper money as real money. Nevertheless, both interpretations provide strong support for the DARA hypothesis.

TABLE 3.5 The Absolute Risk-Aversion Coefficient γ_1: $S_{k,t} = \gamma_0 + \gamma_1 W_{k,t-1} + e_{k,t}$

	Number	Number of significant contributions*	Average R^2
Positive γ_1	49	34	.58
Negative γ_1	13	4	.33
Total	62	38	.53

* At a 5% significance level.

Note: Table taken from Levy, Copyright 1994 Kluwer Academic Publishers, with permission.

The IRRA hypothesis is tested in a similar way. If this property holds, the larger the k^{th} investor's wealth W_k, the smaller the investment *proportion* allocated to risky assets (stocks) and the larger the proportion of wealth allocated to bonds. To test this hypothesis, we run for each investor k the following regression:

$$R_{k,t} = \gamma_0 + \gamma_1 W_{k,t-1} + e_{k,t}$$

where $R_{k,t}$ is the *proportion* (or percentage) of wealth invested in the *riskless* asset in period t by the k^{th} investor ($R_{k,t} > 0$ implies lending, $R_{k,t} < 0$ implies borrowing), with the other variables defined as before. If the IRRA property holds, one expects the coefficient γ_1 to be positive, since the larger the investor's wealth, the more safety he or she wants, hence a larger *proportion* is expected to be invested in the riskless asset. Once again, since in this experiment there is a separate time-series regression for each subject, there are 62 regressions. Table 3.6 summarizes the results of these regressions.

The results of Table 3.6 do *not* support the IRRA hypotheses. There are 41 negative coefficients and only 21 positive coefficients, hence about two-thirds of the subjects took more proportional risk with an increase in their wealth and invested a lower proportion of their wealth in the riskless asset, which contradicts the IRRA hypothesis. About half of the coefficients in each category are statistically significant. Thus, the IRRA is rejected, because out of 62 subjects, only 11 show a significant IRRA behavior. The other 51 subjects reveal either CRRA or DRRA. These results are in sharp contradiction to what was hypothesized by Arrow (1971, p. 97) and what was found in the previous two experimental studies.

To sum up these results, about half of the coefficients (29 out of 62) are not significantly different from zero, supporting CRRA. The other half (see significant coefficients) either supports IRRA or DRRA, with more falling in the DRRA category.

TABLE 3.6 The Relative Risk-Aversion Coefficient γ_1: $R_{k,t} = \gamma_0 + \gamma_1 W_{k,t-1} + e_{k,t}$

	Number	Number of significant contributions*	Average R^2
Positive γ_1	21	11	.43
Negative γ_1	41	20	.38
Total	62	33	.40

* At a 5% significance level.

Note: Table taken from Levy, Copyright 1994 Kluwer Academic Publishers, with permission.

3.5.2. Actual Wealth and the Game's Wealth

Each subject has real wealth in the form of property (apartments, cars, savings, etc.) as well as income from work. The subjects also have their wealth from the game. Thus, it is interesting to see whether the real wealth, which is much larger relative to the wealth from the game, affects the investment behavior. To be more specific, one can test the effect of the actual wealth as well as the wealth from the game on the relative risk aversion property by running the following cross-sectional regression:

$$R_k = \gamma_0 + \gamma_1(RW)_k + \gamma_2 I_k + \gamma_3 W_k + e_k$$

Similarly, one can test the DARA hypothesis by running the following regression:

$$S_k = \gamma_0 + \gamma_1(RW)_k + \gamma_2 I_k + \gamma_3 W_k + e_k$$

where $k = 1, 2 \ldots 62$, R_k is the *percentage* invested in the *riskless* asset by the k^{th} investor, $(RW)_k$ and I_k stand for the real wealth and monthly income of the k^{th} investor, respectively, W_k is the k^{th} investor's wealth in the game, and S_k is the dollar *amount* invested in the *risky* asset by the k^{th} investor. Note that to test DARA and IRRA we advocate that time series regression is more appropriate. However, when actual wealth and actual income of the subjects serve in the regression as explanatory variables, there is no way but to run cross-sectional regressions because these two variables are constant for each subject and are kept constant for all trading rounds. Thus, like other cross-sectional results, the results reported here have some qualifications. Since these are cross-sectional regressions, we run them once for each trading round for a total of 10 times. The results are given in Tables 3.7 and 3.8.

Table 3.7 corresponds to the relative risk aversion regression test. It reveals that 8 out of the 10 γ_1 coefficients are negative; hence there is a tendency to decrease the proportion of riskless asset with larger real wealth, a result that seems, with the actual wealth as an exploratory variable, to contradict the IRRA hypothesis. The same result holds for γ_2: 8 out of the 10 coefficients are negative. However, none of the γ_1 or γ_2 coefficients are significantly different from zero, which lends support to the hypothesis that neither actual wealth $(RW)_k$ nor the actual income I_k significantly affect the investment decision. A completely different result is obtained for the experiment wealth (i.e., for γ_3). This coefficient is negative in 9 out of the 10 cases and is significant at a 5% significance level in 8 out of the 9 negative cases. This implies that the larger the wealth from the experiment, the smaller the wealth proportion allocated to the safe assets, in contradiction to the IRRA hypothesis. Thus, in the cross-sectional analysis (which gives some indication of absolute and relative

TABLE 3.7 The Cross-Section Regression Results: $R_k = \gamma_0 + \gamma_1(RW)_k + \gamma_2 I_\kappa + \gamma_3 W_k + e_k$

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Round 9	Round 10
R-square	0.42	0.23	0.02	0.12	0.47	0.36	0.44	0.41	0.28	0.32
R^2-adjusted	0.37	0.17	−0.06	0.05	0.43	0.31	0.40	0.36	0.22	0.26
γ_1	−0.000001	1.0E-06*	−1.0E-06	−2.0E-06	−3.8E-06	−3.5E-06	−2.8E-06	−2.6E-06	−2.7E-07	8.5E-07
T-value	−1.08	0.84	−0.12	−0.32	−1.65	−1.54	−1.24	−1.25	−0.11	0.32
γ_2	0.0001	−9.62E-05	7.7E-06	−0.0002	−0.0001	−0.0002	−0.0003	−0.0002	−0.0006	−0.0004
T-value	1.17	−0.49	0.006	−0.23	−0.31	−0.55	−0.83	−0.62	−1.79	−1.14
γ_3	0.0003	−4.2E-05	−5.5E-05	08.1E-05	−2.1E-05	−.1E-05	−1.7E-05	−.1E-05	−5.6E-06	−1.1E-05
T-value	4.98	−3.19	−0.84	−2.18	−5.52	−4.17	−5.11	−4.74	−3.01	−3.70

* E-06 means multiplied by 10^{-6}.

Note: Table taken from Levy, Copyright 1994 Kluwer Academic Publishers, with permission.

TABLE 3.8 The Cross-Section Regression Results: $S_k = \gamma_0 + \gamma_1(RW)_k + \gamma_2 I_\kappa + \gamma_3 W_k + e_k$

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Round 9	Round 10
R-square	0.38	0.44	0.04	0.17	0.93	0.96	0.92	0.96	0.96	0.90
R^2-adjusted	0.32	0.39	−0.03	0.10	0.92	0.95	0.91	0.95	0.95	0.89
γ_1	−5.33	3.12	1.14	8.29	−14.07	−9.45	−1.10	−2.64	5.98	−25.90
T-value	−1.19	0.48	0.02	0.17	−0.67	−0.49	0.04	−0.13	0.27	−0.86
γ_2	0.03	−0.03	0.04	2.11	0.42	0.37	0.37	0.30	0.23	0.17
T-value	1.09	−0.85	0.12	0.35	3.24	3.13	2.36	2.46	1.72	0.92
γ_3	−9.49	2.24	3.15	5.16	4.81	3.48	4.46	3.59	2.95	3.95
T-value	−4.57	5.21	1.19	2.60	21.70	27.64	19.58	27.84	27.96	17.69

Note: Table taken from Levy, Copyright 1994 Kluwer Academic Publishers, with permission.

risk-aversion measures, subject to the preceding criticism of cross-sectional analysis), we strongly reject the IRRA hypothesis. On the contrary, there is support for the DRRA hypothesis.

Table 3.8 reports the absolute risk aversion results. As in the IRRA test, we find, once again, that the experiment wealth is the main explanatory variable of the subjects' behavior. The coefficient γ_1 is nonsignificant in all 10 rounds. γ_2 is positive in 9 out of 10 rounds of trade, but it is significant in only 4 rounds. Thus, to some extent, the real income does affect the investment decision: the larger the income, the greater is the tendency to invest more in the risky assets, supporting DARA. However, the main factor explaining the subjects' behavior is the experiment wealth: in 8 out of the 10 rounds of trade, γ_3 is positive and highly significant with relatively large T-values. Thus, the results reported in Table 3.8 strongly support DARA, and the experiment wealth is the strongest explanatory variable.

With the qualifications that we employ cross-sectional rather than time series analysis, we can draw the following conclusions from Tables 3.7 and 3.8:

a. The actual wealth and income have a small or no impact on the investment decision.
b. DARA is strongly supported.
c. IRRA is rejected with a tendency to support DRRA.

Testing the absolute and relative risk aversion should be analyzed with one risky asset and one riskless asset, where the risk profile of the asset is kept unchanged. In this experiment there are 20 risky assets and the level of risk taken can be charged in two ways: (a) by changing the proportion invested in the riskless asset and (b) by changing the composition of the risky assets. To focus on the risk taking, we also regress the standard deviation of the portfolio selected by each subject on the subject's experimental wealth. Thus, the regression is of the form

$$\sigma_{k,t} = \gamma_0 + \gamma_1 W_{k,t-1} + e_{k,t}$$

where $\sigma_{k,t}$ is the standard deviation of the k^{th} investor's portfolio on date t and $W_{k,t-1}$ is his or her wealth on date $t-1$. If IRRA exists, we expect γ_1 to be negative, because the higher the wealth, the less risk is taken. The results of this test are similar to the results reported previously. Out of the 62 subjects, 39 reveal positive γ_1, hence contradicting IRRA, and only 23 reveal negative γ_1. Out of the 62 coefficients, only 6 are positively significant, supporting IRRA, and 17 are negatively significant, contradicting IRRA and supporting DRRA.

Overall, only 23 of the γ_1 coefficients are significant, while 39 are not significantly different from zero, which is consistent with CRRA. Thus, when we correct for the fact that many risky assets are involved, CRRA is the dominating feature of the subjects' behavior. This result also occurs when actual wealth and income, as well as the experimental wealth, serve as explanatory variables. The coefficients of actual wealth and income are insignificant in all 10 rounds (recall that with actual wealth and income we switch from time series regression to cross-sectional regression), and the coefficient of the experimental wealth is positive in 9 out of the 10 rounds and negative in 1 round. In 7 out of the 10 rounds, the coefficient is insignificant, namely, not different from zero, confirming once again CRRA. To sum up, correcting for the fact that many risky assets are involved once again shows that the actual wealth and actual income either show CRRA (because the coefficients are not significantly different from zero), or what may be more reasonable and consistent with previous results is that actual wealth and income were completely isolated from the investment decision making in the game.

The findings that the actual wealth and income have no effect on the investment decision are especially strong in light of the fact that the actual wealth was on average \$35,641 and the average monthly income was \$962, while the wealth from the experiment was on average less than \$100 per subject. How can one account for this result? There are two possible explanations:

a. The subjects created their own "frame" and probably treated the paper money (more than \$30,000) as real money.

b. Classical analysis of decision making under risk advocate that one should compare the distributions resulting from integrating the prospects under consideration with the rest of the assets. Thus, investors should prefer prospect X over prospect Y only if $W + X$ is preferred to $W + Y$, where W is the wealth held by the investor and X and Y are the returns on the two prospects, respectively. Therefore, by expected utility theory, wealth and income should affect the investor's decision. Tversky (1989) claims that in sharp contradiction to expected utility and portfolio theories, people tend to segregate their decisions and evaluate risky projects in terms of X and Y rather than in terms of the final wealth, $W + X$ and $W + Y$. Moreover, people tend to undertake "mental accounting" and make separate decisions rather than combining the outcomes from all decisions, as theory tells us they should do. It seems that in Levy's experiment the investors behave as if they have separate "accounting departments" for their actual wealth and their wealth from the game. Even though, theoretically, investors should treat all assets together, the subjects seem to separate their decisions, and each "department" has its own records as well as its own investment decisions—that is, when subjects played the game, they completely ignored the other wealth they had; the utility is defined, so to speak, only on the game's wealth. The analogy to actual investment is that when investors select their own investment in stocks and bonds, for example, they ignore the wealth they hold in the form of houses, cars, human capital, and so on.

3.6. THE IMPLICATION OF THE FINDINGS REGARDING PREFERENCES TO MICROSCOPIC MODELING

First, actual wealth and income, as well as assets from other sources, are treated separately. This "mental department" behavior can simplify microscopic modeling. In the LLS microscopic simulation model, which is described in Chapter 7, we assume that investors make decisions in the capital market (i.e., allocate a fraction of their wealth to the capital market), and other assets (e.g., housing, cars, human capital) are factors that can be ignored in their investment decisions. The "mental accounting" hypothesis suggested by Tversky (1989) was supported by the experiment conducted by Levy (1994).

In a microscopic simulation model, one needs to assume a preference that the subject tries to maximize. There are theoretical and empirical arguments for various preferences. If the investment is for a very long period of time, some researchers argue that on theoretical grounds it is optimal to maximize the geometric mean in each period, implying a log utility function. Other researchers argue theoretically against the geometric mean rule, yet they use power (myopic) utility functions that include the log preference as a special case. Regardless of whether a myopic or log utility function is used, both imply DARA and CRRA. The only difference between log preference and other power utility functions is the value of the relative risk aversion coefficient.

The other two approaches to analyze the shape of investors' preferences are empirical and experimental. The empirical study of Friend and Blume supports the DARA and the CRRA with a relative risk-aversion coefficient of about 1 to 2. However, the cross-sectional survey can only give us an indication of the risk measures, because a more precise test must be done with time series data.

Indeed, in several experimental tests, a time series analysis was conducted. Gordon *et al.* (1972) and Kroll, Levy, *et al.* (1988a) strongly support the DARA and more mildly support either the IRRA or CRRA. Taking into account the fact that no financial penalty exists in these two experiments, we conclude that the observed tendency for IRRA is an artifact of the experimental design and that CRRA is probably a better description of preferences, as obtained in the Gordon *et al.* study for wealth levels of $200,000 and more. The only experiment known to us that tests absolute and relative risk aversion, where financial penalties and rewards exist, is the one conducted by Levy (1994).

This study found once again that there is strong evidence for the DARA hypothesis, but the IRRA hypothesis is strongly rejected. Investors tend to show decreasing relative risk aversion (DRRA) or, at best, constant relative risk aversion (CRRA), but by no means IRRA. This evidence enhances Arrow's assertion that DARA is observed in the daily behavior of investors and that IRRA has less intuitive evidence. Thus, there is an agreement regarding DARA and there is disagreement regarding IRRA, with a tendency to support CRRA. In the time-series analysis, which is more relevant for such type of tests, the dominating risk characteristics are DARA and CRRA.[3]

In an attempt to analyze the implication of the experimental findings to the shape of the investors' utility functions, let us discuss first the most commonly employed utility functions in economics and in finance:

a. The quadratic function:

$$U(W) = W - \alpha W^2$$

b. The negative exponential function:

$$U(W) = -e^{-\alpha W}$$

c. The logarithmic function:

$$U(W) = \ln(W)$$

d. The power function:

$$U(W) = W^{1-\alpha}/(1-\alpha)$$

e. The adjusted logarithmic function:

$$U(W) = \ln(W + A)$$

f. The adjusted power function:

$$U(W) = (W + A)^{1-\alpha}/(1-\alpha)$$

where α is a risk preference parameter and A is a constant.

Table 3.9 provides the features of these utility functions.

Empirical and experimental evidence suggests that preferences generally have the DARA and CRRA properties. As seen in Table 3.9, the power utility function (and the logarithmic utility function, which is a special case of the power utility function) is the only common function that has these properties.

[3] For three more experiments related to the topic discussed in this chapter, see Harrison (1986), McKee (1989), and Coady (1995).

TABLE 3.9 Pratt's Classification of Utility of Wealth Functions

	Absolute risk aversion	Relative risk aversion
Quadratic	Increasing	Increasing
Negative exponential	Constant	Increasing
Logarithmic	Decreasing	Constant
Power (myopic)	Decreasing	Constant
Adjusted logarithmic	Decreasing	Increasing
Adjusted power	Decreasing	Increasing

In light of the fact that DARA and CRRA probably best describe investors' behavior, the next question is: What are all of the possible preferences that are consistent with these two characteristics? Are there functions other than the power function that have DARA and CRRA? It turns out that the power utility function $U(W) = W^{1-\alpha}/(1-\alpha)$ is the *only* function that is consistent with these two properties. This is summarized in the following theorem:

THEOREM 3.1. *If CRRA and DARA prevail, the only possible utility function is the power function.*

Proof. First note that CRRA ⇒ DARA, and that for a power utility function, CRRA holds. What is left to prove is that CRRA implies a power utility function.

Relative risk aversion is defined as

$$R_R(W) = -WU''(W)/U'(W)$$

CRRA implies that

$$\partial R_R(W)/\partial W = 0$$

Therefore

$$R_R(W) = \alpha \ \text{(a constant)}$$

or

$$U''(W)/U'(W) = -\alpha/W$$

Taking the integral of both sides, this can also be rewritten as

$$\ln U'(W) = -\alpha \ln W + C$$

where C is an integration constant.

Define $A = e^C$, and take the exponent of both sides to obtain

$$U'(W) = AW^{-\alpha}$$

or

$$U(W) = \frac{AW^{1-\alpha}}{1-\alpha}$$

Recall that preferences are invariant to a linear transformation of the utility function; we can therefore take $A = 1$ to obtain the power utility function:

$$U(W) = W^{1-\alpha}/(1-\alpha)$$

where $\alpha > 0$ for risk averters and $\alpha < 0$ for risk lovers.

Given the power utility function, the next question is: What is the typical value of the risk aversion coefficient α? To estimate α, we use the results of Gordon *et al*. In their game number 2, the subject can either win 50% or lose 30%, and the riskless interest rate is assumed to be equal to zero. Therefore, the investor with a utility function U faces the following problem of expected utility maximization:

$$\begin{aligned} &\text{Max}\{[\tfrac{1}{2}U(W \cdot X \cdot 1.5 + W(1-X) \cdot 1)] \\ &\quad + [\tfrac{1}{2}U(W \cdot X \cdot 0.7 + W(1-X) \cdot 1)]\} \end{aligned}$$

where X, the investment proportion in the risky asset, is the decision variable. For the power utility function, this maximization problem can be rewritten as

$$\begin{aligned} &\text{Max}\Big[\tfrac{1}{2}(W \cdot X \cdot 1.5 + W(1-X) \cdot 1)^{(1-\alpha)} \\ &\quad + \tfrac{1}{2}(W \cdot X \cdot 0.7 + W(1-X) \cdot 1)^{(1-\alpha)}\Big]/(1-\alpha) \end{aligned}$$

Note that for a power utility function, W is irrelevant for the expected utility maximization, because $W^{1-\alpha}$ is a common term. Using numerical methods, we solve for each parameter $1-\alpha$, the value X that maximizes the expected utility. The results are summarized in Figure 3.1. Gordon *et al*. found that for most relevant wealth levels of \$75,000 and above, the investment proportion in the risky asset falls in the range $0.47 \le X \le 1.07$. Using the results of Figure 3.1, this implies that $(1-\alpha)$ falls approximately in the range $-0.4 \le (1-\alpha) \le 0.4$, which implies that $0.6 \le \alpha \le 1.4$. Because α is positive, risk aversion prevails. Also, note that for $\alpha = 1$, which is included in the range, the power utility function coincides with the logarithmic utility function, $U(W) = \log(W)$.

It is interesting to note that Friend and Blume (1975), who employ cross-sectional *empirical* survey data, come to a similar conclusion that there is CRRA with a parameter α in the neighborhood of 1.0 to 2.0, a result similar to our findings.

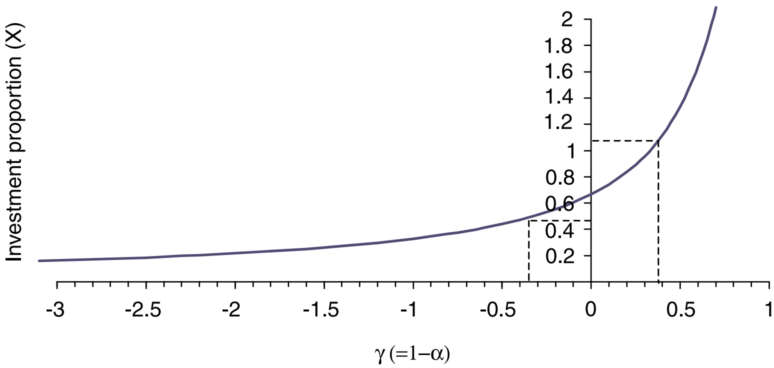


FIGURE 3.1 Investment proportion in the stock as a function of the risk aversion measure.

We use this important experimental finding in the LLS microscopic simulation model. In this model we assume a power (myopic) utility function and focus on values of α in the range found in the experiments.

3.7. SUMMARY

The results of any microscopic simulation model depend on the assumptions made. Therefore, it is important that at least regarding key features one should make realistic assumptions—that is, assumptions that do not contradict the actual investment behavior. This is especially true if one hopes to obtain quantitative predictions from the model. A key feature to which we have devoted this chapter is the preference of investors.

Empirical studies, as well as experimental studies, confirm that risk aversion dominates the market. Moreover, there is also almost complete agreement that DARA prevails. Regarding relative risk aversion measures, there are contradicting views and contradicting evidence. Latané (1959), Hakansson (1971), and Markowitz (1976) advocate the log utility function (for long-run investors), which implies constant relative risk aversion (CRRA), with a risk aversion parameter of 1. Samuelson and Merton (1975) disagree with the argument about the relationship between the investment holding period and the log function. Nevertheless, they employ a power function (see also, Samuelson, 1989, 1990, and 1997), which, once again, is characterized by CRRA. Arrow, on the other hand, speculates that IRRA prevails in the market, but he has no strong evidence or proof of this claim.

Absolute and relative risk aversion measures are very hard to test empirically, because time series studies of each individual investor should be carried out at various levels of wealth. Because such data is not available, Friend and Blume (1975) substitute the time series test with a cross-sectional test. They conclude that DARA and CRRA prevail with the relative risk aversion coefficient in the range of 1 to 2 and may even be in excess of 2.

Experimental studies by Gordon, Paradis, and Rorke (1972), as well as Kroll, Levy, and Rapoport (1988a) support DARA and, to some extent, the IRRA property. However, taking into account that in these two studies there is no financial penalty in the case of a bad performance, IRRA may be an artifact of the experimental setup. Levy (1994), who includes financial penalty and financial reward, finds DARA and strongly rejects the IRRA in favor of CRRA. When there are deviations from CRRA, he finds more DRRA than IRRA, which confirms the hypothesis that the IRRA found in the previous studies is due to the absence of financial penalty.

Based on the experimental and empirical studies, it is reasonable to assume DARA and CRRA. We proved in this chapter that the only utility function with the DARA and CRRA property is the power utility function.

The value of the relative risk aversion parameter seems to be in the range of 0.6 to 2.0. Of course, one should keep in mind that these are results relating to "typical" investors. Deviations from these values or even from CRRA and DARA may be possible.

Another property discovered in the experimental studies is that investors have "mental departments." They invest in one segment of the market while ignoring the other investment components they have. This allows us to focus in microscopic simulation models on assets available in the capital market, ignoring other assets such as human capital, houses, cars, and so on.

4

INEFFICIENT CHOICES AND INVESTORS' IRRATIONALITY

4.1. INTRODUCTION

Economic theory is based on the assumption that investors and consumers are rational and very "efficient machines," namely, that they make the best choices for themselves. Laboratory tests reveal that investors' behavior is much more complicated relative to the behavior assumed in most economic theories. As a result, a growing number of economists and psychologists are interested in the actual (or laboratory) economic behavior in contrast to the normative predictive behavior advocated by various models. Indeed, Plott (1979), Smith (1976, 1982), and Wilde (1980) argue in one form or another:

> It is important for economic science for theorists to be less own-literature orientated, to take seriously the data and disciplinary function of laboratory experiments. (see Smith, 1982, p. 924)

Most economic and financial models, particularly equilibrium asset pricing models, explicitly or implicitly assume that investors are efficient and rational. There is no clear characterization in the literature of investors' irrationality. In this book we define *weak* irrationality as deviation from expected utility maximization. We define *strong* irrationality as the (more severe) violation of the monotonicity axiom (preferring more to less). To illustrate the weak irrationality definition, consider the capital asset pricing model (CAPM). In deriving the CAPM it is assumed that all investors are risk averters and that the distributions of rates of return are normal. Assuming also full information on the various relevant parameters, and with the other standard assumptions (e.g., no taxes, no transaction costs), the CAPM risk-return relationship is derived. Suppose that indeed the preceding assumptions are intact. Then, to obtain the CAPM, one has also to assume that investors are rational—that they *act* to maximize their expected utility (they are not weakly irrational)—and also that investors are efficient, namely, they know how to achieve this goal. Thus, in the CAPM it is implicitly assumed that investors do not make mistakes and always choose their portfolios from the efficient set, namely, they choose portfolios located on the capital market line (CML). The rationality of the investors and their knowledge of how to achieve their goals are crucial to the CAPM derivation. Nevertheless, generally there is not much discussion in the financial literature of these two implicit assumptions.

We will present in this chapter evidence that some investors are either irrational or do not know how to select the best investment for themselves. In light of these findings, the crucial question is: How are equilibrium asset prices determined when the majority of investors are indeed efficient and rational, but a minority of them are irrational or inefficient? In such a case, the CAPM equilibrium model breaks down and one cannot tell what would be the assets' equilibrium prices. Microscopic simulation (MS) comes to the rescue in this case, and price determination as well as the dynamic price behavior can be investigated by these simulations.

Another example of a possible contradiction to the expected utility theory can be formulated in the stochastic dominance (SD) framework.[1] Suppose that there are two prospects, F and G, and investors have to make a choice to invest in either F or G, when F dominates G by first degree stochastic dominance (FSD). By the monotonicity axiom, the expected utility of each investor, regardless of his or her preferences, is higher with F relative to G. Yet, some investors may choose G, either because they do not know how to choose efficiently or because their goal is not to maximize expected utility. In this case investors are strongly irrational: their behavior contradicts the expected utility theory and, in particular, the monotonicity axiom asserting that the investors prefer more wealth over less wealth.

Regarding the CAPM and stochastic dominance examples, one may claim that investors may not know what expected utility is, what stochastic dominance is, what CAPM is, and so on, which implies that these theoretical models cannot be intact. However, if all the investors are rational and

[1] For Stochastic dominance rules, see H. Levy (1992) and (1998).

efficient, then they are expected to invest "as if" they know these concepts (see Friedman, 1953a). If they do not behave as if they know the theory, then the models break down. We emphasize that even if a small proportion of investors do not behave as if they know the theory, the models break down and their asset pricing relationships no longer hold. However, in such a case, MS suggests a framework to analyze the effect of such deviations from rationality or efficiency on asset pricing.

Another example of a possible violation of the theory is related to arbitrage models. Arbitrage models are very strong and the results of these models are intact even if a small or large proportion of the investors are irrational or do not behave as if they know the theory. To be more specific, in arbitrage models, to deviate from the equilibrium model's results one has to assume that *all* investors are either strongly irrational or inefficient. Let us elaborate with an illustration of the Black and Scholes (1973) option pricing model, which is derived using "no arbitrage" arguments. Suppose that there are no-transaction costs, and that the volatility of the stock, σ, is known to all investors with certainty. As the other four parameters (t, E, S, r) are given, the option must be priced correctly. Otherwise, arbitrage profit is available. Suppose now that for some investors the monotonicity property is violated (they prefer less money to more money, hence they are *strongly irrational*) or alternatively suppose that they do not know how to exploit the certain profit in case of option mispricing (hence they are inefficient). Does it imply that the Black and Scholes model will not be intact? Unlike the previous cases, the answer here is negative. It is sufficient that there is *one* efficient and rational investor to guarantee that the Black and Scholes pricing model will be intact. Otherwise, this investor can have an infinite profit with certainty! He or she will guarantee that the option will be priced correctly. Thus, to invalidate most models (like the CAPM) it is sufficient that *some* investors are weakly irrational or inefficient. To invalidate the arbitrage models, it is required that *all* investors are either inefficient, strongly irrational, or both.

Obviously, the option pricing may deviate from the Black and Scholes model even if *all* investors are efficient and rational for other reasons. For example, if the market is inefficient (transaction costs) or investors disagree on the future volatility of the stock, the Black and Scholes model breaks down. But this is not the issue we discuss in this chapter, because we are focusing on possible deviations from the classical models due to irrational choices or inefficient selection of choices by the investors, and not on the effects of market frictions. Finally, in the experiments reported later in this chapter, it is hard to distinguish between irrationality and inability to choose correctly; the end result is a wrong choice, which could be due to either of these two factors or to both. Thus, we identify errors in the subjects' choices due to these two factors combined.

In Chapter 2 we focused on experiments in which the subjects have to choose between pairs of simple bets, where each bet is usually composed of only two outcomes. This setting is too simple and does not even come close to the complex setting that exists in the capital market. Yet using this simple framework has its pros and cons: on the one hand, researchers can draw important conclusions regarding some specific characteristics of the subjects' behavior. These specific features of the subjects' behavior are difficult to test in a more complex setting. To be more specific, in the simple experimental setting of Chapter 2, one can test separately the subject's attitude toward risk, probability distortions, and the role that the initial wealth and change in wealth (in contrast to total wealth) play in the subject's decision-making process. On the other hand, the simple bets framework has a drawback; one may claim that in a more complex price setting, it is possible that other investment behavior and distortions will emerge, hence the simple setting is not representative.

In Chapter 3 we discuss the general feature of the utility function and wealth changes. In the present chapter we focus on more complex settings that gradually come very close to mimicking the actual capital market setting. We start with a relatively simple experiment, which includes one or two stocks, and build up to investigating the subjects' stock selections and the pricing of multiple risky assets where each subject's decision directly affects the other subjects' wealth. Thus, in the most complex setting, the pricing of the risky assets is determined collectively by all subjects participating in the experiment, exactly as done in the actual stock market. This chapter focuses on investors' inefficiency and irrationality, leading to risky asset pricing, with an emphasis on the risk-return relationship as determined in the marketplace by the subjects themselves.

The results found in the experiments are later employed in the LLS microscopic simulation model. For example, experiments reveal that investors look for trends in the historical rates of return, and they base their investment decisions on these historical returns. This finding is later incorporated into the LLS model, where we analyze the impact of this and other features on the dynamics of asset prices (see Chapter 7).

4.2. INVESTORS' INEFFICIENCY AND IRRATIONALITY

Do people choose their investments efficiently? What proportion of choices are clearly wrong choices? These questions are hard to answer based on empirical *ex post* data, but they are easy to analyze in a laboratory experiment. In this section, we discuss two experiments that directly address the issue of "right" and "wrong" investment portfolio choices. The main features of investors' irrationality found in the experiments will be incorporated later on in the LLS model.

4.2.1. The Gordon, Paradis, and Rorke Experiment

In the Gordon, Paradis and Rorke (1972) experiment discussed in Chapter 3, the researchers report on the right and wrong investment choices made by the subjects. We now discuss the subjects' choices in this experiment.

The subjects had to choose one of five available risky assets given in Table 3.1 (presented in the previous chapter). They could also borrow and lend at zero interest rate with some restrictions on the amount of borrowing. Namely, subjects were not allowed to lever up to a position where they could end up with negative wealth. Table 4.1 reveals the two possible outcomes for each of the five games with the maximum allowed borrowing.

Let us illustrate with the outcomes of gamble number 1. In this gamble, for each dollar invested there is an equal probability to end up with either \$1.30 or \$0.80. For an investment of \$100,000 there is an equal chance to end up with either \$130,000 or \$80,000. If the investor borrows \$400,000 and invests the \$500,000 available, he or she will end up with either \$250,000 ($= 500{,}000 \cdot 1.30 - \$400{,}000$) or zero ($= \$500{,}000 \cdot 0.8 - \$400{,}000$). The cash flow in the other gambles is calculated in a similar way. In gambles 4 and 5, borrowing is not allowed because the cash flow in a case of a loss is zero; therefore, by borrowing, negative wealth may occur, which is not allowed in the Gordon *et al.* experiment. For example, in gamble number 5, if the investor wins, he or she receives \$100 per \$1 invested (see Table 3.1 in Chapter 3); hence for \$100,000, the income will be \$10,000,000. However, if the investor loses, he or she receives nothing, and therefore, to avoid bankruptcy, borrowing is not allowed in this gamble. Gordon *et al.* analyze the efficiency of choices by the Markowitz (1952a) mean-variance (M-V) rule. However, because the outcomes of the

TABLE 4.1 Outcomes for Maximum Investment on the Five Gambles[a]

Gamble number	Amount of money invested	Wealth Outcome if Win	Wealth Outcome if Lose
1	\$500,000	\$250,000	0
2	\$333,333	\$266,667	0
3	\$166,667	\$250,000	0
4	\$100,000	\$250,000	0
5[b]	\$100,000	10^7	0

[a] The investor's initial wealth is \$100,000, and he borrows the difference between his investment and \$100,000 at a zero interest rate.

[b] Here we add the same calculation of returns for game number 5. It was not reported by Gordon *et al.*

Note: Table taken from Paradis *et al.*, 1972, with permission.

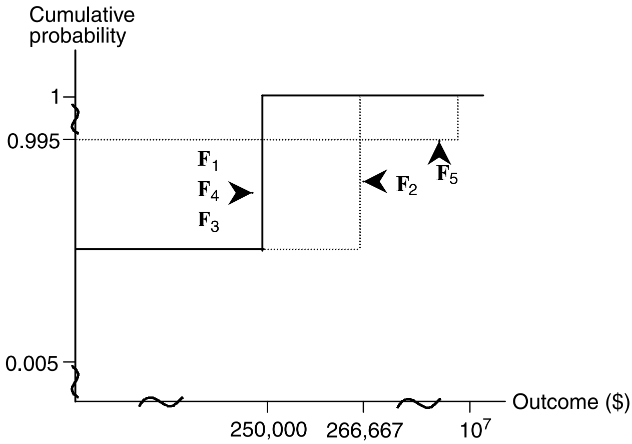


FIGURE 4.1 The cumulative distributions of the five gambles given in Table 4.1. (From Paradis *et al*., 1972, with permission).

bets are discrete, the returns are not normally distributed and the M-V criterion is not optimal. Luckily, the bets in Gordon *et al*. are very simple and we can interpret their results in the expected utility framework with no need to assume normality of the returns.

Figure 4.1 provides the cumulative distribution of the five options given in Table 4.1. As can be seen, distribution 2 dominates by FSD all other distributions with the exception of investment 5 because[2]

$$F_2(X) \leq F_i(X) \text{ for all } X \text{ and for } i = 1, 3, 4$$

For $i = 5$, the two distributions cross, hence neither gamble 2 nor gamble 5 dominate the other by FSD. Note that in the first four gambles, there is an equal probability of winning or losing, while in gamble 5 the probability of winning is 0.005 and the probability of losing is 0.995 (see Table 3.1 of Chapter 3). Note that F_2 dominates by second degree stochastic dominance (SSD).

Table 4.2 contains the subjects' choices as obtained in the Gordon *et al*. experiment. As this table shows, 108 (10 + 20 + 78) investment decisions out of 374 were wrong—namely, in about 29% of the investment decisions, an FSD inferior investment has been selected. As mentioned previously, Gordon *et al*. analyze these choices in the mean-variance

[2] Note that if there is a mix of distribution 2 with the riskless asset (denoted by 2_α), which dominates some other mix of distribution 1 with the riskless asset (denoted by 1_β), then for any mix of the option 1 and the riskless asset there is a mix of option 2 with the riskless asset, which dominates it by FSD. In such a case, we say that option 2 dominates option 1 by first degree stochastic dominance with the riskless asset, FSDR. For more details, see H. Levy (1998).

framework (which is inappropriate to employ in the absence of normality); however, their results regarding subjects' wrong decisions turned out to be correct in the expected utility framework as well. In fact, the result is even stronger than presented in their paper (hence the errors committed by the subjects are more severe) because investments 1, 3, and 4 are not inferior to investment 2 only by M-V; they are also inferior by FSD (i.e., for all utility functions). Thus, there is no need to assume normal distributions of returns and risk aversion to conclude that 108 investment decisions (29%) were wrong.

4.2.2. The Kroll, Levy, and Rapoport Experiment

Analysis in the M-V Framework

In a simple experimental setting where subjects had to choose investing either in stock A or in stock B and were also allowed to borrow and lend at a riskless rate, Kroll, Levy, and Rapoport (1988a) examined the subjects' behavior, with emphasis on the following issues:

a. It was tested whether the subjects select uncertain prospects efficiently by the mean-variance criterion. If returns are normally distributed, in the face of risk aversion, the mean-variance investment criterion coincides with the maximum expected utility criterion (see H. Levy, 1998). While Kroll *et al*. focus on the M-V criterion, we elaborate in this chapter showing that because of the possibility of bankruptcy, the distributions are not exactly normal, hence the M-V criterion is not optimal. Thus, we refine and correct the analysis of Kroll *et al*.

b. Given the information that stock price changes are drawn randomly from a given distribution, do the subjects look for past and irrelevant information, presumably to improve their investment decision?

c. In a random walk setting, do the subjects identify imaginary trends in rates of returns and make an investment decision based on these subjective trends?

TABLE 4.2 Frequency Distribution of Gambles Selected

Gamble number	1	2	3	4	5[a]	0[b]	Total
Frequency	10	253	20	78	11	2	374

[a] In every one of these cases, the individual was in lower depths. Lower depth is defined by Gordon *et al*. as a situation in which the subject's wealth is reduced to zero due to overconsumption or bad luck in the gamble.

[b] Two individuals chose not to play on one trial.

TABLE 4.3 The Structure of the Experiment by Kroll, Levy, and Rapoport

Session I	Session II
Two sets, each of 10 portfolio games (altogether 20 games)	Two sets, each of 10 portfolio games (altogether 20 games)
Five portfolio investment decisions per game (trials)	Five portfolio investment decisions per game (trials)
Total number of investment decisions	Total number of investment decisions
$2 \cdot 10 \cdot 5 = 100$	$2 \cdot 10 \cdot 5 = 100$

Note: Table taken from Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission.

In this section we discuss issue a. Issues b and c will be discussed in the next section.

The experiment was conducted with 15 male and female students. All subjects had taken at least one course of statistics; hence they were familiar with the notion of correlation, statistical independence, and the normal distribution. The subjects volunteered to serve in a computer-controlled portfolio selection experiment with a monetary payoff that was determined by their investment performance in the experiment. Each subject participated individually in two self-paced sessions, each of which lasted 3 to 90 minutes. Two to seven days elapsed between the two sessions. Each session consisted of two sets of 10 portfolio games, each with a maximum of 5 trials per game. Therefore, a session included 20 games for a maximum of 100 trials (i.e., investment decisions). The two sets differed from one another only in the initial amount of investment capital given to the subject. The wealth effect on risk taking as revealed in this experiment is discussed in Chapter 3. In this section we focus on the "right" and "wrong" investment choices.

Table 4.3 summarizes the structure of this experiment. As there were 15 students, we have altogether a maximum of 3000 investment decisions ($15 \cdot 200$, see Table 4.3). The subjects who borrowed money could go bankrupt during the experiment, and as some of them indeed went bankrupt, the number of investment decisions is slightly smaller than 3000. In a case of bankruptcy, the subject was eliminated from the experiment with a zero realized rate of return. Hence, an outcome of -150% or -250% are identical from the subject's point of view, meaning terminal zero wealth. In each trial, the subjects were allowed to borrow up to five times the wealth possessed at this stage. Thus, as in actual investment in the capital market, there are some restrictions on the amount of borrowing, which is a function of the borrower's equity. The restrictions on borrowing in the Kroll *et al.* experiment, however, are not as severe as the restrictions in the Gordon *et al.* experiment discussed earlier.

TABLE 4.4 The Mean and Standard Deviation of the Two Risky Assets

	Asset A	Asset B
Mean, μ	5%	7%
Standard deviation, σ	4%	12%

Note: Table taken from Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission.

The subjects had to choose from asset A or asset B, whose rates of return are drawn randomly (on each trial) from normal distributions with the parameters given in Table 4.4. The rate of return on the riskless asset is 3%. Subjects could mix either asset A or asset B with the riskless asset to obtain a portfolio rate of return:

$$\tilde{R}_{A,\alpha} = \alpha\tilde{R}_A + (1-\alpha)r$$

or

$$\tilde{R}_{B,\alpha} = \alpha\tilde{R}_B + (1-\alpha)r$$

where $\tilde{R}_A$ and $\tilde{R}_B$ are the rates of return on assets A and B, respectively, r is the riskless interest rate, and α is the proportion of wealth invested in the risky asset where $\alpha > 0$. If $\alpha > 1$, borrowing takes place. The subjects were not allowed to diversify between assets A and B, hence the correlation between these two assets and portfolio construction are irrelevant.

Because the returns are drawn at random from normal distributions (a fact known to the subjects), it seems that Markowitz's (1952a) well-known mean-variance (M-V) criterion is optimal. To be more specific, with normal distributions one can safely assert that if

$$\text{(a) } E(X) \geq E(Y) \text{ and (b) } \sigma_x \leq \sigma_y$$

and with at least one strong inequality, then $EU(X) \geq EU(Y)$ for all risk-averse utility functions, $U \in \mathbf{U}_2$, where $\mathbf{U}_2$ is the set of all utility functions with $U' \geq 0$ and $U'' \leq 0$ (the opposite is also true: if $EU(X) \geq EU(Y)$ for all risk-averse utility functions, then a and b in the equation hold).

Secondly, with normal distributions, if

$$\text{(a) } E(X) \geq E(Y) \text{ and (b) } \sigma_x = \sigma_y$$

then $EU(X) \geq EU(Y)$ for all nondecreasing utility functions, risk averse, as well as risk seeking ($U \in \mathbf{U}_1$ where $\mathbf{U}_1$ is the set of all functions with $U' > 0$), because in such a case (with normal distributions), X dominates

Y by FSD.[3] Kroll *et al*. analyze the results in the M-V framework, which apparently coincides with expected utility maximization. But, in fact, it does not, because the bankruptcy possibility induces a violation in the normality of the distribution. We will elaborate on this issue and suggest another analysis of the results from Kroll *et al*. Before we turn to the results, note that we do not assert that investors know this theory; however, we expect them to make the correct choices by the "as if" argument discussed before.

Figure 4.2 illustrates the various investment possibilities that the subjects faced in the M-V framework. By mixing assets *A* and the riskless asset, an infinite number of feasible portfolios are generated, all of which are located on the line segment *rA* (see Fig. 4.2). Similarly, by mixing *B* with the riskless asset, an infinite number of portfolios are generated, all residing on the line segment *rB*. Point A^* in Figure 4.2 is obtained by borrowing five dollars for each dollar of equity (the subjects could not borrow more than five times their wealth) and investing all of the six dollars in stock *A*. Similarly, point B^* is obtained by investing six dollars in stock *B*, five of which are borrowed at rate *r*. If a subject selects a portfolio on the line segment *rA* or on the line segment *rB*, he or she is diversified between one of the stocks and the riskless asset. Being exactly

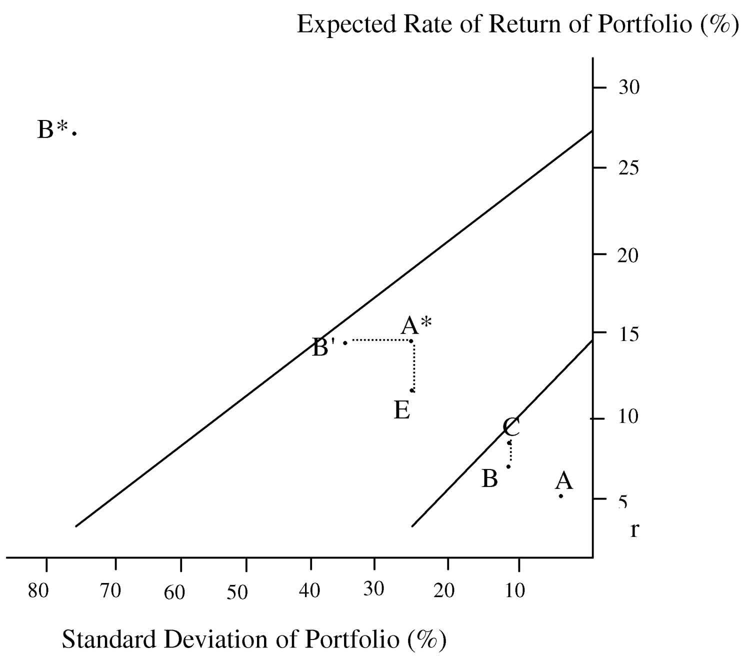


FIGURE 4.2 Efficient frontier for the two-stock experiment. (From Kroll *et al*., 1988a, Copyright Academic Press, Inc., with permission).

[3] For more details, see H. Levy (1992, 1998).

at point A or point B indicates neither borrowing nor lending. Being to the right of point A or point B on the corresponding lines indicates that a levered portfolio has been selected. Finally, point B' corresponds to the case where for each dollar of equity the subject borrows two dollars and then invests all three dollars in stock B.[4]

Employing the M-V rule, Kroll *et al.* show that not all feasible portfolios are M-V efficient as revealed in Figure 4.2. In particular, all the portfolios that lie on the segment rB' are inefficient because there are portfolios on the segment rA^* with higher (or equal) mean and smaller (or equal) standard deviation. The M-V efficient frontier consists of the line segment rA^* and $B'B^*$, not including point B'.

Are investors "efficient machines" who know how to choose efficient portfolios? To answer this question, the feasible set of portfolios was partitioned into ten categories:

1. Point r: Investing all the capital in the riskless asset.
2. Segment rA: Diversifying between stock A and the riskless asset.
3. Point A: Investing all the capital in stock A.
4. Segment AA^*: Borrowing up to five dollars for each dollar of equity and investing in stock A.
5. Point A^*: Borrowing the maximum of five dollars for each dollar of equity and investing all the capital in stock A.
6. Segment $B'B^*$: Borrowing more than two dollars (see footnote 4) per each dollar of equity and investing all funds in stock B.
7. Point B^*: Borrowing the maximum of five dollars for each dollar of equity and investing all the capital in stock B.
8. Segment rB: Diversifying between stock B and the riskless asset.
9. Point B: Investing all the capital in stock B.
10. Segment BB': Borrowing between zero and two dollars (see footnote 4) for each dollar of equity and investing all funds in stock B. Unfortunately, Kroll *et al.* do not distinguish between segments BE and EB', which, as we shall see, are of crucial importance.

Suppose that the distributions are indeed normal. Then, if one is willing to assume only that utility functions are nondecreasing (i.e., $U \in \mathbf{U}_1$ and risk seeking is also permitted), then segments rB and BE are inefficient, because for any portfolio located on this segment (e.g., portfolio B), there is a portfolio on segment rA^* (see portfolio C) with the same standard deviation and a higher mean; hence, all portfolios located on line rE are FSD inefficient.[5]

[4] Note that $E(R_{A^*}) = 6 \cdot 5\% - 5 \cdot 3\% = 15\%$, $E(R_B) = 7\%$, and $r = 3\%$ (see Table 4.4). Because at point B' we have $E(R_{B'}) = E(R_{A^*}) = 15\%$, we have $\alpha E(R_B) + (1 - \alpha)r = ER_{A^*} = 15\%$ or $\alpha \cdot 7\% + (1 - \alpha) \cdot 3\% = 15\%$, hence $\alpha = 3$, which implies that \$2 are borrowed for each \$1 of equity.

[5] Recall that for normal distribution if $\sigma_1 = \sigma_2$, FSD coincides with the dominance by the M-V rule (see H. Levy, 1992, and 1998).

For portfolios located on segment EB', there are portfolios on line rA^* with the same mean and a smaller standard deviation; hence, with normality, these EB' portfolios are inferior for all risk averters (i.e., by second degree stochastic dominance, SSD). In the study by Kroll *et al.* there is no distinction between FSD and SSD, and all portfolios on line rB' are considered inefficient because the assumption of risk aversion is commonly accepted.

Table 4.5 presents the frequencies of the various portfolio selections aggregated over subjects and games. The frequencies are classified by session (1 or 2) and category. The total number of choices is 2975 rather than 3000 because a few bankruptcies occurred during the experiment.

The most striking result in Table 4.5 is the relatively large percentage of inefficient portfolios selected. There are 366 (24.6%) inefficient M-V portfolios in session 1 and 426 (28.7%) in session 2. It is interesting that the percentage of wrong investment choices is in the same magnitude as found in the Gordon *et al.* experiment discussed earlier; it is about 29% in the Gordon *et al.* experiment and about 27% in the Kroll *et al.* experiment.

A more detailed analysis of the M-V inefficient portfolios shows that most of them fall on the line segment BB'. In a total of 550 out of 792 inefficient choices (69.4%), the subjects borrowed less than two dollars per each dollar of equity and invested the funds in stock B. Kroll *et al.* claim that the subjects did not realize that by borrowing from two to five times their starting capital for the trial and investing the funds in stock A, they could generate portfolios that were superior by the M-V rule (i.e., segment CA^* dominates segment BB', see Figure 4.2).[6]

Risk of Bankruptcy

Kroll *et al.* speculate that the subjects did not want to go bankrupt and hence avoided segment CA^*, which involves intensive borrowing. Let us analyze this possibility under the assumption that instead of expected utility maximization the subjects' goal was minimization of the risk of bankruptcy. The subjects knew that on any given trial, bankruptcy would occur if the investment decision on the preceding trial yielded negative capital. Since being on the line segment CA^* implies a relatively large amount of borrowing (200 to 500%), the subjects preferred being in segment BB' with a lower rate of borrowing (0 to 200%), which *seemed* to imply a lower probability of bankruptcy. This reasoning, if indeed it took place, is fallacious, because the probability of bankruptcy with, say, portfolio A^* is smaller than that with portfolio B', despite the larger amount of borrowing involved in portfolio A^*.

[6] Borrowing \$2 per \$1 of initial wealth corresponds to point C in Figure 4.2.

TABLE 4.5 Frequency of Portfolio Choices by Line Segment and Session

	The location of the selected portfolio on the 10 various segments											
	1 Point *r*	**2 Segment *rA***	**3 Point *A***	**4 Segment *AA****	**5 Point *A****	**6 Segment *B′B****	**7 Point *B****	**8 Segment *rB***	**9 Point *B***	**10 Segment *BB′***	**Total of inefficient choices[a]**	**Total number of decisions**
Session 1	24	135	1	421	15	507	21	140	25	201	366	1490
Session 2	29	58	9	361	19	537	46	72	5	349	426	1485
Across	53	193	10	782	34	1044	67	212	30	550	792	2975

[a] These are inferior investment decisions by the M-V criteria: segments *rB*, *BB′*, and point *B*.

Note: Table taken from Kroll *et al*., 1988a, Copyright Academic Press, Inc., with permission.

Let us elaborate on this issue. If \$5 are borrowed and \$6 are invested in stock A, the mean dollar return is

$$\mu = \$6 \cdot 1.05 = \$6.30$$

and standard deviation in dollar terms is $\sigma = \$6 \cdot 4\% = \$0.24$. The subject who borrows \$5 has to pay back $\$5 \cdot 1.03 = \$5.15$. Therefore, with point A^* the probability of bankruptcy is

$$\begin{aligned} \Pr(w \le \$5.15) &= \Pr\left(\frac{w - \$6.30}{\$0.24} \le \frac{\$5.15 - \$6.30}{\$0.24}\right) \\ &= \Pr\left(z \le \frac{-\$1.15}{\$0.24}\right) = F(-4.79) \cong 8.3 \cdot 10^{-7} = 0.000083\% \end{aligned}$$

where w is the wealth from the game before the borrowed money and interest are paid back. Similarly, with point B', which implies \$2 of borrowing and investing in B, we have

$$\mu = \$3 \cdot 1.07 = \$3.21$$
$$\sigma = \$3 \cdot 12\% = \$.36$$

The probability of bankruptcy with strategy B' (with \$2 of borrowing, and a payment of interest and principle of \$2.06) is, therefore,

$$\begin{aligned} &\Pr(w \le \$2.06) \\ &\quad = \Pr\left(\frac{w - \$3.21}{\$0.36} \le \frac{\$2.06 - \$3.21}{\$0.36}\right) \\ &\quad = \Pr\left(z \le \frac{-\$1.15}{\$0.36}\right) = F(-\$3.19) \cong 0.071\% \end{aligned}$$

Despite the fact that with strategy A^*, \$5 are borrowed and with strategy B' only \$2 are borrowed, the probability of bankruptcy with strategy B' is much larger than with strategy A^*. Thus, the amount of borrowing is misleading regarding the chance of bankruptcy. Thinking that intensive borrowing increases the risk of bankruptcy (which is wrong in our case) probably led investors to select portfolios from the M-V inefficient segment BB'.

The possible error of the subjects with regard to the chances of bankruptcy may be related to the argument that "framing" affects the investors' decision (see Tversky and Kahneman, 1981). If one would present the investors the choice of net returns (e.g., the net returns corresponding to points A^* and B'), presumably most subjects would choose point A^*. However, when one tells the subject that they need to borrow 500% of their wealth to achieve point A^*, it might frighten them; hence the subjects may choose point B' with less borrowing. Let us elaborate on the framing issue. The percentage mean rate of return corresponding to point A^* is

$$6 \cdot 5\% - 5 \cdot 3\% = 15\%$$

and standard deviation of $6 \cdot 4\% = 24\%$.

Regarding point B' we have a mean of

$$3 \cdot 7\% - 2 \cdot 3\% = 15\%$$

and standard deviation of $3 \cdot 12\% = 36\%$.

Thus, if the subjects would have to choose from

A^*: Drawing a rate of return from a normal distribution with $\mu = 15\%$ and $\sigma = 24\%$

or

B': Drawing a rate of return from a normal distribution with $\mu = 15\%$ and $\sigma = 36\%$

with zero return at bankruptcy (whenever the return is zero or negative), most of them will probably choose A^*. However, by framing it differently and letting the subjects borrow 500% of their wealth to achieve point A^*, we suspect that some portion of the subjects would probably prefer B' to A^*.[7]

Analysis of Kroll, Levy, and Rapaport in the Expected Utility Framework

Kroll, Levy, and Rapaport employ the M-V rule and justify this by the normality of the return distributions. However, because the returns are truncated due to the possibility of bankruptcy, the normality assumption is violated; hence the M-V rule is not optimal. To be more specific, returns of -200%, -400%, or -150% all imply bankruptcy and the investor is indifferent to these three results. All three results imply zero wealth and elimination from the experiment. Thus, from the investor's point of view

$$P(x \leq 0) = P(x = 0)$$

which implies a mixed distribution that is discrete at $x = 0$ and continuous for $x > 0$. Because the normality is violated, the M-V rule is not optimal, and it is not obvious that all portfolios located on segment rB' are inefficient in the expected utility framework. Actually, we will show later

[7] Note that although there was a constraint on the amount borrowed, it was binding in 101 trials (3.4%) only, as shown in columns 5 and 7 of Table 4.5. It is therefore reasonable to speculate that if the subjects were allowed to borrow unlimited funds, the basic findings of Table 4.5 would remain unchanged.

that all portfolios located on rE are inefficient but portfolios located on segment EB' may be efficient.

Let us first compare portfolios A^* and E (see Figure 4.2) with density functions $f_{A^*}(x)$ and $f_E(x)$ and cumulative distributions $F_{A^*}(x)$ and $F_E(x)$. If the distributions are normal with possible negative realized returns, it is obvious that A^* dominates E because $E_{A^*}(x) > E_E(x)$ and $\sigma_{A^*}(x) = \sigma_E(x)$; in this case portfolio A^* dominates E for risk lovers and risk averters alike (see H. Levy, 1998). Namely, A^* dominates E by FSD. However, as explained earlier, in the Kroll *et al*. experiment the normality assumption breaks down because of the possibility of bankruptcy. The M-V rule is therefore not optimal, and it is not obvious that portfolio A^* dominates portfolio E. Nevertheless, despite the invalidity of the M-V rule, in this case we will show that the truncated returns corresponding to portfolio A^* dominate the truncated returns corresponding to portfolio E by FSD (for risk averters and risk lovers alike). To see this claim, let us first calculate the returns on these two portfolios.

Portfolio A*

For each \$1 of initial wealth, \$5 are borrowed, and \$6 are invested in stock A, whose mean return is 5% and standard deviation is 4%. Therefore, in net dollar terms the mean of the untruncated return distribution is

$$\mu = \$6 \cdot 1.05 - \$5 \cdot 1.03 = \$1.15$$

where \$5 · 1.03 is the principle and interest paid on the borrowed money.

Similarly, the standard deviation of the untruncated return distribution is

$$\sigma = \$6 \cdot 4\% = \$0.24$$

To have bankruptcy, we need to deviate \$115/0.24 ≅ 4.79 standard deviation to the left, hence the probability of bankruptcy with portfolio A^* is approximately 0.000083%.

Portfolio E

The (untruncated) standard deviation of portfolio A^* is 24% (see the previous calculation). The (untruncated) standard deviation of portfolio B is 12%. Therefore, to reach portfolio E with 24% standard deviation, we need to borrow \$1 and invest \$2 in B. Thus, in dollar terms we have for portfolio E

$$\mu = \$2 \cdot 1.07 - \$1 \cdot 1.03 = \$1.11 \text{ (see Table 4.4)}$$

and

$$\sigma = \$2 \cdot 12\% = \$0.24$$

Bankruptcy occurs if we deviate \$1.11/\$0.24 = 4.625 standard deviation to the left; hence the probability of bankruptcy is about 0.000187%. Figure

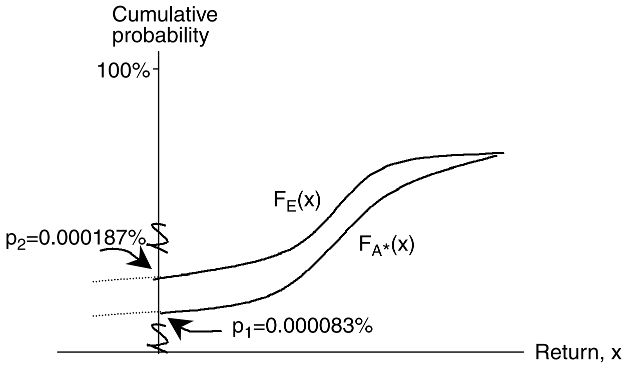


FIGURE 4.3 The cumulative distribution corresponding to portfolios E and A^* with bankruptcy.

4.3 draws the cumulative distribution of terminal wealth with possible bankruptcy corresponding to portfolio's A^* and E, respectively, when the truncation is accounted for. Note that two main results emerge from this figure:

a. $F_E(0) > F_{A^*}(0)$; hence there is a higher probability of bankruptcy with strategy E relative to strategy A^*.
b. $F_E(x) \geq F_{A^*}(x)$ for all values x (and strong inequality holds for at least one x); hence A^* dominates E by FSD (i.e., for all $U \in \mathbf{U}_1$).

Note that for $x > 0$ the cumulative distributions are the normal distributions, and because A^* has a higher mean and the same variance as E, it is located to the right, as drawn in Figure 4.3.

By the same argument, we claim that for every portfolio located on *rE* there is a portfolio located on *rA** which dominates it by FSD; this argument holds when there is no need to assume normality and when the possibility of bankruptcy is accounted for.

Unfortunately, Kroll *et al.* fail in their interpretation regarding portfolios located on segment EB', because even though such portfolios are inferior by the M-V rule, they may constitute a rational choice and the M-V rule is misleading in this case. To see this, compare portfolios B' and A^* and recall that Figure 4.2 provides the location of the various stocks according to the parameters given to the students when bankruptcy is *not* taken into account. With normal distributions, A^* dominates B' by the M-V rule as well as by SSD. However, because B' has a larger left tail than A^*, and because with bankruptcy all these negative results turn to zero, we will find that with bankruptcy, $E_{A^*}(x) < E_{B'}(x)$ (recall that with no bankruptcy both have equal means); hence A^* does not dominate B', because a necessary condition for dominance by FSD, SSD, and the M-V rule is that the dominating portfolio will have a mean that is equal or greater than the inferior portfolio (see H. Levy, 1998).

Kroll *et al*. report 550 choices from segment BB'. Unfortunately, there is no distinction between line *BE* and EB'. Thus, some of these portfolios are not necessarily inefficient, but we do not know how to allocate these 550 choices to efficient and inefficient choices. However, regarding the 212 choices of segment *rB* and the 30 choices of point *B*, we can safely assert that these are inferior choices by FSD (i.e., inferior for risk averters and risk seekers alike).

To sum up, in both the Gordon *et al*. and the Kroll *et al*. experiments, wrong investment choices were made. Moreover, the subjects not only selected investments that were M-V inefficient, which implies weak irrationality, they also selected investments that were FSD inefficient, which implies strong irrationality as defined in this chapter. Yet, we would also like to emphasize that in the Gordon *et al*. experiment, about three-quarters of the choices made were correct choices, and in the Kroll *et al*. experiment at least three-quarters of the choices made were also correct choices! Thus, it would be interesting to see the effect on asset pricing in a case where most investors make the correct choices but some investors make wrong investment choices. MS provides a framework to analyze such questions.

4.3. THE "HOT HAND" IN BASKETBALL AND LOOKING FOR TRENDS IN THE STOCK MARKET

The "Hot Hand" Illusion: Gilovich, Vallone, and Tversky's Study

Basketball players and fans alike tend to believe that a basketball player's chance of hitting a shot are greater following a hit than following a miss on the previous shot. A survey of 100 basketball fans reveals that

> 91% of the fans believed that a player has "a better chance of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots"; 68% of the fans expressed essentially the same belief for free throws, claiming that a player has "a better chance of making his second shot after making his first shot than after missing his first shot"; 96% of the fans thought that "after having made a series of shots in a row players tend to take more shots than they normally would"; 84% of the fans believed that "it is important to pass the ball to someone who has just made several (two, three, or four) shots in a row." (Gilovich, Vallone, Tversky, 1985, pp. 297–298)

Table 4.6, taken from Gilovich, Vallone, and Tversky (1985), reveals the runs test for nine players of the Philadelphia 76'ers. All the *Z* values are not significant, with the exception of one which is significantly negative. This means that in eight out of the nine cases there is no relationship between previous hits and the chance for success in the next shot, and in

TABLE 4.6 Runs Test — Philadelphia 76'ers

Players	Hits	Misses	Number of runs	Expected number of runs	*Z*
Clint Richardson	124	124	128	125.0	−0.38
Julius Erving	459	425	431	442.4	0.76
Lionel Hollins	194	225	203	209.4	0.62
Maurice Cheeks	189	150	172	168.3	−0.41
Caldwell Jones	129	143	134	136.6	0.32
Andrew Toney	208	243	245	225.1	−1.88
Bobby Jones	233	200	227	216.2	−1.04
Steve Mix	181	170	176	176.3	0.04
Daryl Dawkins	250	153	220	190.8	−3.09[a]

[a] $p < .01$.

Note: Table taken from Gilovich *et al.*, 1985, Copyright Academic Press, Inc., with permission.

one case the opposite holds: a series of hits increases the chance of a miss in the next shot, contrary to the "hot hand" hypothesis.

What can we learn from this example? People tend to create trends (of a series of hits) even if they do not exist in reality. The same may be true in the stock market; investors may believe that there are trends, even though changes in asset price over time are completely random. If this is also the case in the stock market, investors would make investment decisions based on past rates of return. However, there is an important difference between trends in the stock market and trends in basketball games. In basketball games, the (wrong) beliefs do not affect the next result (hit or miss). In the stock market, wrong beliefs many create demand (or supply) for the stocks which, in turn, affect the asset prices. Thus, at least in the short run, assets may be incorrectly priced due to a self-fulfilling prophecy, because investors act based on trends that actually do not exist! For instance, it is possible to observe two relatively high rates of return in a row, from which investors mistakenly conclude that the next rate of return will also be relatively high. They will purchase the stock, hence artificially creating another positive rate of return, even though such a positive rate of return is not economically justified. Let us turn to discuss the experimental findings regarding trend seeking.

The Quest for *Ex Post* Information: The Kroll, Levy, and Rapoport Study

In the experiment by Kroll *et al.* (1988a), the subjects could press a key on the computer to obtain the rates of return on stock A and stock B for the 15, 30, 45, or 60 preceding trials, where each press provided additional 15 *ex post* results. The subjects were told that the rates of return were drawn *randomly* from normal distributions and the parameters of the distributions were given to them. Therefore, theoretically, they should require no *ex post* information in order to select an optimal portfolio. The future rates of return are independent of past rates of return, hence all information on past rates of return is useless from an economic point of view, and any search for such information is worthless. Yet each of the subjects requested information at least once. It may be recalled that each subject in each session participated in 20 games of 5 trials each. With 15 subjects and 2 sessions, the maximum number of opportunities to request information on each trial is 30. For example, if all 15 subjects request information in trial 1 in both sessions 1 and 2, we have 30 requests for information in the trial.

Table 4.7 shows that when the experiment commenced, 28 of the 30 subjects requested information about past behavior of the two risky assets before investing their capital. With growing investment experience, the mean number of trials on which information was requested diminished by more than half. Additionally, even though the subjects were instructed that all 20 games in a session were identical (with the exception of the initial capital) and, in particular, that the distributions of rates of return on the

TABLE 4.7 Number of Times Information was Requested by Trial and Game

	Trial					
Game	**1**	**2**	**3**	**4**	**5**	**Mean**
1	28	11	5	8	6	11.6
2	24	2	6	7	7	9.2
3	23	7	3	6	6	9.0
4	22	4	4	6	4	8.0
5	21	2	6	5	3	7.4
6	21	3	2	5	6	7.4
7	22	5	4	2	3	7.2
8	20	3	5	2	3	6.6
9	19	2	2	0	2	5.0
10	16	5	5	4	3	6.6
11	14	4	5	2	3	5.6
12	16	1	4	4	4	5.8
13	16	2	2	2	4	5.2
14	16	1	2	2	2	4.6
15	17	2	6	1	4	6.0
16	15	2	1	0	1	3.8
17	16	6	3	1	3	5.8
18	18	3	4	4	0	5.8
19	15	3	2	1	3	4.8
20	15	2	3	3	2	5.0
Mean	18.70	3.50	3.70	3.25	3.45	

Note: Table taken from Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission.

two assets, A and B, denoted by $\tilde{R}_A$ and $\tilde{R}_B$, were stationary, at the beginning of each game they renewed their demand for information. Later on, as the game proceeded, they drastically reduced their demand for *ex post* returns (see trials 2 through 5 in Table 4.7).

Several possible explanations for the requests for information are suggested. It is possible that some or all of the subjects did not believe the instructions concerning the shape or parameters of the distributions of rates of return given to them. They needed the information to verify the instructions. It is also possible that although they had completed at least one introductory course in statistics, the subjects did not comprehend the implications of random sampling from normal distributions, and therefore requested information about past price movements. Yet another possible reason for requesting information is that it was provided free of charge (however, the time they spent represents "alternative" costs; hence it is not absolutely free of charge). Although these explanations are complementary, the most plausible explanation is that the subjects did not believe or comprehend that the sampling was random and independent. As many of the subjects testified in a postexperimental interview, they attempted to discover patterns or trends in the rates of return. We now turn to this issue.

Establishing Strategies Based on "Trends" in a Random Walk Setting

In the Kroll *et al.* experiment, the rates of return are independently and identically distributed, and the parameters of the return distributions are known. Therefore, past performance of the risky assets is of no relevance for future investment. Yet, as we shall see, the subjects did make investment decisions that are affected by the *ex post* rates of return. A series of negative rates of return or of positive rates of return should not affect the future investment decision.

Discarding the first two trials of game 1 on each session and allowing sequential effects across games, there are altogether 98 opportunities for each subject in each session to switch from one risky asset to another. Out of a maximum of 2940 switches over subjects and sessions ($98 \times 2 \times 15$), the subjects actually switched 445 times from stock A to B and 447 from B to A. Altogether, the risky asset was switched in about 30% of all trials. To explore the reasons for switching from one stock to another, the switches from A to B and from B to A were classified by the outcome of the preceding two trials as follows:

Scenario I: Switches from A to B where:	Scenario II: Switches from B to A where:
1. $R_{A,t-1} > R_{B,t-1}$ and $R_{A,t-2} > R_{B,t-2}$	1. $R_{B,t-1} > R_{A,t-1}$ and $R_{B,t-2} > R_{A,t-2}$
2. $R_{A,t-1} > R_{B,t-1}$ and $R_{A,t-2} < R_{B,t-2}$	2. $R_{B,t-1} > R_{A,t-1}$ and $R_{B,t-2} < R_{A,t-2}$
3. $R_{A,t-1} < R_{B,t-1}$ and $R_{A,t-2} > R_{B,t-2}$	3. $R_{B,t-1} < R_{A,t-1}$ and $R_{B,t-2} > R_{A,t-2}$
4. $R_{A,t-1} < R_{B,t-1}$ and $R_{A,t-2} < R_{B,t-2}$	4. $R_{B,t-1} < R_{A,t-1}$ and $R_{B,t-2} < R_{A,t-2}$

Table 4.8 presents the frequencies of switches between risky assets by category and type of switch. The subjects actually switched 445 times from stock A to stock B and 447 times from stock B to stock A. The table reveals that the frequencies are not distributed evenly over the four categories.

In about 67% of all cases, switches occurred in category 1 and 2, implying a negative recent performance effect. In categories 1 and 2 of scenario I, corresponding to the first row in Table 4.8, 254 switches occurred from A to B when A was recently outperforming B (either in $t-1$ or in both $t-1$ and $t-2$). Similarly, the first two categories in row 2 of Table 4.8, reveal that 344 switches occurred from B to A when B was recently outperforming A. Thus, it seems that a relatively high frequency of switches occurred in categories 1 and 2, indicating that the subjects believed that a relatively good past performance will be followed by a relatively bad performance, and a relatively poor performance in the past will be followed by a relatively good performance. Though there is a relatively high proportion of switches in categories 1 and 2 supporting this hypothesis, there is an exception in category 4 of scenario I: B performs better in $t-1$ and $t-2$ and 149 switches occurred from A to B. This number of switches is the highest of all categories in scenario I. This is consistent with the argument given in Chapter 2, that investors look at past rates of return and assume that more of the same will occur in the future. However, the relatively high frequency of switches in categories 1 and 2 support just the opposite view. Thus, when there are sequences of outperformance ($++$ or $--$) the subjects probably do not believe the random walk information given to them, and they create their own subjective beliefs. When there are mixed performances in the last two trials ($+-$, or $-+$, see categories 2 and 3), the shifts between the two stocks are almost even. However, even in this case, the performance in period $t-1$ is more

TABLE 4.8 Frequency of Switches between Stocks as a Function of the Outcomes of the Last Two Trials

	Category				
	1	2	3	4	Total
			Scenario I		
Switch from A to B	121	133	42	149	445
			Scenario II		
Switch from B to A	202	142	42	61	447
			Total number of switches in both scenarios		
	323	275	84	210	892

Note: Table taken from Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission.

important than the performance in $t - 2$, with a tendency to sell the stock with a good performance in period $t - 1$ and to switch to the other stock.

From these results we can conclude the following:

a. Past performance is important to the subjects for investment decision making, even when they are told that rates of return are drawn randomly from a given distribution.

b. The subjects tend to create subjective beliefs even though they are told that rates of return are drawn independently and randomly from given distributions whose parameters are given to them.

c. Some subjects believe that the future will be "more of the same" as the past, hence they tend to buy stocks that are characterized by two succeeding outperformances (see category 4 in the first row of Table 4.8). Other subjects base their decisions on the opposite view asserting that good performance follows bad performance (and bad performance follows good performance).

The results corresponding to category 4 mentioned earlier are the *raison d' être* for the operation and may be the reason for the success of the "behavioral" mutual funds (see Chapter 2). These funds base their operation on the investor errors mentioned earlier.

As the subjects were told that returns are drawn at random from given distributors with well-known parameters, there is no justification to switch between stocks due to past performance. Yet we identify two groups of switches and maybe two groups of investors. In one group, investors believe that the past will repeat itself (see category 4 in scenario I). We can consider this group as those investors who make a mistake and therefore artificially drive stock prices (in actual trading) upward (or downward) with no economic justification. The other group contains the sophisticated investors who take advantage of these errors: like the behavioral funds discussed in Chapter 2, these sophisticated investors buy the stocks with bad performance in the past because they are underpriced due to the errors done by the first group.

In the LLS microscopic simulation model described in Chapter 7, we model these experimental findings and assume that some of the investors base their investment decisions on past rates of return. We assume that these investors maximize expected utility and use the distribution of past rates of return to estimate the future distribution of rates of return. For example, a stock with + + rate of return series will be in demand, creating probably another "+" in the third period. Thus, unlike the "hot hand" wrong belief in basketball, in the stock market the wrong perception of investors may create a trend—that is, a prophecy can become self-fulfilling, at least in the short run. In the LLS model we assume that there is a subpopulation of investors who believe that past returns are representative of future returns. We would like to stress that changing the proportion of this subgroup in the total population and introducing other investor types (such as the sophisticated contrarians) are straightforward steps in the MS framework.

4.4. CORRELATIONS AND THE PORTFOLIO INVESTMENT DECISION: HOW CLOSE ARE INVESTORS TO THE EFFICIENT FRONTIER?

In the mean-variance framework all investors, regardless of their preferences, should diversify between a given portfolio (the "market portfolio") located on the efficient frontier and the riskless asset. This property is well known as the separation theorem, which is necessary to derive the CAPM. In an experiment with three risky assets and a riskless asset, Kroll, Levy, and Rapoport (1988b) tested the separation theorem and the investment efficiency of the subjects participating in the experiment. In a sense, it is an extension of their previous experiment (1988a), but in this study the experiment is closer to actual investment decision making under uncertainty because portfolio diversification among the risky assets was allowed.

The subjects were undergraduate students. There was a monetary reward contingent on performance but no losses were possible. All subjects had at least one comprehensive course in statistics. In this study, participants were required to make several investment decisions with the following tasks:

a. Diversification among three stocks without the riskless asset.
b. Diversification among three stocks with the riskless asset.

The correlation between stocks B and C varied in various groups from $\rho_{B,C} = 0$, $\rho_{B,C} = 0.8$, and $\rho_{B,C} = -0.8$. The correlation of other pairs of stocks was zero. The effect of correlation on diversification strategy was tested.

The main results in this experiment are as follows: most subjects diversified between the three risky assets, and when the monetary reward was increased tenfold, the investors' diversification and their performance substantially increased. These two results tend to support portfolio theory. However, in this study there are some results with negative implications for investment and portfolio theory: even though the subjects were informed that rates of return were drawn randomly from normal distributions with known parameters (given to them), they repeatedly requested information on past rates of return. An even more disturbing result (from a portfolio theory point of view) is that changing the correlation $\rho_{B,C}$ in the range $-0.8 \leq \rho \leq 0.8$ does not substantially affect the diversification policy. Moreover, the introduction of the riskless asset did not enhance homogeneity in the diversification of the three risky assets across investors.

These results strongly contradict the separation theorem and the CAPM. Finally, the portfolios selected by the subjects were quite far away to the right of the M-V efficient frontier, indicating either inefficiency or irrationality of the subjects participating in this experiment.

The improvement in the subjects' performance with an increase in the reward hints at the importance of making the experiment as close as possible to reality. Indeed, Kroll and Levy (1992) repeated the Kroll *et al.* experiment (with the same framework and the same three risky assets) with the following improvements, in an attempt to make the experiment closer to actual investment conditions:

a. There was a reward and a penalty that resulted as a function of performance (in the Kroll *et al.* study there was a reward but not a penalty).

b. The subjects were MBA students specializing in finance, not a random sample of university students. Hence, they were more interested in the subject matter and probably more representative of investors in the stock market. However, it is important to note that the subjects completed the experiment before they were exposed to the computational tools of portfolio selection theories (e.g., before studying how to employ a quadratic program or how to use any software to solve for the M-V efficient frontier). We would like to stress that in reality not all people are investors in the stock market. Therefore, MBA students are more relevant to test investment theory than a random sample of university students, because a high proportion of the latter group may not invest in the stock market at all. The fact that, on average, the sample of subjects of the Kroll *et al.* study did not invest according to portfolio theory is not sufficient to reject portfolio theory, because a high proportion of the subjects may never actually invest in the stock market. Portfolio theory does not imply that all people have to invest in the stock market.

c. Finally, each subject's diversification policy and the value of his or her portfolio at the end of each step were made public, hence the successful investor in the past could be mimicked by other investors. This is consistent with actual market conditions because the *ex post* diversification and performance of the experts (e.g., mutual funds) are publicly available.

Indeed, with this setting, which is closer to reality, the experimental results are much more in favor of portfolio theory than the results of Kroll *et al.* To be more specific, the subjects behaved almost "as if" they knew a quadratic programming technique to solve the mean-variance efficient frontier, because they selected portfolios that were very close to the theoretical frontier. Also, in contrast to the Kroll *et al.* study, in this study the subjects responded to changes in the correlation coefficient, and in the right direction: the lower the correlation between assets B and C, the higher the investment proportion in these two assets, exactly as predicted by portfolio theory.

Figure 4.4 summarizes the efficiency of the investment choices in the Kroll *et al.* experiment and in the Kroll and Levy experiment. Given the parameters of the three stocks, first the M-V efficient frontier and the optimal unlevered portfolio (portfolio O, see Figure 4.4) were derived. Each subject selected a portfolio that can be drawn as a point on the M-V space. The mean and the variance of the weighted average portfolio held by all subjects can be calculated (the actual "market portfolio"). This has been done twice, once without the riskless asset and once with the riskless asset. Of course, the actual market portfolio is expected to be to the right of the efficient frontier. The two most relevant portfolios are the market portfolios L of the Kroll and Levy experiment and portfolio $\bar{L}$ of the Kroll *et al.* experiment, where L indicates that the riskless asset is available (for lending and borrowing alike). The optimum portfolio is the tangency portfolio O. As can be seen from this figure, portfolio L is quite close to portfolio O, unlike portfolio $\bar{L}$, which is located far away to the right, in the inefficiency zone.

From the comparison of these two experiments, one can draw the following conclusions, which can be employed in MS models:

a. Testing an investment theory should be done with actual investors or potential investors. Results of subjects selected randomly, who do not

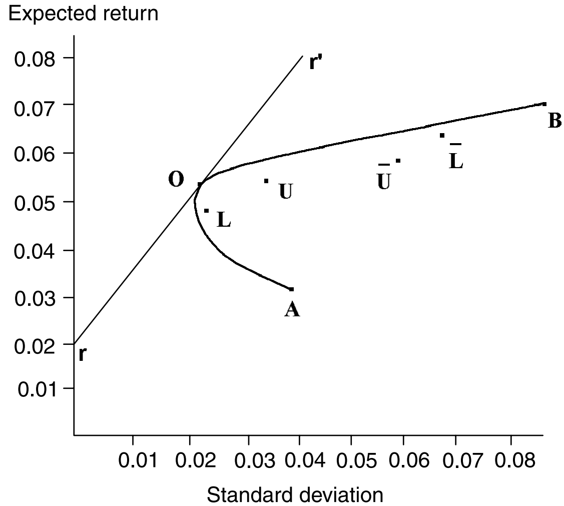


FIGURE 4.4 The efficient frontier and actual portfolio selected by the subjects. *Notes*: O—optimal, L—leveraged, U—unleveraged, "bar" (¯) designate a result from the Kroll *et al.* study. (From Kroll *et al.*, 1988a, Copyright Academic Press, Inc., with permission).

invest and probably never will, are irrelevant. Thus, in MS market models one can safely assume that the investors do have some degree of investment efficiency. They do know, at least partially, to read and correctly understand the parameters given to them.

b. When the reward is increased or when penalty is also introduced, the investment choices, including the response to changes in correlations, conform more closely with the theory of portfolio selection. Because in actual markets, reward and penalty do exist, one can safely assume in MS models that investors utilize the gain from diversification and do not ignore correlations.

c. When realistic market conditions governed the experiment, a portfolio, L, was selected. This portfolio differs from portfolio O; however, it is close to it. This implies that investors behave *almost* "as if" they know the theory. We emphasize the *almost*, because they still do not select portfolio O. This means that investors are not perfectly "efficient machines," and that they do not always succeed in fully exploiting the gain from diversification. Thus, some degree of error in the investment choice processes exists in the market.

Indeed, in the LLS model we rely on these findings: we assume that investors are almost "efficient machines" but some degree of error or irrationality exists, which is consistent with the difference between portfolios L and O in Figure 4.4.

4.5. TESTING THE CAPM: AN EXPERIMENTAL SETTING WITH *EX ANTE* PARAMETERS

All the experiments discussed so far tested the existence of the components of the theory of portfolio selection and of rational choices. Consistent choices, reaction to change in correlations, diversification, and efficient choices were tested. Yet none of these tests directly examine the risk-return equilibrium relationship as implied by the CAPM.

As in testing relative and absolute risk aversion, in testing the CAPM, experimental studies are superior to empirical studies: the empirical studies are based on *ex post* data (because *ex ante* parameters are not available); hence they are difficult if not impossible to employ in testing the CAPM (see Roll, 1977).

To the best of our knowledge there are two experiments that directly test the CAPM. One by Levy (1997) and one by Bossaerts, Kleiman, and Plott (1999). We devote this section to Levy's study, yet we would like to mention at the outset that the experimental design and the results of Bossaerts *et al*. are a little different. While Levy shows strong empirical support for the CAPM, the study by Bossaerts *et al*. asserts that

> slow but steady convergence towards the CAPM is discovered. The convergence process, however, halts before reaching the actual equilibrium. There is ample evidence that subjects gradually move up in mean-variance space, in accordance with the CAPM. Yet, adjustment stops as if the remaining trading time was insufficient to complete all the transactions that are needed to guarantee improvements in positions. We conjecture that this is due to subjects' hesitance in the face of market thinness. Because the convergence process halts, statistical tests reject the CAPM.

Thus, either time constraints or market thinness are the main factors by which Bossearts *et al*. explain the invalidation of the CAPM. The study by H. Levy (1997), which supports the CAPM, overcomes these difficulties as well as the difficulty of the empirical studies that aim to test the CAPM. We will extend and explain this point later on in this section.

4.5.1. The Experiment

The subjects in this experiment were first-year MBA students at the Hebrew University of Jerusalem who were specializing in finance. Since the subjects had not taken investment or portfolio courses, they did not know how to derive an optimal mean-variance portfolio or the CAPM. They were exposed only to the notion of risk as measured by the variance of returns through a basic finance course. The experiment was not mandatory and did not affect the subjects' course grade. In all, 64 of the 67 students chose to participate in the experiment. Since the subjects could lose money,[8] they were told at the beginning of the experiment to check their savings and other income resources in order to make sure that they had enough money to cover possible losses.[9]

Each subject was given an initial endowment of $30,000 and the ability to invest in the stocks of 20 pure equity firms. There was a total of 10 rounds of trade in these stocks. The *book value* of each firm's assets at the beginning of the experiment was given to the subjects.

[8] Creating an experimental setup where the subjects can actually make or lose real money (as opposed to a setup in which subjects are asked hypothetical questions) greatly affects the experimental results. Kroll, Levy, and Rapoport (1988b) and Kroll and Levy (1992) show that the more the subjects have at stake, the closer their decisions will be to those predicted by normative models. See also Fiorina and Plott (1978), Plott (1986), and Smith (1991).

[9] Theoretically, subjects should consider a portfolio that includes their private assets as well as the experiment's assets. However, this and other experiments reveal that people tend to undertake "mental accounting" and make separate decisions rather than combine the outcome from all decisions, as theory tells us they should do. In this experiment, neither the subject's wealth (which was on average $35,641) nor the subject's monthly income (which was on average $962) affected investment behavior (see Levy, 1994).

The following information was given to the subjects: the firm's assets are growing or declining in each period at random, where the random variable is normally distributed. The mean and the variance of the random variables differ across firms and were reported to the subjects. For example, denoting the book value of the i^{th} firm at period t by $V^B_{i,t}$, the book value of firm 1 at the end of period 1 is given by $\tilde{V}^B_{1,1}$, where

$$\tilde{V}^B_{1,1} = V^B_{1,0}\left(1 + \tilde{R}^B_{1,1}\right) = \$78{,}930\left(1 + \tilde{R}^B_{1,1}\right)$$

the superscript B stands for book value, the first subindex indicates that we are dealing with firm $i = 1$, the second subindex denotes the period ($t = 1$ or $t = 0$), and \$78,930 is the book value (which varies across firms) of the assets of firm 1 at the beginning of the experiment. $\tilde{R}^B_{1,1}$ is a random variable drawn in the first period from a normal distribution with a mean of 3% and standard deviation of 4%, reflecting the operating earnings of firm 1. The superscript B indicates that these are rates of return that the firms earn on the book value of assets—namely, the operational profit of the firm. Similarly, at the end of the second period the book value of this firm's assets grows to \$78,930 $(1 + \tilde{R}^B_{1,1})$ $(1 + \tilde{R}^B_{1,2})$, where $\tilde{R}^B_{1,2}$ is a random variable drawn in the second period from the same normal distribution. Since there are 10 periods (or trading rounds) in the experiment, the i^{th} firm's asset value at the end of the 10^{th} period is given by

$$\tilde{V}^B_{i,10} = V^B_{i,0} \prod_{t=1}^{10} \left(1 + \tilde{R}^B_{i,t}\right)$$

where $\tilde{R}^B_{i,t}$ is the random variable corresponding to firm i in period t. As in a real-world environment, the subjects also observe the market prices of the stocks as determined in each trading round, and can use this information in their investment decisions in the next trading round.

The subjects were told that $\tilde{R}^B_{i,t}$ is the cash rate of return on the assets and not an accounting term. There are no taxes, and the book value of the asset $\tilde{V}^B_{i,t}$ is equal to the market value of the asset should the firm liquidate its assets in period t. However, it was fully known that none of the firms would liquidate their assets during any of the first nine rounds. The subjects trade the stocks of the firm and in each trading round determine the market value of the firm, which could be different from the book value. This scenario is very similar to closed-end mutual funds. When the shares of a closed mutual fund are sold or bought in the market, their market price, as determined by the supply and demand, could be different from the market net assets' value per share (or the net asset value known as NAV) of the mutual funds (hence we observe a premium or discount in the market). The subjects were told that at the end of the 10^{th} round all firms would liquidate and $V^B_{i,10}$ which is the liquidation value of these assets, would be distributed among subjects proportionately to the number of shares of the firm held by each subject. The risk-free interest rate was constant at $r = 2\%$, and there was no constraint on borrowing or lending. The subjects' potential profit or loss was structured as follows.

The Subjects' Financial Profit or Loss

Each subject received at time t_0, \$30,000 of "paper" money to buy stocks or bonds. If the subject did not go bankrupt at some point during the experiment, at the end of the 10th round the wealth of the k^{th} subject was given by $W_{k,10}$

$$W_{k,10} = \sum_{i=1}^{20} (N_{i,k}/N_i)\tilde{V}^B_{i,10} - B_{k,9}(1 + r)$$

where $\tilde{V}^B_{i,10}$ is the liquidation value of the i^{th} firm at the end of the 10th period, N_i is the number of shares issued by the i^{th} firm, $N_{i,k}$ is the number of shares of the i^{th} firm held by the k^{th} investor ($i = 1, 2, \ldots, 20$ firms, $k = 1, 2, \ldots, 64$ subjects), and $B_{k,9}$ is the amount of outstanding borrowing (or lending) of the k^{th} investor at the end of the 9th trading round. Thus, $B_{k,9}(1 + r)$ is the amount of money the subject should pay back to the bank, where r denotes the interest rate. Note that if the subject lends money at the 9th round, $B_{k,9} < 0$, hence $-B_{k,9}(1 + r) > 0$. To calculate the actual reward of each subject, each \$1000 of "paper" money represents \$1; hence the actual financial reward of the k^{th} subject at the end of the 10th round is R_k:

$$R_k = W_{k,10}/1000$$

where $W_{k,10}$ is as defined previously. At the end of *each* trading round, the net market value of the assets of each subject is examined. If it is negative, the subject goes bankrupt and pays out of his or her own pocket money to the experimenter. The penalty paid by the investor should he or she go bankrupt at period t is equal to the amount of his or her (negative) net market value of assets divided by 1000:

$$W^m_{k,t}/1000 = [V^m_{k,t} - B_{k,t-1}(1 + r)]/1000$$

where $W^m_{k,t}$ is the net *market* value of the k^{th} investor's wealth at the end of round t, $V^m_{k,t}$ is the *market* value of the stocks (as determined by the demand and supply of the subjects to distinguish from the firm's book value), which the k^{th} investor holds at the end of period t, and $B_{k,t-1}$ is the amount borrowed at period $t - 1$. For example, if the market value of the stocks held by investor k at the end of the 5th period is $V^m_{k,5}$ = \$100,000, and the subject's outstanding borrowing is $B(1 + r)$ = \$1,000,000, his or her wealth is \$100,000 − \$1,000,000 = \$ −900,000, and he or she must pay $\frac{900{,}000}{1{,}000}$ = \$900 (recall that we divide by 1000, such that each \$1000 of "paper" money represents \$1 for the calculation of the actual profit or loss). When a subject goes bankrupt, he or she is eliminated from the experiment for the remainder of the trading rounds. Thus, a penalty is paid by the subject whenever the net wealth is negative. However, the reward is paid off only at the end of the 10^{th} trading round.

Thus, the reward is directly indexed to the subject's performance in the experiment. The average reward was about \$70, with the highest reward equal to \$543 and the lowest reward equal to \$33. The subjects were very conservative in their borrowing policy, being net lenders: on average, 26% of the portfolio was invested in bonds. This explains why no subject went bankrupt.

4.5.2. Competitive Equilibrium Pricing Models

The CAPM

According to the Sharpe-Lintner CAPM, the mean rate of return on the i^{th} asset, μ_i, is related to beta by the following linear relationship:

$$\mu_i = r + (\mu_m - r)\beta_i$$

where r is the riskless interest rate and μ_m is the mean rate of return on the value-weighted market portfolio (m stands here for the market portfolio). Beta is given by $\beta_i = \text{Cov}(\tilde{R}_i, \tilde{R}_m)/\sigma_m^2$, where $\tilde{R}_i$ and $\tilde{R}_m$ are the market rate of return on the i^{th} asset and on the market portfolio, respectively (for more details see Sharpe, 1964, and Lintner, 1965a).

The one-period market rate of return to the investor in period $t + 1$ is given by $R_{i,t+1} = (V_{i,t+1}/V_{i,t}) - 1$, where $t = 1, 2, \ldots 9$; namely, we have nine market rates of return for each stock, when $V_{i,t}$ is the market value of the i^{th} firm at period t. For example, the 1st rate of return on the i^{th} stock is $V_{i,2}/V_{i,1} - 1$ and the 9^{th} rate of return is $V_{i,10}/V_{i,9} - 1$. The 10^{th} period rate of return to the investor is calculated in a different way; it is $V^B_{i,10}/V_{i,10} - 1$, where $V^B_{i,10}$ is the liquidation value of the firm.

Unlike the first 9 trading rounds, the 10^{th} round can supply an accurate test of CAPM with *ex ante* parameters. If at time t (for any round t) the distributions of $\tilde{V}_{i,t+1}$ and the market portfolio $\tilde{V}_{m,t+1}$ are known, the CAPM can be tested on an *ex ante* basis. The CAPM asserts that by having information on these distributions corresponding to period $t + 1$, trading in the assets in period t determines the current equilibrium price $V_{i,t}$, which simultaneously determines the *ex ante* parameters μ_i, μ_m, and β_i. To see this, recall that $\mu_{i,t}$ and $\beta_{i,t}$, which are the parameters corresponding to period $t + 1$ and which are relevant for the decision at period t, can be rewritten as

$$\mu_{i,t} = E\left[\tilde{V}_{i,t+1}/V_{i,t}\right] - 1$$

and

$$\beta_{i,t} = \text{Cov}\left(\tilde{V}_{i,t+1}/V_{i,t}, \tilde{V}_{m,t+1}/V_{m,t}\right)/\text{Var}\left(\tilde{V}_{m,t+1}/V_{m,t}\right)$$

Obviously, for determined market values $V_{i,t}$ and $V_{m,t}$, and known distributions of $\tilde{V}_{m,t+1}$ and $\tilde{V}_{i,t+1}$, the future parameters μ_{it} and β_{it} are not estimates based on historical data, but they are the precise *ex ante* values based on the future distributions of rates of return that are determined collectively by the investors.

In practice, however, the distributions of $\tilde{V}_{i,t+1}$ and $\tilde{V}_{m,t+1}$ are unknown; therefore, empirical tests of CAPM employ *ex post* parameters as estimators of these *ex ante* parameters. Observations are taken from the past to estimate beta and then the following standard cross-sectional regression is employed:

$$\overline{R}_i = \gamma_0 + \gamma_1 \hat{\beta}_i + e_i$$

where $\overline{R}_i$ is an estimate of μ_i and $\hat{\beta}_i$ is an estimate of the beta corresponding to future rates of return.[10] The various procedures for conducting cross-sectional empirical tests share a feature in that they all employ some sort of method for estimating β_i and μ_i based on historical rates of return. In most empirical studies, γ_1 is significantly positive (i.e., beta measures risk), with an R^2 of about 20%. However, in a recent study by Fama and French (1992), γ_1 is not significantly different from zero, hence the conclusion of Fama and French is that beta does not measure risk.

Experimental paradigms have their advantages and disadvantages. With CAPM there is a major advantage to the experimental framework: it can be designed so that the distribution $\tilde{V}_{i,t+1}$ is known precisely. If the CAPM is intact, the equilibrium values $V_{i,t}$ must be determined such that all *ex ante* pairs of μ_i and β_i lie on a straight line or very close to it. This is impossible to design with empirical tests of the CAPM because the *ex ante* distributions of $\tilde{V}_{i,t+1}$ are unknown.

The experiment conducted in Levy's study consists of 10 trading rounds. One can employ the 10 market rates of return to estimate historical mean return and betas, as in the empirical study, and then run the second-pass cross-sectional regression to test CAPM. This is not much different from what has been done before. However, the 10th round

[10] Roll (1977) showed that using the same *ex post* data to estimate β_i and to test the CAPM may raise a problem; if a mean-variance efficient portfolio is employed to measure β_i, then $R^2 = +1$ is obtained as a tautology; it neither confirms nor refutes the CAPM. The test suggested by H. Levy (1997) does not suffer from this drawback.

provides an ideal condition for testing CAPM (and other models), and in this respect the experimental framework is superior to empirical studies. Let us elaborate.

When investors make a decision at the 10th round they determine the market value of each firm $V_{i,10}$, knowing that at the end of the period the random value of the firm will be $\tilde{V}^B_{i,10} \equiv V^B_{i,9}(1 + \tilde{R}^B_{i,10})$, where $V^B_{i,9}$ is the book value of the firm at $t = 9$, as reported to the investors before the last round takes place, and $\tilde{R}^B_{i,10}$ is a random variable whose *distribution* is known to the investors. Recall that at the end of the 10th period the firm's assets are liquidated and the cash obtained from the liquidation is equal to $V^B_{i,10}$. Thus, since $V^B_{i,9}$ is known and reported to the investors before they conduct the last round of trade, and since the distribution of the random variable $\tilde{R}^B_{i,10}$ is known, the *distribution* of $\tilde{V}^B_{i,10}$ is known before the last trade takes place. Therefore, before the 10th trading round takes place, investors know precisely the *distributions* of the future cash flow they will get at the end of the 10th period. If the CAPM holds, the investors should determine equilibrium market prices in the last trade, $V_{i,10}$, such that the *ex ante* future pairs of parameters (μ_i, β_i) would lie approximately on a straight line, known as the CAPM security market line.

Calculating the parameters for the cross-sectional regression in the 10th round is done as follows: the market rate of return on the ith stock in the last round is given by $\tilde{R}_{i,10} = [\tilde{V}^B_{i,10}/V_{i,10}] - 1 = [V^B_{i,9}(1 + \tilde{R}_{i,10})/V_{i,10}] - 1$, where $V_{i,10}$ is the market value as determined at the *beginning* of the last period and $V^B_{i,10}$ is the liquidation value at the *end* of the last period. In determining the market value $V_{i,10}$, the subjects determine simultaneously all the relevant *ex ante* parameters. The mean one-period rate of return is $\mu_{i,10} = [(V^B_{i,9}/V_{i,10})E(1 + \tilde{R}^B_{i,10})] - 1$. Similarly, the variance is given by $\sigma^2_{i,10} = (V^B_{i,9}/V_{i,10})^2\sigma^2(\tilde{R}^B_{i,10})$. Since the expected values and variances of $\tilde{R}^B_{i,10}$ were given to the subjects in determining the equilibrium value $V_{i,10}$, the subjects simultaneously determine the future mean and variance of the rate of return on each stock. Denoting the market proportion at the 10th round in the ith firm by $X_{i,10} = V_{i,10}/\Sigma_{i=1}^{20} V_{i,10}$, the *market portfolio* rate of return corresponding to the last round is given by $\tilde{R}_{m,10} = \Sigma_{i=1}^{20} X_{i,10}\tilde{R}_{i,10}$, where the subscript m denotes a market portfolio. Having $\tilde{R}_{m,10}$ and $\tilde{R}_{i,10}$, one can calculate the market portfolio *ex ante* variance for the 10th period and the beta, which are necessary for testing the CAPM.

With the *ex ante* pairs of parameters $(\mu_{i,10}, \beta_{i,10})$, for each stock one can run a cross-sectional regression of the form

$$\mu_i = \gamma_0 + \gamma_1 \beta_i + e_i$$

which tests CAPM with *ex ante* parameters. If CAPM is invalid, the investor would determine $V_{i,10}$ in such a way that μ_i would not be positively related to β_i. However if β_i is an important risk index, $V_{i,10}$ would be determined such that μ_i and β_i are positively related.

The Generalized CAPM

The CAPM predicts that the subjects would hold some proportion of all of the 20 risky assets available, namely, that they would hold a portfolio located on the mean-variance efficient set. Due to the existence of fixed or variable transaction costs (Levy, 1978), an uneven flow of information (heterogeneous expectations; Merton, 1987), or constraints on borrowing (Markowitz, 1990), in practice, the capital market is segmented; therefore, not all investors hold all available risky assets. Indeed, transaction costs, asymmetrical information, or constraints on borrowing are the main motivation for the development of the generalized CAPM (GCAPM). The CAPM implies that in equilibrium all investors should hold the market portfolio composed of all risky assets. Realizing that in practice this is not the case, Levy (1978), Merton (1987), and Markowitz (1990) developed the GCAPM, which is a variant of the CAPM. The GCAPM is an equilibrium risk-return relationship in a *segmented market*. It states that the k^{th} investor may hold only n_k risky assets, creating a portfolio whose return is R_k with a mean of μ_k and risk $\beta_{i,k}$, where $\beta_{i,k}$ is the beta of the i^{th} asset, calculated against the return on the k^{th} investor's optimum portfolio R_k. In this case, the k^{th} investor, and all other investors who hold portfolio R_k, have a segmented market of risky assets with a "little CAPM" or segmented CAPM, in which the following risk-return relation holds (for details see Levy, 1978, and Merton, 1987):

$$\mu_i = r + (\mu_k - r)\beta_{i,k} \qquad i = 1, 2, \ldots n_k$$

This equilibrium relationship should hold for all assets included in this segment. Obviously, the i^{th} stock may be held in more than one segment, and hence it may be included in many portfolios R_k. Denoting the wealth invested by the k^{th} investor in the stock market by T_k, the preceding equilibrium relationship can be rewritten as

$$\mu_i = r + \beta_i^*$$

where

$$\beta_i^* \equiv \sum_k T_k(\mu_k - r)\beta_{i,k} / \sum_k T_k$$

which is a weighted average of all the $\beta_{i,k}$'s (recall that each investor holds a different set of assets with the i^{th} asset, and hence has a different $\beta_{i,k}$). We refer to β_i^* in what follows as the GCAPM beta.

In empirical testing of the CAPM, the GCAPM has an advantage over CAPM since it describes the actual behavior of investors who hold a relatively small number of assets in their portfolio. Hence, GCAPM is expected to provide better explanatory power than CAPM. However, its disadvantage in comparison with CAPM is technical in nature, since it is more difficult to test empirically. To test GCAPM, one has to first know the portfolio composition of each investor (to calculate $\beta_{i,k}$), and the wealth of each investor, T_k; this kind of data is usually inaccessible.

An experimental study once again provides an ideal framework to test GCAPM. At the end of each trading round, the precise portfolio composition of each investor, as well as his or her invested wealth, T_k, is known. For the 10th trading round one can calculate the rate of return on the i^{th} asset, and having the k^{th} investor's portfolio composition at the beginning of the last round of the experiment, one can calculate the mean rate of return on the k^{th} portfolio held, μ_k, as well as $\beta_{i,k}$. The k^{th} portfolio rate of return is

$$\tilde{R}_{k,10} = \sum_{i=1}^{n_k} X_{i,10,k}\tilde{R}_{i,10}, \text{ with mean } \mu_{k,10} = \sum_{i=1}^{n_k} X_{i,10,k}\,\mu_{i,10}$$

where $X_{i,10,k}$ is the investment proportion of the k^{th} investor held in the i^{th} stock at the beginning of the 10th period.

Having a portfolio $\tilde{R}_{k,10}$, $\beta_{i,10,k}$ can be calculated as follows:

$$\beta_{i,10,k} = \text{Cov}\left(\tilde{R}_{i,10}, \tilde{R}_{k,10}\right)/\sigma^2_{k,10}$$

Since the investor's wealth T_k (at each round) is also available, the GCAPM can be tested. The GCAPM was tested for the 10th trading round. The results of the CAPM and GCAPM are reported in the next section.

4.5.3. The Results

In the experiment just discussed, an ideal condition to support CAPM is created: the investors did not pay transaction costs, symmetrical information regarding the available assets was given to all subjects free of cost, and unlimited borrowing and lending was allowed. In spite of this, investors were found not to hold all 20 stocks in their portfolio but specialized in a relatively small number of stocks.

The striking result is that none of the subjects held all 20 assets through all rounds. On average, a little fewer than 5 assets were held in either the 10th trading round, or the average of all 10 rounds. It is interesting that in a comprehensive survey, Blume, Crockett, and Friend (1974) found that, on average, 3.41 assets are included in the American investor's portfolio. One explanation is that transaction costs may limit the number of assets held in the portfolio. In our experiment, there are no direct transaction costs; still, small portfolios are held. This implies that there are other costs: time required to carry out calculations, keeping track of all 20 assets, and so on. Obviously, not holding all assets implies that the selected portfolios are not mean-variance efficient. This result is demonstrated in Figure 4.5, which corresponds to the 10^{th} trading round. For this round the means and variances of all assets are known precisely. Hence one can derive the *ex ante* mean-variance frontier. Since the investment strategy of each subject is known, one can also calculate the mean and variance of each selected portfolio by the k^{th} subject. As can be seen, all portfolios selected by the subjects are interior, not falling on the frontier, and the market portfolio is interior.

When the subjects were asked at the end of the experiment about the reasoning for this less-than-perfect diversification strategy, they claimed that it was very hard to follow and analyze all 20 assets; therefore, even if there are no direct transaction costs, there are costs in the form of time spent to analyze the investment behavior and in the human capability to handle information regarding many stocks. Thus, the subject probably believed that a "little diversification goes a long way"[11] because most of

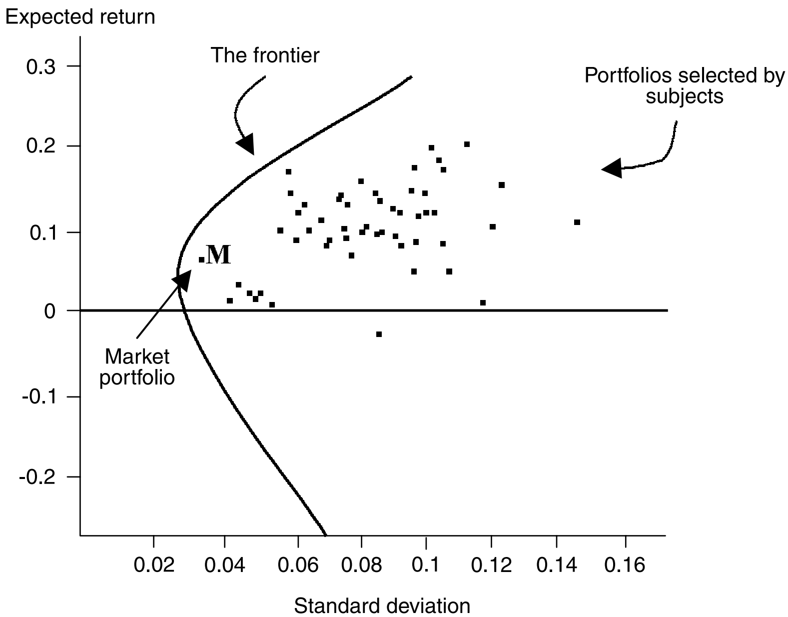


FIGURE 4.5 Efficient frontier and the selected portfolios in round 10. (From Levy, Copyright 1997 Blackwell Publishers, with permission).

[11] Kroll, Levy, and Markowitz (1984) show that diversifying in a relatively small number of assets approximately maximizes the expected utility even though there are many assets in the market. This notion of "a little diversification goes a long way" was also found in our experiment.

the portfolio risk-reduction is obtained by holding only a few assets. When the marginal transaction cost (not necessarily out-of-pocket direct costs) is larger than the benefit from the marginal risk reduction, the investor does not add more assets to the portfolio. This lack of diversification supports the GCAPM's framework.

The next step is to contrast the CAPM and the GCAPM models directly by examining the *ex ante* explanatory power of each.

The following results are obtained regarding β (CAPM) and β^* (GCAPM): in the 10^{th} trading round, $\beta^* = 4.850$, with a T value of 6.49 and R^2 of 71%, and $\beta = 0.065$, with a T-value of 6.33 and R^2 of 70%.

Thus, β^* as well as β have an explanatory power much higher than is normally obtained in empirical studies. While in empirical studies the R^2 is generally below 20% (with individual stocks, not portfolios), here it is relatively high (about 70%).

It is interesting to note that these experimental results, though not consistent with Fama and French's (1992) empirical study, conform with recent empirical studies by Amihud, Christensen, and Mandelson (1992) and Jagannathan and Wang (1993), which conclude that CAPM is alive and well. Jagannathan and Wang (1993) assert

> When human capital is also included in measuring wealth, the CAPM is able to explain 28% of the cross-sectional variation in average returns in the 100 portfolios studied by Fama and French. When, in addition, betas are allowed to vary over the business cycle, the CAPM is able to explain 57%.

Similarly, Amihud *et al.* (1992) conclude that "Beta is still alive and well."

The main findings of Levy's study are as follows: investors do not diversify efficiently in all stocks, hence their portfolios are not located on the M-V efficient frontier. Yet even with this "little diversification," the CAPM and the GCAPM have a very strong explanatory power, with an R^2 of about 70%. Thus, irrationality or inability to make the optimal choices does not mean that the theoretical models necessarily break down or, in our specific case, that beta does not measure risk. It does measure risk with the qualification that even in an ideal situation, when R^2 is expected to be 100%, we obtain a less than perfect fit, due to the irrational choices or due to inability to achieve the desired goal. Another interesting result is that the subjects focus on a relatively small number of assets. We employ this principle in the LLS model: one riskless asset and one or several risky assets are included in the portfolio, despite the fact that thousands of assets are available in the market.

4.6. SUMMARY

Even when all assumptions underlying a given theoretical model are intact (e.g., normal distribution and risk aversion), the model's results may not hold because some portion of the investors are either irrational or do not know how to achieve their desired goal. The assumption that investors are supermen or very "efficient machines" may not hold in practice.

In various experiments, the following results have been found:

a. In the Gordon *et al.* experiment, about 29% of the investment decisions are wrong because investors select investments that are inferior by FSD, SSD, or M-V. In the Kroll *et al.* study, the proportion of wrong choices is between 8.1% (= 242/2975, see Table 4.5) and 27% (= 792/2975).

b. Investors repeatedly request information on past rates of return and believe in trends (the "hot-hand" hypothesis) even when they are told that the rates of return are drawn at random from given distributions.

c. Investors diversify in only 3 to 5 stocks out of 20 available stocks, even when there are no transaction costs. Thus, diversification in only a small number of assets takes place.

d. Even in an ideal situation, the investors choose inefficient portfolios located to the right of the mean-variance efficient frontier. Yet the fundamental risk-return CAPM or GCAPM linear relationship is intact, explaining about 70% of the variability of the rates of return. However, recall that we have here an ideal situation for the investors, and if all were rational and efficient we would obtain 100% explanatory power (i.e., a perfect fit). Thus, we conclude that some investors are irrational or do not know how to achieve their investment goals.

The interesting question is: What is the effect of irrationality (or errors in the decision making) of a portion of the investors on equilibrium models and price dynamics? Following the experimental findings, in the LLS microscopic simulation model, which is described in Chapter 7, we introduce investors who employ past rates of return as an indicator for the future rate of return distribution, and we allow for deviations from the optimal expected utility strategy. Moreover, in some cases, we have two groups of investors parallel to the experimental findings: one group can be considered as "efficient machines," who select an efficient portfolio, and the other group commits errors, hence investors in this group may select inferior portfolios. In such a scenario, there is no theoretical model that analyzes equilibrium prices and, in particular, the price dynamics. MS provides a framework to analyze such realistically complex scenarios.

5

THE MICROSCOPIC SIMULATION METHOD

5.1. INTRODUCTION

Imagine a stock market with only three investors: a fundamentalist, a portfolio rebalancer, and a technical trader. Suppose that you had perfect knowledge about the strategies employed by these three investors. For example, assume that

a. The fundamentalist employs Gordon's dividend model to calculate the stock's fundamental value and buys some quantity of stock (which depends on the investor utility function) whenever the stock is undervalued.
b. The portfolio rebalancer always invests 70% of his wealth in the stock.
c. The technical trader buys some quantity of stocks after three consecutive positive returns and sells after two consecutive negative returns.

For the simplicity of this example, assume for the time being that everything is deterministic. Given certain initial wealth allocations to these three investors, could you determine the price dynamics in this stock market?

Conventional modeling techniques would suggest writing down an equation (or equations) for the price as a function of time and trying to solve it. It is possible, in principle, to write a set of equations for the stock price at time t. These equations would have to involve past returns and the wealth of each investor at time t (which is a function of the entire sequence of prices and buy-sell orders from time 0 up to time t). Usually, these equations will not admit analytical solutions. Thus, the conventional approach would most likely not work for this simple model, let alone for more complicated and non-deterministic market models. Is there any way to estimate the price dynamics in this market?

Yes, there is a way. And it is simple: one can take three people, give them "paper" money according to the initial wealth allocations, assign a role of one of the investors to each person and explain the trading strategy that he has to follow, let them trade, and record the sequence of market-clearing prices. This is, in principle, what is done in microscopic simulation. The only difference is that instead of using people with paper money, the simulation is performed on the computer. In other words, the three agents with their trading strategies are represented in a computer program. This program simulates the actions of each trader at each time period and calculates and records the price at each period.

The main idea of the microscopic simulation (MS) method is to study systems by representing and keeping track of each individual microscopic element of the system on the computer (in the previous example, the microscopic elements are the investors) and by recording macroscopic variables of interest (such as prices, volume, etc. in the preceding example).

The great advantage of the MS methodology is that systems that do not yield to analytic treatment can easily be studied. Moreover, in the context of economics and finance, employing MS allows the relaxing of assumptions that are made for the sake of analytic tractability, and modeling investors' behavior as empirically and experimentally observed. A disadvantage of the MS approach is that it is more difficult to reach general conclusions from simulations than from analytical results. Namely, one can find out exactly what happens in a deterministic system with a given set of parameters, by running a single simulation. However, if the system is not deterministic, or if the initial conditions are not precisely known, one may need to collect statistical data from many simulations.

The focus of this book is MS in economics and finance. As an introduction to the MS method, this chapter presents a detailed example of one of the original and classical applications of MS: the application to nuclear fission systems—A-bombs and nuclear reactors. The purpose of this example is to illustrate the main ideas of MS in a simplified framework. With this relatively simple example in mind, the reader can better digest the MS in finance and economics discussed in later chapters. In Section 5.2 we describe the problem of designing nuclear devices and how MS is applied in these situations. In Section 5.3 we discuss some of the general considerations that should be taken into account when using MS. In Section 5.4 we compare and contrast the MS method with more traditional analytical "continuum" methods. We conclude in Section 5.5 with some philosophical remarks about the MS approach and reductionism in different fields of science.

5.2. A SIMPLE EXAMPLE OF A MICROSCOPIC SIMULATION APPLICATION: NUCLEAR FISSION

One of the basic MS methods was invented by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), who were involved in the (A-bomb) Manhattan project (and the subsequent thermonuclear reactions projects). This was not a coincidence: the developers of nuclear fission projects were in desperate need of a reliable, robust research tool, capable of describing, explaining, and predicting the macroscopic phenomena (e.g., nuclear explosion) emerging from a multitude of microscopic elementary interactions (e.g., individual atom decays).

The collective properties of a large set of uranium atoms are quite dramatic: below a certain mass and density, and for certain geometric configurations, the system is an inert slob of metal. However, given the appropriate geometry, mass, and density conditions, the system undergoes a collective explosive process named the "chain reaction." The transition between these two regimes is very sharp, and the smallest changes (say a 0.01% change in density) can turn in a fraction of a second a quiet metal piece into an exploding A-bomb.

Nuclear power plants (reactors) exploit the narrow intermediate dynamical region between these two extremes and require a precise control of the system. In fact, in the absence of better computational design and control tools, the first nuclear reactor, built in 1942 by E. Fermi under the grandstands of the University of Chicago Stadium, was found *a posteriori* to be very close to the critical (exploding) conditions. Nowadays, the optimal geometrical configuration and the safe steering of the nuclear reactors are carefully designed using MS, such that the chain reaction does not explode or dwindle but is rather kept in a steady regime with a steady energy output: controlled fission.

The exposition that follows parallels (very schematically) the modern way in which nuclear plants are actually designed and safety controlled by MS (Figure 5.1 displays such a professional design; we include it just for a general impression, but the reader should ignore the details and the names of the different design elements). However, our present objective is not to

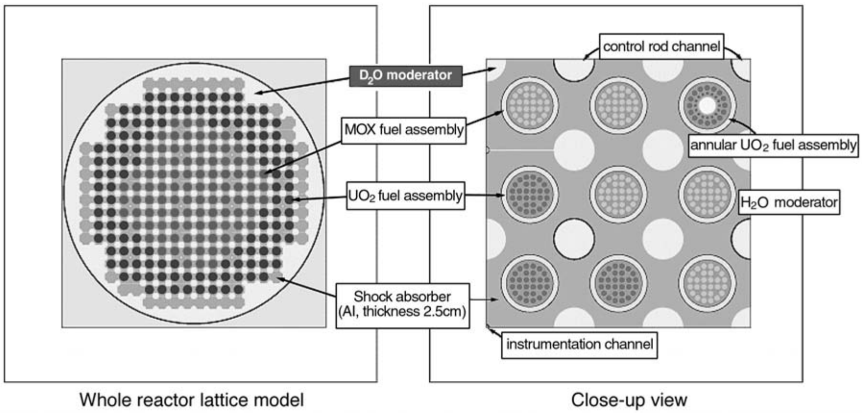


FIGURE 5.1 Cross-sectional view of a nuclear reactor constructed by computer. (From The Japan Atomic Energy Research Institute, 1997, *Persistent Quest-Research Activities 1996*, p. 51, with permission).

provide a faithful description of the actual computer programs used in practice. Rather, we use the nuclear chain reaction example to illustrate how one can predict macroscopic phenomena (in this case a nuclear explosion) using microscopic simulations.

The main microscopic phenomenon allowing the atomic reactors and atomic bombs to function is the fission of the Uranium 235 nucleus (from now on U235) when hit by a neutron. To follow the example, the reader does not need to know what a neutron is or what U235 is: the only things he or she has to know is that upon fission U235

a. Splits into smaller parts
b. Liberates energy
c. Emits (typically 2 to 3) new neutrons

Without any additional information, it is clear that if there are other U235 nuclei in the neighborhood, the neutrons resulting from one fission reaction may hit some of the neighboring U235s and produce new fission reactions, as shown in Figure 5.2. If there are enough U235s in the neighborhood, the neutrons resulting from the second generation of fissions may hit U235 nuclei again, and so on. The result will be a chain (or rather "branching tree") of reactions in which the neutrons resulting from one generation of fission events induce a new (larger) generation of fission events by hitting new U235 nuclei.

To get an intuitive feeling of the numbers involved, one can assume (quite realistically) that two of the neutrons produced in each and every fission reaction are hitting within 10^{-3} seconds two other U235 nuclei and cause them to fission. Then, the number of neutrons in the system would

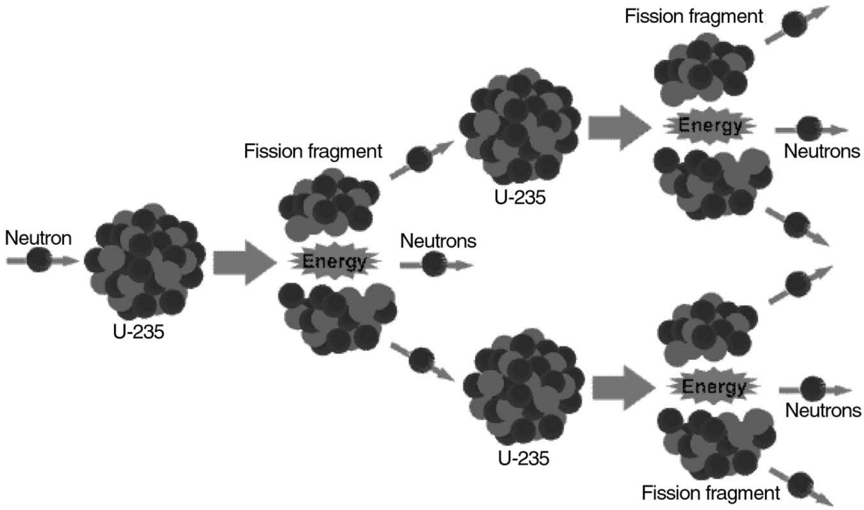


FIGURE 5.2 The beginning of a chain of nuclear fission reactions. From Clyde, Schleier-Smith, Tseng, Latham, Higley (1996).

double every 10^{-3} seconds. After only 0.08 seconds their number will reach 2^{80} neutrons. By that time, 2^{80} U235 nuclei, which are about 1kg of Uranium, would have fissioned. (We neglect here the time necessary to trigger the first fission in the chain: we assume there is always around one neutron from "the neighborhood" to "pass by" and start the entire chain.) Eventually, almost the entire available U235 population is exhausted and their corresponding energy is emitted: this is the atomic explosion. On the other hand, if the density of U235s in the sample is relatively low such that a neutron is more probable to exit the sample before hitting a U235, even if a reaction is started, it will die out and no explosion will occur.

In a nuclear power plant, the objective is to keep the uranium in a macroscopically active state but without exploding. This is called controlled fission. To this effect, the number of neutrons "flying around" has to be kept rigorously under control: even deviations of 0.1 % in the rate of neutrons hitting U235 nuclei may lead to macroscopically significant effects. In a reactor, the U235 dense regions are separated by neutron-absorbing regions ("control rods" made of cadmium or graphite). This allows regulation of the number of neutrons capable of carrying on the chain reaction: the more control rods, the less neutrons will hit U235s and vice versa.

The question when designing a nuclear reactor is whether a given configuration and density of U235s and control rods will lead to an explosion, to controlled fission, or to the halt of the reaction. While it is possible to obtain analytical results for some simplified settings, for general configurations MS is the best tool to answer this question.

Let us demonstrate this with a simplified model of the nuclear fission in an atomic reactor. Imagine one is given a square mesh representing space (Fig. 5.3). On each site/square of the mesh there might be placed any number (including 0) of U235 nuclei (denoted by U's) or neutrons (denoted by n's), or a control rod (denoted by O). We assume the nuclear reactions are governed by the following set of simplified dynamical microscopic rules:

1. From one time step to another, the n's can jump to a neighboring site (let's assume for simplicity that they may jump with equal probability (one-fourth) up, down, left, or right to any of the four neighbors of the site they are currently occupying).
2. If an n is placed on a border site, it may leave the system (e.g. if an n is placed on the right-most column, there is a probability of one-fourth that it will jump further to the right and leave the system).
3. If an n and a U happen to meet on the same site, the U fissions and emits a second n. Operatively, this is represented by substituting on the mesh the fissioning U by one n.
4. If an n reaches a site in which a control rod, O, is located, the n is absorbed (disappears).

The objective of the present MS exercise is to estimate whether the global system will explode or not, or, more generally, the number of fissioned U's during a given time period. The MS basics are best explained by pictures describing the time evolution of the model rather than by formulae. Therefore, let us describe by a sequence of "cartoons" the MS of the chain reaction.

Given an initial U and O configuration, as shown in Figure 5.3, one starts the microscopic simulation with the assumption that one stray neutron, n, from the environment hits one of the U's. One can imagine that the computer extracts the coordinates of the U, which undergoes the initial fission event from a virtual urn.

Let's assume that in the present case the initial fissioning U is the one at the position (5, 6) (i.e., a neutron hits the U placed at the intersection of column 5 with row 6). We represent this in Figure 5.4 by placing an n next to the U on the box with the appropriate coordinates. As a consequence, the U on site (5, 6) fissions. This is represented operationally in Figure 5.5 by substituting the fissioned U with an n on the same mesh site. Consequently, the new configuration of the system will present two n's on the

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				U			O			
5	U									
6					U					
7				O			O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.3 An initial setting of the U nuclei (and absorbing bars O).

site (5, 6). Each of the two neutrons resulting from the fission site starts moving around by jumping randomly to neighboring locations (the jumping direction—up, down, left, right—is chosen at each time randomly with equal probability). Let us discuss one possible realization of the dynamics. In Figure 5.6 one neutron jumps from the initial square (5, 6) to the left neighbor (4, 6) while the other neutron jumps to the neighbor from below (5, 7). At the third time step (Fig. 5.7), one sees the neutron from (4, 6) moving up to (4, 5) and the neutron from (5, 7) moving right to (6, 7). At the fourth time step (Fig. 5.8), the neutron from (4, 5) jumps to (4, 4) and

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				U			O			
5	U									
6					Un					
7				O			O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.4 The triggering neutron is placed at site (5, 6).

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				U			O			
5	U									
6					nn					
7				O			O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.5 The fission of the U at site (5, 6) leaves two n's at the location.

triggers the fission of the U residing on that site. This results in two neutrons in Figure 5.8 on the site (4, 4). The neutron that is located at (6, 7) at the third time step moves to the right at the fourth time step and is absorbed by the absorbing bar O. This explains its absence from Figure 5.8.

If one continues the simulation, there are good chances that the two neutrons at (4, 4) will keep moving and hit other U's (maybe the ones at (3, 2) or (2, 3)), producing more neutrons, and so on, until most of the U's are fissioned. If one wishes to avoid this, one may decide to change the

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				U			O			
5	U									
6				n						
7				O	n		O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.6 One of the neutrons from (5, 6) moves to the left, the other moves down.

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				U			O			
5	U			n						
6										
7				O		n	O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.7 The neutron from (4, 6) moves to the left; the neutron from (5, 7) moves right.

geometry of the system by introducing more neutron-absorbing material *O* at various sites.

By appropriately introducing and removing neutron-absorbing bars, one can control the reactor: start or stop the reaction or keep it at a constant rate (or let it explode). In the case of the bomb, to accelerate the reaction, one would try to keep as many neutrons as possible in the system. In particular, one can prevent the neutrons from leaving the system by

	1	2	3	4	5	6	7	8	9	10
1					U					
2			U					U		
3		U								
4				nn			O			
5	U									
6										
7				O			O			U
8				U				U		
9		U				U				
10									U	

FIGURE 5.8 The *n* moving from (4, 5) to (4, 4) encounters a *U* and "multiplies." The *n* from (6, 7) moves to (7, 7) and is absorbed.

placing at the boundary neutron-reflecting plates (i.e., neutrons leaving the system will be reintroduced).

By repeating the simulation with different realizations of the random moves of the *n*'s (and of the site of the initial *n* hit), one will generally obtain for the same system different results. In particular, the number of *U*'s that undergo fission may vary from one realization to the other. For instance, if at stage Figure 5.8, the neutron at (6, 7) instead of moving to the right would have chosen one of the other stochastic possibilities (e.g., moving up), then it would not have fallen on the absorbing site and its life might have continued with additional *U*'s being fissioned (and additional "descendent" *n*'s being created). The prediction about the behavior of the macroscopic system can be obtained by running many simulations with different microscopic realizations, counting the number of *U*'s fissioned in each simulation, and statistically processing these data. In particular, for a given system configuration one can estimate $P(k)$, the probability for the fusion of k nuclei.

For small systems, this probability will be influenced by small details (e.g. the fission of 2 *U*'s represents 15% of the 13 *U*'s present in the previous example). However, for very large systems the data are more clustered and the statistics over the various microscopic realizations of the same macroscopic system will return results that are less sensitive to the details. In computational terms, this situation means that for large systems, a very small number of MS runs is enough to inform us with high confidence of the expected macroscopic behavior of the system. Obviously, simulating a nuclear reactor with a realistic number of nuclear particles (in the order of 2^{80} particles) is not feasible because of computational constraints. One of the trade secrets of MS is to find the optimal size of the simulation system: small enough to allow its handling by the computer and large enough to provide results representative for the real macroscopic system.

In the previous example, the model of nuclear interactions is extremely oversimplified: we assumed the *n* can move only in four directions, we ignored other products of the fission, and so on. This has been done partly for the sake of simplifying the example. However, we are also trying to make a deeper point. It turns out that the macroscopic behavior of most statistical systems with a large number of interacting "microscopic" elements, is insensitive to many details of the microscopic interaction. For instance, in the previous example, if we were to allow the *n*'s to move diagonally as well, the macroscopic properties of the system would not have changed. Another trade secret of microscopic simulators is the ability to figure out which elements of the dynamics are essential and which details can be neglected without affecting the macroscopic results. We will touch on this point in the next section.

As a parting remark to this first encounter with the MS method, let us compare it with the alternative approach of analytical treatment. One may insist that given all the positions of the U's and O's, and the probabilities for the various n moves, one can, in principle, compute directly the probability $P(k)$ of having k fissions of U's during the stochastic chain reaction process defined by the elementary interactions 1 through 4. However, this is practically impossible. Even if the n's can move only in four directions, following a single n for 10 time periods requires calculating 4^{10} possible paths. With two n's this becomes 4^{20}, and so on. Obviously, calculating all possible outcomes is not feasible, even for such a small and simplified system. Instead, one can use the MS that enacts on the computer a limited set of realizations of the stochastic process and extracts information by performing statistical measurements on this limited set of instances.

5.3. CONSIDERATIONS IN APPLYING MICROSCOPIC SIMULATIONS

The idea of MS in economics and finance is very simple. However, in implementing this methodology many technical questions may arise: What are the "microscopic elements" that should be used as the basic building blocks of the model? Individual investors? Households? Funds? Institutions? Which microscopic details should be modeled, and which can be neglected? How many microscopic elements should be simulated? Perhaps 10 investors? What about 10^6 investors? In this section we comment on some of these issues.

5.3.1. Identifying the Relevant Microscopic Elements

In the previous section we argued that complex macroscopic systems can be described, explained, and predicted by the MS of the interactions between their "elementary atoms." How does one decide what these atoms should be? If they are chosen too close to the macroscopic level, the strength of the resulting scientific explanation is diminished: explaining that a car works because it has wheels and an engine is not very illuminating about the heart of the matter: Where does it take the energy and how does it transform it into controlled mechanical motion? If the atoms of explanation are sent to too fine scales, one gets bogged into irrelevant details: for example, if one starts explaining the quantum mechanical details of the oxidation of hydrocarbon molecules (fuel burning), one may never get to the point of the car in the previous example.

It is a common feature of the human perception and understanding that for each task and system there is an appropriate scale to represent it.

The scale should be fine enough such that all salient features would be represented on it and coarse enough that these features will not be cluttered by unessential details.

The decision of the appropriate scale and objects for representing a system is crucial for the outcome of the research. In fact, the most severe and costly failures of traditional science are due to the mistaken identification of the appropriate level of representation. Some of these failures can often be traced up to the traditional partition of science into disjoint departments. The result of these artificial boundaries is that the phenomenology of a problem often falls within one discipline while the appropriate concepts/objects for finding and representing its solution lie outside disciplinary boundaries. The emerging macroscopic collective subsystems are in a sense conceptual fictions in as far as they contradict in detail the microscopic underlying laws. Their very existence is in ostensive contradiction with the microscopic overt structure of the system. In deciding whether to describe the macroscopic dynamics in terms of such collective variables, one is in a conceptual dilemma: to describe an ocean storm in terms of ill-defined, ephemeral "waves" or in terms of the well-defined, perennial (but useless) water molecules?

The relevant microscopic element can be defined by the property that going into finer details of description does not change or add to the understanding of the macroscopic dynamics. For example, constructing a stock market model by using as elementary building blocks the cells making up the human bodies of the investors trading in the market may be technically correct, but it is entirely superfluous. Starting with individual investors as elementary building blocks is detailed enough. Sometimes, the relevant microscopic element is not clear. In the context of financial markets, should the elementary building block be an individual investor? A household? A mutual fund? Or perhaps an investment institute?

If one has strong intuitions about the relevant microscopic elements of the system investigated, these intuitions can be tested and verified by simulating the system at the level of detail of the suggested building blocks and at finer-detail levels. If the finer-detail level simulations yield the same result as the suggested building-block level simulations, then indeed, the extra details are unnecessary. In other situations, the relevant microscopic building blocks are unclear but they can be identified through the simulation itself. In many instances, MS models have been formulated at one level of detail, while the simulated dynamics revealed that the relevant building blocks are actually much coarser. For instance, in a recent simulation of diabetes 1, it was found that instead of describing the dynamics of the illness at the level of the immune cells, it is sufficient to use as elementary objects the pancreatic islets (Louzon, Solomon, Atlan, and Cohen, 2000).

5.3.2. Identifying the Relevant Interactions (Universality)

The popularity of the MS method in the fields of statistical physics and particle physics is partly due to the discovery that many of the large-scale properties of the macroscopic systems are independent on the details of the microscopic elementary interactions. This is related with a series of phenomena that go under the name of criticality, scaling, and universality. Recent works (Shnerb, Louzon, Bettelheim, and Solomon, 1999) have shown that these properties arise naturally in many statistical systems. The macroscopic properties of such systems are therefore robust to changes in the details of the microscopic interaction.

In the context of financial models there are questions like “When is the next crash going to happen?” which, while macroscopically interesting, do depend on microscopic details. In such situations, the issue of how much microscopic detail is necessary has to be dealt with on a trial-and-error basis. A high sensitivity of the results to the microscopic details might signal a genuine instability of the actual system under study. Even in this situation, MS may be quite useful in characterizing (although not predicting in detail), the behavior of the system under study. For instance, the exact timing of a market crash may not be predictable, but MS studies may indicate correctly the states of the market when a crash is likely to happen and the intervention necessary to avoid it.

5.3.3. System Size Considerations

When investigating a large stochastic system with many interacting elements (such as a stock market), one has to decide how many microscopic elements (investors in the stock market example) to include in the simulation. There is a trade-off: on the one hand, the more elements simulated the more realistic the model. On the other hand, more elements mean more calculations and therefore more time-consuming simulations. For example, in nuclear fission simulations it is impossible to simulate a system with a realistic number of typically 10^{24} elements. The nuclear fission system described in Section 5.2 can be thought of as a coarse model of reality: each U represents a cluster of many atoms. The model can be refined by making the cells finer (for example, by splitting each cell into four smaller cells) and splitting each U into four smaller U's. Fortunately, in many systems the fact that the model is a coarse version of reality does not affect the macroscopic properties much: a small number of simulated elements produce macroscopic behavior that is also characteristic of the more refined real system, which has a larger number of microscopic elements.

If the macroscopic properties of the system do depend on the number of simulated elements, one can still use MS to obtain an estimation of the properties of the real system, even if it has a very large number of elements. To be specific, one can find the relationship between the number of elements and the macroscopic property investigated within a certain range, and then one can extrapolate to estimate the properties of a system with a realistically large number of elements. In certain cases, there are analytical ways to estimate the error of the MS originating in "finite size effects"—the fact that the simulated system has fewer microscopic elements than the real system.

A recent analysis of the system size effects in the MS of markets has attracted attention to the following problem: How many of the investors are really influencing the dynamics of the S & P index (Stauffer, 1998, 1999)? The advent of many new Internet traders raises the issue of the influence of the number of traders on the market behavior.

5.4. THE EFFECTS OF DISCRETENESS—COMPARISON WITH THE ANALYTICAL "CONTINUUM" APPROACH

The traditional approach to the study of dynamical systems such as the nuclear fission previously discussed was to express them in terms of (partial) differential equations and to solve those by analytic methods. To do so, one has to assume that the system is continuous. More precisely, one has to substitute the discrete systems of n's and U's by continuous functions of position and time $fn(r, t)$ and $fU(r, t)$ representing the density of elements in a particular space-time neighborhood of (r,t). The microscopic evolution rules of the discrete system (of the type 1 through 4 in Section 5.2) translate in the continuum formalism into partial differentials acting on the $f(r, t)$'s.

It is difficult in the age of digital computers to convincingly defend this change from discrete to continuum quantities (especially when this is at the price of missing quite crucial features of the system). Certainly, this affects rather negatively our intuitions on the system: we never meet in real life density distributions of our friends, spouses, or children but rather individuals of those sorts, and we treat them accordingly. However, most of the classical mathematical analysis tools depend on the continuity assumption. In the presence of computers, we are starting to probe the essential features that were missed by this assumption and, in particular, the details of the birth of macroscopic collective objects out of the microscopic noisy individuals.

In spite of this criticism of the density field approach and its kindred, one has to admit that much of the knowledge one has now on dynamical systems originates in the continuum approach. We couldn't live a second in (let alone make sense of) real life without our capability to isolate, identify, and "call by collective names" large spatio-temporal collections of microscopic individual entities. Therefore, it is worth making clear the conditions in which one can trust it. Let us return to the square mesh picture of the previous section. Let $Rn(r,t)$ be the number of microscopic agents of type n (neutrons) in the unit box located at r. Let $RU(r,t)$ be the number of U's in the same box. To approximate the integers Rn and RU by real numbers, one has to perform averages that involve a large number of them. This can be realized either by averaging over large enough volumes or over long enough time periods. This means that in defining the corresponding continuous fields $fn(r,t)$, $fU(r,t)$, one will have to average over space-time intervals large enough to involve a macroscopic number of elements. On the other hand, to capture with enough precision the relevant dynamical features, the averages have to be taken over space-time intervals that are smaller than spans over which significant changes in the R values take place. The continuum/averaging approach breaks in the cases in which these two conditions are contradictory. For instance, for cells in the immune system, the density is a few thousands units per cm-cube, which means the continuity assumption breaks already for phenomena in which there are significant spatial variations at distances of the order of 1 mm. In financial systems, where the price dynamics obviously consists of jumps (taking place at the individual trade times) rather than a continuous flow, the continuity assumption is invalid for phenomena where there is significant price variation from one individual trade to another.

There exists a remarkable class of systems in which the discretization is crucial, and the assumption of continuity leads to invalid results. It was shown that if the microscopic agents are auto-catalytic (i.e., the existence of an agent induces the creation of a similar agent at the same location), then the slightest microscopic noise is amplified to macroscopic inhomogeneity (Shnerb, Louzon, Bettelheim, and Solomon, 1999). This leads to the generic emergence of localized macroscopic objects with collective complex properties. Assuming microscopic continuity entirely misses this effect.

Averaging is not always a liability. Even in the MS studies, one often chooses to analyze the end results of many simulations by statistical means. The point is that the average of the end results is not necessarily equal to the end result of the average dynamics. For instance, by sending ships over the ocean one may find that in average a certain percent is lost by the end of the trip due to violent storms. The average survival expectation would be greatly overestimated by the naive assumption that the ships sail over an averaged ocean (with all the big waves canceled by the big gaps in between them). Much of the salient properties that make life, ecology, economy, and markets possible revolve around this "petite difference."

In conclusion, while it is a very powerful tool, the continuum approach is only an approximation of the discrete microscopic reality. Whenever the conditions of the approximation do not hold, or whenever the system is analytically intractable, one has to revert to the original discrete microscopic system and to the MS method.

5.5. PHILOSOPHICAL REMARKS

Until recently, science was divided in "hard" versus "soft" disciplines. With a slight oversimplification, "hard" sciences were fields where the problems were formulated and solved within a mathematical language based mainly on (partial differential) equations. The archetypal "hard" science was physics and for any discipline to progress quantitatively it was supposed to emulate its example. In fact, chemistry became "harder" as the use of the Schroedinger equations to describe chemical processes became possible. Biology, recognizing the impossibility of reducing its matter to equations, developed a defensive attitude of mistrust against theoretical explanations while economics had often chosen the opposite road: to sacrifice detailed quantitative confrontation with empirical data for the sake of closed tractable models. The psychology and creativity studies have chosen to ignore the mechanistic/analytic methods altogether sometimes even taking pride in the ineffable, irreducible, holistic character of their subjects. This was natural, as the mechanisms allowing the understanding of phenomena in those fields are most probably not expressible in the language of equations.

Fortunately it turns out that equations are not the only way to express understanding quantitatively. In fact, with the advent of computers one can express precisely and quantitatively almost any form of knowledge. For example, adaptive agents can "learn" certain tasks by just inputting lists of "correct" and "incorrect" examples, without any explicit expression of the actual criteria of "correctness" (sometimes unknown even to the human "teacher"). This relaxation of the allowed scientific language and the focus on the emergence of collective objects with spatio-temporal (geometrical) characteristics renders the scientific discourse more congenial to the daily cognitive experience and to practical applications. One can now formulate in a precise, "hard" way any "naive" explanation by simulating (enacting) in the computer its postulated elements and interactions. One can then verify via computer experiments whether these postulates lead indeed to the effects observed in nature. As opposed to the human mind, the computer can deal simultaneously with a macroscopic number of agents and elementary interactions and therefore can bridge between microscopic elementary causes and macroscopic collective effects, even when they are separated by many orders of magnitude.

The MS method provides a framework within which one can systematically follow the birth of the complex macroscopic phenomenology out of the simple elementary microscopic laws. The MS paradigm consists of deducing the macroscopic objects (macros) and their phenomenological complex emergent laws out of a multitude of elementary microscopic objects (micros) obeying simple rules. The macros and their laws emerge then naturally from the collective dynamics of the micros as their effective global large-scale features. By emergence we mean the appearance of macroscopic collective objects with properties that are not present (or even definable) at the level of their microscopic components. In the next chapter, we review examples from various fields in the social sciences in which MS is used in order to investigate the dynamics of complex systems.

6

MICROSCOPIC SIMULATIONS IN VARIOUS FIELDS

6.1. INTRODUCTION

Microscopic simulation (MS) was first developed and applied in the physical sciences. The A-bomb microscopic simulation described in Chapter 5 was one of the first applications of this method. The success of MS in the physical sciences in solving problems that are analytically intractable has caused many other fields to adopt the MS methodology over the years. Although the MS methodology has been applied to very diverse systems ranging from systems of atomic particles to systems of cars, animals, or even humans, the idea of the MS method in all these different applications is the same: by modeling the behavior of the elementary building blocks of the system (atomic particles, cars, or individuals) and their interactions (e.g., when a neutron hits a uranium nucleus, the nucleus splits and neutrons are released), and by representing these building blocks in a computer simulation, one can analyze the dynamics of the macroscopic properties of the system.

In this chapter we review some of the applications of MS in a range of fields. The aim of this review is to place the MS methodology in a wider perspective. There is no pretension of establishing credit or of exhaustiveness. Such an attempt would be bound to failure given the great variety of fields over which the applicability of MS has spread. The naturalness of the ideas behind MS has lead to their rediscovery over and over again in different fields. As a result, different names were given to this methodology: microscopic simulation, microsimulation, agent-based simulation, and so on. There is no unique universally agreed meaning to these terms, and, in fact, there is no unique universally agreed way to apply these ideas. This is not necessarily a drawback because it allows adapting the MS paradigm in the most fitting way to each problem at hand.

In Section 6.2 we describe the application of MS to traffic flow and freeway planning as well as to the study of the generic emergence of traffic jams. In Section 6.3 we describe two population models that present space-time localization features. In Section 6.4 we mention a few MS methods that have been used for a long time in social sciences, especially in welfare planning. In Section 6.5 a more ambitious aim is considered: to explain from simple microscopic interactions the diversification of the population of individuals. This means that the behavior of the individual agents rather than being measured from the real population statistics is predicted from simple assumptions. In Section 6.6, the use of MS techniques shows how the expectations of the individuals on the behavior of the other individuals have often the property of reinforcing *a priori* unstable (and inefficient) global situations. Sections 6.7, 6.8, and 6.9 deal respectively with applications in archeology, marketing, and biology.

Though most of the book is devoted to MS in economics and finance, this chapter focuses on MS in other fields. We recommend Moss de Oliveira, de Oliveira, and Stauffer (1999) for further examples in various biological, financial, and political systems and Anderson, Arrow, and Pines (1988) for examples from the economic realm.

6.2. TRAFFIC FLOW MICROSCOPIC SIMULATIONS

One of the most erratic, costly, and far from equilibrium systems affecting the day-by-day life of millions of people is the traffic flow on freeways and inside cities. Assuming traffic continuity is usually inappropriate: often the transition from freely flowing traffic to jams is discontinuous both in space and in time (in fact, this is often the source of chain accidents on the freeway). Moreover, the statistical knowledge and the intervention possibilities in traffic systems justify a sustained engineering effort in simulating, understanding, and controlling traffic. Following Douglas and Lewis, (1970) a general view on the MS paradigm from the point of view of road traffic has evolved:

> Simulation modeling is an effective tool for analyzing a variety of dynamic problems whose process is too complex to describe in analytical terms. Many entities interact simultaneously during the process that can not be represented logically and mathematically. Simulation models mimic the interactions and behaviors of the systems and integrate these behaviors and interactions to produce a detailed, quantitative description of the system performance. Simulation models execute on a digital computer as software and represent a real-world system in an experimental fashion . . . . Traffic simulation software creates a scenario (highway network configuration, traffic demand) as an input model to describe the system operations in statistical, graphical, and animated form. The numerical results provide the analyst with detailed quantitative description of what is likely to happen and to provide insight of why the system is behaving this way.

The theoretical study of traffic flow was performed using statistical mechanics techniques by Biham, Middleton, and Levine (1992) (see also Nagel, Esser, and Rickert, 2000). Their model, and the vast bulk of studies following it, is in strong relation with the theoretical understanding of phase transitions in statistical mechanics systems. While they do not model particular roads and intersections, they describe and explain convincingly the generic features of the traffic flow and its points of instability (e.g., the outbreak of jamming):

> A simple model that describes traffic flow in two dimensions is studied. A sharp jamming transition is found that separates between the low density dynamical phase in which all cars move at maximal speed and the high density jammed phase in which they are all stuck. Self organization effects in both phases are studied and discussed.

Further MS studies of traffic flow may lead to a deeper understanding of the origins of traffic jams and may suggest generic techniques to avoid them.

6.3. POPULATION DYNAMICS, MOBILITY, AND SEGREGATION

One of the earliest models of population dynamics was the classical segregation model by Schelling (1978). Schelling's model is not intended as a realistic tool for studying the actual dynamics of specific communities as it ignores economic, real-estate, and cultural factors. Rather, the aim of this very simplified model is to explain the emergence of macroscopic single-race neighborhoods even if the individuals are nonracist. More precisely, Schelling found that the collective effect of neighborhood racial segregation results even from individual behavior that presents only a very mild preference for same-color neighbors. For instance, even the minimal requirement by each individual of having (at least) one neighbor belonging to one's own race leads to the segregation effect. This model is implemented and described on the Internet Education Project page of the Carl Vinson Institute of Government (1999) as follows:

> Schelling created a simple model in which a given simulated agent (or set of agents) prefers that some percentage of her neighbors be of her own "color." If an agent is residing on a square or cell and the agent does not have at least that percentage of her own kind nearby, the agent moves to another square or cell. What Schelling discovered was that even when the agents in the simulation had only a weak preference for nearby agents being of the same "color," the result after several moves (or chances for each agent in turn to move to another square) was a fairly high level of segregation.

More operationally, the simulation starts with a square mesh, or lattice, (representing a town) which is composed of cells (representing houses). On these cells reside agents, which are either "blue" or "green." One fixes then the crucial parameter: the minimal percentage of same-color neighbors that each agent requires. Each agent, in his or her turn, examines the color of all his or her neighbors. If the percentage of neighbors belonging to his or her own group is above the "minimal percentage," the agent does nothing. If the percentage of neighbors of his or her own color is less then the minimal percentage, the agent moves to the closest unoccupied cell. The agent then examines the color of the neighbors of the new location and acts accordingly (moves if the number of neighbors of his or her own color is below the minimal percentage and stays there otherwise). This goes on until the agent is finally located at a site in which the minimal percentage condition holds. After a while, however, it might happen that following the moves of the other agents, the minimal percentage condition ceases to be fulfilled and then the agent starts moving again until he or she finds an appropriate cell. As mentioned earlier, the main result is that even for very mild individual preferences for same-color neighbors, after some time the entire system displays a very high level of segregation.

A more modern, developed, and sophisticated reincarnation of these ideas is the Sugarscape environment described by Epstein and Axtell (1996). The model considers a population of moving, feeding, pairing, procreating, trading, warring agents and displays various qualitative collective events that their populations incur. By employing MS one can study the macroscopic results induced by the agents' individual behavior. This model has been widely covered in the media during the recent years. Wright (1997) describes the Sugarscape model:

> In Sugarscape, dots representing people or families move around a digital landscape in search of food—sugar. Whether they live or die depends on whether they find enough food to satisfy their "metabolic" needs. The dots, or "agents," are given a range of abilities—such as how far they can "see" over their virtual

> landscape when searching for food—and are programmed to obey certain rules. In the most basic scenario, the agents look for the richest source of sugar, and go there to eat it. But they are in competition with each other and with groups of agents programmed with different rules and abilities. By modifying the rules governing how the agents interact, Axtell and Epstein can make them either fight or cooperate. They can allow the agents to accumulate wealth, excess sugar, and measure their "health" by how much sugar they eat. By introducing mating, the researchers make the agents pass on their abilities—and the rules they obey—to their offspring.

The Sugarscape model makes a forceful point of the power of MS. It also opens the way to a much wider range of further studies that aim to translate the visual and anecdotal conclusions of particular simulation runs into generic quantitative scientific predictions that are to be compared with the empirical data.

6.4. MICROSIMULATION IN SOCIAL SCIENCE

In a more pragmatic context, the social scientists starting with Orcutt in the 1950s have collected the life statistics data of many taxpayers and fed them back into microsimulation programs. They used these programs to predict and plan the collective future behavior of the population from which the statistical data were collected (see Orcutt, Caldwell, and Wertheimer, 1976). The idea is simple and natural: suppose one wishes to estimate the cost to the government of lowering the retirement age from 65 to 60. One can do so if one has the appropriate life statistics that allow this estimation. In particular, one needs to know which people will retire under the new law in excess to the ones that would have retired under the old law. Moreover, for each of them one has to estimate what change the retirement will cause to his or her health state, contribution to the gross national product, income, taxes, and expenditures. Of course, it also would be important to estimate how long those who retire are going to live after retirement, and so on. Obviously, nobody can know all these numbers in advance for each person. Therefore, the microsimulation method assumes that the behavior of the various taxpayers in the future will be similar to the past behavior reported by the statistical poll. This method, as applied to the tax and subsidies policy, is described in Caldwell (1993):

> Conceived by Guy Orcutt in 1957, dynamic microsimulation is a "bottom-up" strategy for modeling the interacting behavior of decision makers (such as individuals, families and firms) within a larger system. This modeling strategy utilizes data on representative samples of decision makers, along with equations and algorithms representing behavioral processes, to simulate the evolution through time of each decision maker, and hence of the entire population of decision makers ... By summing over individual responses, the short-term and long-term consequences of the policy or trend change for the population can be traced.

This approach has been applied in various forms by quite a number of groups. Another description of the practical implementation of the microsimulation in social planning is given by Beebout (1999):

> Simply put, microsimulation is a type of computer program that simulates how a welfare program would operate under proposed changes and how participants would be affected. A microcomputer houses information about the welfare population, welfare program rules and operations, and how low-income families behave. The welfare population is represented by a micro database of administrative records, survey records, or a combination of both . . . The computer then processes each remaining family in the database, counting the families who are eligible and who participate, to produce the caseload estimate. It also adds up each family's benefits, producing the cost estimate.

By comparing these two definitions of microsimulation, one sees here a very typical feature about MS methods: it is very difficult to give a definition that would be specific enough and still contain all the views, techniques, and ideas that people associate with it.

A version of microsimulation that emphasizes the links to stochastic methods is presented in the Demographic-Sociological Microsimulation Program (1976), which

> simulates the interaction between demographic rates and sociological rules. A human population (with a specified ancestral genealogy, if desired) is simulated through time by the stochastic occurrence of demographic events to individuals in the population, in accordance with user-specified vital rates and sociologically defined decision rules. SOCSIM permits the user an approximation to experimental control—i.e., by specifying rates and rules, the users can explore what patterns of behavior might be expected to occur in a population characterized by those rates and rules.

A leading microsimulation model created for use in macroeconomics, called Aspen, is described in Pryor (1999):

> Aspen uses agents to represent the various decision-making segments in the economy. Each segment may have one or more agents representing it in an analysis. One of Aspen's key features is its ability to realistically reproduce the process used by actual economic agents to maximize utility or profit. Agents in Aspen not only can communicate with one another but also make "real-life" decisions. Through use of evolutionary learning techniques, the agents adapt their behavior according to changing economic conditions and past experience. In other words, they become smarter as they move through time.
>
> . . . Macroeconomic variables, such as gross domestic product, inflation (CPI), and the unemployment rate are computed as aggregate results of innumerable decisions by the individual economic agents.

The large bulk of work that has been produced in the social sciences by the microsimulation approach is exemplified in the recent book by Troitzsch, Mueller, Gilbert, and Doran (1996):

> This book gives an overview of the state of the art in five different approaches to social science simulation on the individual level. The volume contains microanalyti-

cal simulation models designed for policy implementation and evaluation, multi-level simulation methods designed for detecting emergent phenomena, dynamical game theory applications, the use of cellular automata to explain the emergence of structure in social systems, and multi-agent models using the experience from distributed artificial intelligence applied to special phenomena.

6.5. THE OUTBREAK OF COOPERATION, INDUCTIVE REASONING, AND INVESTMENT STRATEGIES

While the microsimulations described in the preceding section are specific and are based on empirical statistics, other researchers have addressed more generic aspects of social systems pertaining to the emergence of macroscopic cooperation in societies of interacting citizens. The beginnings of such studies were quite pessimistic as to the future of human societies. Not only was it not clear how social cooperation and structure may emerge, but in fact it seemed that a society of rational individuals is bound to fall apart (Hardin, 1968). For example, assume that it costs an effort equivalent to K cents to bring a unit of junk to the junk basket rather than throwing it on the sidewalk. Suppose, moreover, that each unit of junk on the sidewalk causes an average discomfort equivalent to $L < K$ cents to everyone in the community. It is clear that in these conditions, every time a rational citizen has a piece of junk on his hands he will promptly discard it on the sidewalk. Pretty soon the entire population of rational citizens will swim in sewage.

An elegant (although partial) solution to this paradox is offered by the iterated "prisoner dilemma" game of Axelrod (1988). The solution emphasizes the difference between "one-shot" decisions, which are made independently of the actions of others, and decisions that are related to the actions of other individuals and are made repeatedly. The MS method treats entire populations of interacting agents and allows for the study of collective emergent phenomena therein. The original noniterated "prisoner dilemma" game refers to a situation in which two police suspects are confronted with the following dilemma:

- If any of them agrees to serve as state witness ("defects"), he escapes punishment and the other suspect gets 4 years in jail.
- If none of them agrees to be a state witness (i.e., they "cooperate" among themselves), they both get 1 year in jail.
- If both "defect" (accept to be witness), both get 4 years.
- The suspects cannot communicate among themselves.

The apparent rational result provided by the mathematical game theory analysis is that each suspect will keep to his or her own interest (i.e., will "defect") and both will lose (get 4 years in jail).

Let us now describe the *iterated* game in a more social (MS-like) setting: Suppose there are many agents moving in a virtual space. Each time two agents meet, each has to decide whether he or she behaves cooperatively or aggressively. If both behave cooperatively, both get 1 point. If any of them behaves aggressively, he or she gets 2 points and the other gets 0. If both are aggressive, both get 0 points. The agents are endowed with memory: they recall which other agents they met in the past and what was the outcome of their encounters. Based on the behavior of the other individuals, each individual has the freedom to define a behavioral policy. This is the key point: once the game is played for more than one round, the agents have the opportunity to reward cooperation and punish aggression.

This opens the way to the actual study of the optimal strategies by MS computer experiments rather than by analytical reasoning. In fact, the issue was settled ultimately by "tournaments": agents using different strategies were "thrown" into the computer's virtual arena and left to have multiple encounters with their peers. The winning strategies in the long run turned out to be those that were on average more "cooperative" (Axelrod, 1985):

> [To achieve cooperation one] need not assume altruism, but rather selfish, egotistical players that can remember past cooperations and defections. ... [Axelrod] studied extensively the Tit-For-Tat (**TFT**) strategy. TFT is the strategy where the first move is to cooperate. Then simply play the move your opponent played last. This way if a TFT comes across any other player who is "nice" it cooperates. Any strategy is said to be "nice" if it does not defect first.

The main rules that Axelrod found to be efficient were:

- reward cooperation
- punish defection
- do not be greedy (or envious)
- do not be first to defect
- do not be too clever

It is quite a remarkable effect that while a single interaction leads to "noncooperative" results (according to the game theory analytic solution of the *noniterated* prisoner dilemma), the collective effects of many interactions lead to the "outbreak of cooperation."

Another simple game capturing the idea of (co-)evolution is the El-Farol minority game. This game throws a special light on the ways in which the interaction of many individuals may lead endogenously to the preference of otherwise arbitrary behavioral rules. *A priori* all social conventions are equally acceptable. Their validity originates only in the interaction with the rules that the other players happen to adopt at that particular time (Arthur, 1994):

> The Bar Problem
>
> Consider now a problem I will construct to illustrate inductive reasoning and how it might be modeled. *N* people decide independently each week whether to go to a bar that offers entertainment on a certain night. For concreteness, let us set *N* at 100. Space is limited, and the evening is enjoyable if things are not too crowded—specifically, if fewer than 60% of the possible 100 are present. There is no way to tell the numbers coming for sure in advance, therefore a person or agent goes—deems it worth going—if he expects fewer than 60 to show up, or stays home if he expects more than 60 to go. (There is no need that utility differ much above and below 60.) Choices are unaffected by previous visits; there is no collusion or prior communication among the agents; and the only information available is the numbers who came in past weeks. (The problem was inspired by the bar El Farol in Santa Fe which offers Irish music on Thursday nights; but the reader may recognize it as applying to noontime lunch-room crowding, and to other coordination problems with limits to desired coordination.) Of interest is the dynamics of the numbers attending from week to week.
>
> Notice two interesting features of this problem. First, if there were an obvious model that all agents could use to forecast attendance and base their decisions on, then a deductive solution would be possible. But this is not the case here. Given the numbers attending in the recent past, a large number of expectational models might be reasonable and defensible. Thus, not knowing which model other agents might choose, a reference agent cannot choose his in a well-defined way. There is no deductively rational solution—no "correct" expectational model. From the agents' viewpoint, the problem is ill-defined and they are propelled into a world of induction. Second, and diabolically, any commonalty of expectations gets broken up: If all believe few will go, all will go. But this would invalidate that belief. Similarly, if all believe most will go, nobody will go, invalidating that belief. Expectations will be forced to differ.
>
> At this stage, I invite the reader to pause and ponder how attendance might behave dynamically over time. Will it converge, and if so to what? Will it become chaotic? How might predictions be arrived at?

This quotation makes clear that in the situation in which the efficiency of the decision of each player depends on everybody else's decisions, the very concept of optimal strategy may become quite problematic and the existence of an analytic solution even more so. This situation is typical of many social systems, and MS is a natural tool to investigate such systems. The idea of inductive reasoning has been used in the MS of financial market strategies by Palmer, Holland, LeBaron, and Taylor (1994) and Caldarelli, Marsili, and Zhang (1998). More specifically, one may assume a population of traders, each of which adopts various rules for predicting the future trend of the market based on past data. Each agent selects his or her strategy according to the past performance: if a strategy fails, the agent adopts another one. The market dynamics will present trends similar with the ones obtained in the "bar model." Rather than trying to solve this model analytically, people have used MS to investigate the macroscopic properties of such markets. One of the successes attributed to these models is the fact that they lead to nonequilibrium, ever-fluctuating systems, in which the relative performance of various strategies changes with time.

We will see in the next chapters that in order to obtain these nonequilibrium effects one does not necessarily require adaptive/intelligent agents with changing strategies. Indeed, assume each agent has a fixed strategy. If at some stage a strategy is a winner, the carriers of this strategy will get richer. After enough trading periods, they will posses enough wealth to influence the market price significantly by their bids. At that stage, the strategy will cease to win and another strategy will become preferable. The carriers of the new strategy will start gaining at the expense of the carriers of the old strategy. Consequently, the wealth of the new group will increase to the level of influencing the market price. Following this, they will start losing, and so on. Therefore, instead of a strategy selection mechanism based on agents discarding losing strategies, the selection takes place automatically at the market level by the impoverishment of the losing agents (Levy, Persky and Solomon, 1996).

It is likely that in real markets both selection mechanisms take place: at the strategy level and at the agent level. Unfortunately, most of the literature on these minority games and market simulations has been published in physics journals. One of the objectives of this book is to make the relevant MS techniques known and available to researchers in the fields of economics and finance.

6.6. DYNAMICS OF EXPECTATIONS: THE FORMATION OF COALITIONS

One of the most difficult things for western intellectuals to understand is the resilience of dictatorships in different societies in the past and in other contemporary societies. It would seem evident that a leader or a regime cannot stay in power unless a significant part of the population supports that leader or regime. Yet reality seems to systematically deny this belief. To study this problem quantitatively, one has to consider the interaction of many agents, especially their expectations about the behavior of the other agents (Huberman, 1997). Since inductive reasoning is involved (people are trying to guess what others will do in various situations, just like in the bar problem), MS is a main tool in the investigation of such systems. It has been shown that models incorporating such mechanisms lead to the preservation of nonoptimal regimes for long times. When it comes, the transition to the optimal regime is dramatic and discontinuous due to the positive feedback loop between people's behavior and people's expectations. These ideas were used to describe the "outbreak of cooperation" (Glance, 1993).

In political science, the application of MS techniques to international relations and political coalition formation was done by Galam (1990, 1996, 1997, 1998). More specifically, Galam applied it to pre−World War II Europe and to the Cold War between the NATO and the Warsaw Pact.

Galam (1990) has also proposed a simple MS model to show how a minority can stay in power within a democratic hierarchical system that operates through majority rule voting. Each agent in the model is assumed to have an opinion. Cells of four agents are formed randomly to elect a representative, according to the majority within the cell itself. These elected agents constitute the first hierarchical level of the organization. The cell forming process is then iterated at higher levels with committees formed from the agents elected at the previous hierarchical level. In order to enforce a change, a clear majority within the committee is required. In case of a split vote, things will just stay the way they are. The MS shows that this procedure makes it rather hard for change to emerge: even a large majority throughout the population ($> 70\%$) is typically not enough. Galam's model exemplifies, therefore, how democratic voting can lead to a totalitarian regime.

6.7. MICROSCOPIC SIMULATION OF THE NEOLITHIC REVOLUTION

To exemplify the MS formalization of "softer" subjects, let us recount the story of the introduction of agriculture in Europe (*New Scientist*, 1997). Radiocarbon dating of artifacts associated with farming life shows that farming spread from Anatolia (now Turkey) to northern Europe in less than 2000 years (from 8 to 6 thousands years ago). This was termed as the "Neolithic Revolution" and it was associated, among other things, with the spread of the proto-indo-european language in Europe (Renfrew, 1990). It is still under fierce debate if this was accompanied by significant population migration and consequently by genetic stock influx. Among the theories proposed by the various groups (see, for instance, Cavalli-Sforza, Menozzi, and Piazza, 1996) in order to explain the spread of the Neolithic Revolution throughout Europe were the following:

- Learning agriculture from neighbors (and transmitting it to other neighbors)
- Sons/daughters establishing farms next to parents farms
- Farmers moving inside unsowed territory, etc.

Most of the direct data relevant to these individual events is lost. However, the main feature that was established is that the archeological findings are incompatible with a simple diffusion mechanism. Indeed, simple diffusion would imply an expansion of the farming territory proportional to the square root of time, and a fuzzy boundary separating the farming territory from the unsowed territory. In reality, the speed of expansion was constant in time and it advanced along relatively sharp (though irregular) boundaries. Until now, the modeling of this problem involved mainly "continuum" approaches that did not lead to universally accepted conclusions. The crucial role of the individual agents (farmers, hunters) and events (learning farming, establishing farms, moving from one location to another) suggests the use of MS for deciding this debate in a definitive way.

6.8. MICROSCOPIC SIMULATION IN MARKETING

The tulip mania is one of the most celebrated and dramatic bubbles in history. It involved the rise of tulip bulb prices in 1637 to the level of average house prices. In the same year, after an increase by a factor of 20 within a month, the market collapsed back within the next 3 months. After loosing a fortune in a similar event (triggered by the South Sea Company) in 1720, Sir Isaac Newton was quoted to say, "I can calculate the motions of the heavenly bodies, but not the madness of people." It might seem overambitious to try where Newton has failed, but let us not forget that we live 300 years later, have big computers, and have had plenty of additional opportunities to contemplate the madness of people.

We present in this section the use of MS in marketing models in order to reproduce collective economic phenomena such as the Tulip- or Tamagotchi-manias. We will show that global "macroscopic" (and often "catastrophic") economic phenomena are generated by reasonably simple buy-and-sell "microscopic" operations. More specifically, we focus on the space and time dynamics of sales. Examples of intriguing macroscopic phenomena displayed by the process of product marketing are as follows:

- *Sudden death*. The total and almost immediate disappearance of a product (due to the introduction of a new-generation product or to the lack of demand)
- *Wave-fronts*. The geographical spreading of new products through the markets by spatially and temporally localized waves (Bettelheim and Lehmann, 1999)
- The larger profit/survival rate of the second producer of a new product after the failure (or limited success) of the first producer to introduce it (see Goldenberg, Mazursky and Solomon, 1999)

To formulate MS marketing models, it is natural to consider lattices, or square meshes, that may represent in the present context real "geographic" space or a network of interacting markets or individuals. The microscopic elements can represent the product itself, the instruments to buy it, the knowledge of the existence of the product, and so on. The interactions can represent economic processes in which valuables are transformed, lost, or created. Diffusion stands for the transfer of these valuables from one location or business entity to another.

6.8.1. The "Tamagotchi" Model

The two quotes that follow tell best the meteoric story of the Tamagotchi-craze:

> Tokyo, Jan. 24—By dawn today, a line of almost 2,000 people stretched a quarter of a mile through the Ginza shopping district. Hundreds of them had spent the night camped out on the sidewalk in the numbing midwinter cold. When you want a toy chicken badly enough, you will endure anything. (Sullivan, *The Washington Post*, 1997)
>
> Tokyo—Bandai Co. Ltd. said Thursday that it will post a 16 billion yen special loss for 1998/99 due to operations restructuring. The company also revised its parent net forecast to a loss of 14.5 billion yen (about US$121.4 million) from an expected profit of 3.5 billion yen (about $29 million). This is partly due to losses caused by excessive inventory of unsold Tamagotchi, which can now be found for $10 in US toy stores. So much for the US virtual-pet craze, eh? (Ohbuchi, 1999)

In 1997 a very abrupt wave of Tamagotchi sales swept across most of the world and died within months. "Tamagotchi" (egg + watch in Japanese) is a small computer in the shape of a small plastic box with a screen representing schematically a "pet" animal (dinosaur, cat, dog, etc). At particular time intervals, the animal has to be virtually "fed," "cleaned," "put to sleep," and so on, by using the appropriate buttons on the plastic box (otherwise the "pet" shows signs of distress and eventually "dies"). Millions of children and adults around the world were touched at one time by the "Tamagotchi mania." Various countries were reached by the wave at slightly different times, but the overall duration of the sales at each location did not exceed a few months.

The Tamagotchi story is a very clean laboratory for marketing models as it involves very specific characteristics at the microscopic level and very peculiar macroscopic consequences. The MS method has the opportunity to prove its strength by obtaining the macroscopic phenomena from the microscopic properties. The *microscopic* properties at the product level are as follows:

- Nobody really needs a Tamagotchi.
- People want a Tamagotchi only upon actually seeing one.
- The price is relatively small.
- Usually people are satisfied with only one Tamagotchi.
- People get bored with their Tamagotchi after a while.

- The cost of transporting a Tamagotchis between production and selling locations is small.
- The product is always with the individual owning it.

The observed emergent *macroscopic* space-time behavior of the Tamagotchi market included dramatic features such as these:

- Its sudden unexpected spread
- Its not-less-sudden disappearance.

Moreover, the observed *macroscopic* behavior is at variance with the usual market theory that views the spread of new products as a simple diffusion process. Indeed a diffusion process would lead to the following:

- Diffuse (i.e., smooth and continuous) boundaries between the regions where the new product dominates and the virgin regions
- The expansion of the product territory with the square root of time
- A long-lasting dominance of the product in the already dominated area

In reality the situation is very different:

- The spread of the product is defined by sharp boundaries.
- The distance the front advances is proportional to time.
- The initial selling rate is diminished greatly inside the territory already traversed by the wave.

The objective of the MS model is to obtain the *macroscopic* features as collective properties of a multitude of individual interactions that reflect the previously mentioned microscopic properties of the Tamagotchi product.

Consequently, the elementary microscopic objects to be included in the MS are as follows:

- A's, which represent sold product units (Tamagotchi)
- B's, representing the knowledge or conceptual awareness of the product
- C's, representing the availability of money to be allocated for the purpose of buying A's

The elementary objects A, B, and C, are placed on a square mesh whose sites represent spatial locations/customers. The elementary reactions governing their dynamics are as follows:

- If a B and a C are placed on the same site (i.e., there exists on the same site both the awareness of Tamagotchi and the money available to buy it), then there is a certain probability rate for generating an A and a (second) B on that site. In this case, the C will disappear. The interpretation is that a product A is bought at that site and the awareness B is increased. The money C is consumed by the transaction.

- B can spread from one site to a neighboring one. This corresponds to learning of the Tamagotchi by hearing of it from a neighbor (people in the neighborhood of a Tamagochi can see it, hear reports from its owner, be involved in discussions about its health, etc.).
- The disappearance of A's with a certain probability rate. It represents the loss of a Tamagotchi due to wear and tear or by just being discarded by its owner.
- The disappearance of B's (with certain probability rate). This represents the fact that the Tamagotchis can eventually be forgotten.
- The spontaneous appearance of C's on any site (with a certain probability rate). This represents the continuous influx of money due to salaries, allowances, and other incomes.
- The disappearance of C with a certain probability rate. It represents the spending of money on products other than the Tamagotchi or just financial losses.

6.8.2. Tamagotchi Selling Waves

During the MS runs of the above Tamagotchi model, one indeed observes macroscopic long-lasting wave fronts of high A concentration moving across the simulation space. Figure 6.1 displays snapshots of the distribution of A, B, and C taken at the same instant. One sees indeed that the front of Tamagotchi sales (the black front in Fig. 6.1.a) is preceded and followed by a wider region of high Tamagotchi awareness (the black strip with similar shape in Fig. 6.1.b). As the front of sales advances, it leaves behind a region in which the resources allocated for Tamagotchi purchases are depleted (the white region in Fig. 6.1.c).

The dynamical explanation behind this effect is the auto-catalytic character (Shnerb, Bettelheim, Louzon, and Solomon, 1999) of the Tamagotchi sales. More precisely, there exists a positive feedback loop:

Sales of Tamaguchi → Tamagotchi Visibility → Tamagotchi Awareness
→ Tamagotchi Demand → Tamagotchi Sales,
which leads to self-organizing waves.

Note that different products with different microscopic consumer characteristics will present a rather different marketing pattern. For instance, breakfast cereals that are consumed daily and whose demand depends on being hungry rather than on seeing them with a neighbor will display much milder fluctuations.

The important qualitative effect is that the dynamics of the waves are dominated by the first individual sales penetrating a new region, not by a

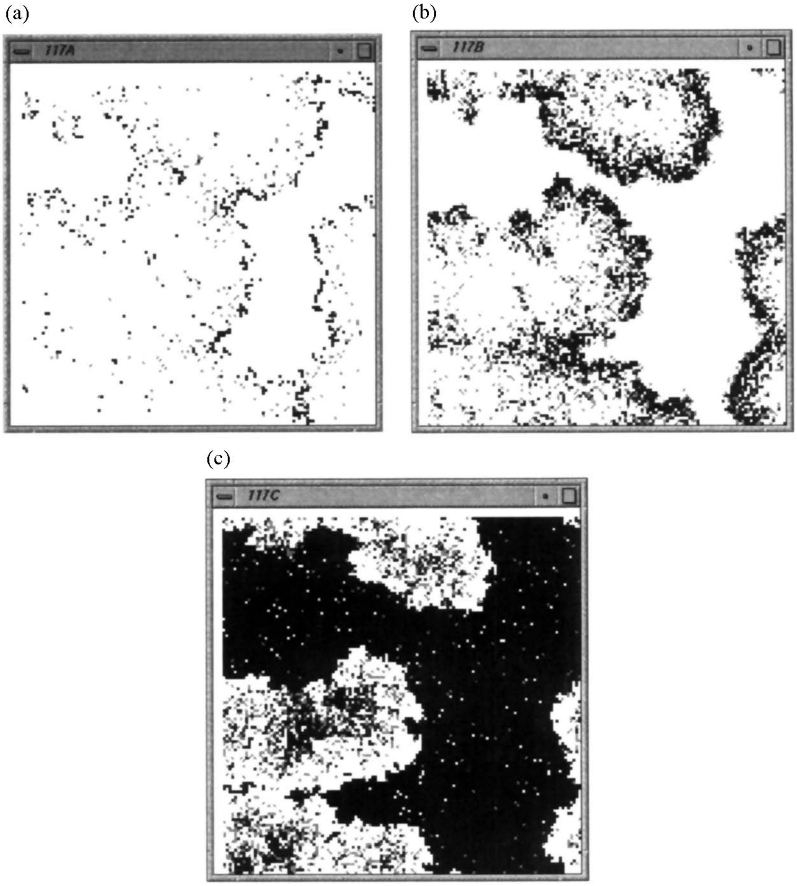


FIGURE 6.1 Contemporary snapshots of the distributions of A, B, and C (respectively) in the "Tamagotchi" model. (From Bettelheim and Lehmann, Copyright 2000 World Scientific Publishing Co. Pte Ltd., with permission).

diffusion of a macroscopic number of B's (awareness of the product) into the region. This emphasizes the importance of MS versus usual market research methods. Indeed, the statistical marketing methods (polls) are sensitive to macroscopic amounts of sales and awareness (B), while, in fact, the real dynamics is triggered at the level of the first individual sales and the first individuals in the area aware of the product.

The fact that the product propagation takes place in waves rather than by normal diffusion is extremely important for the planning of marketing campaigns and the planning of production and investment policies. As opposed to the dissemination by diffusion that is proportional to the square root of time, the wave propagation is linear in time. This affects dramatically the time scales at which one expects the emergence and the decline of a product in a certain market.

Anticipating the demand surge, and the demand collapse, by MS may assist the production planners and the managers in avoiding overstocking or in exploiting in time the imminent emergence of market demands. MS might uncover the potential of these events in situations in which the statistical polls would detect nothing. By the time the polls detect change in demand, it might be too late (as it was the case of Tamagotchi, see Goldenberg *et al.*, 1999).

MS may help decide the optimal place and time for starting a marketing campaign, and whether to be the first to start the marketing of a new product or to wait for the competition to invest in the creation of market awareness (B's). The second producer often has a market advantage over the company that introduces a novel product in the first place (Goldenberg *et al.*, 1999). One possible strategy to make this model useful for practitioners is to study the space-time evolution of past sales in order to establish the connectivity of the various market locations and their capacity in terms of C distribution. This would represent a quantitative microscopic encoding of the market conditions, receptivity and capacity. Once the correct microscopic "map" of the market is obtained from past data, it can be used for new products. One can then estimate the effects of introducing a new product, introducing a cheaper/upgraded version of an existing product, or experiencing an economic/political crisis in certain regions.

Most of the above marketing effects have their counterparts in the behavior of stock prices (booms, crashes, herding, clustered volatility, short time price momentum, long time price reversal, etc.).

A generic model of hits and flops based on the statistical mechanics concept of "percolation" was proposed in Solomon *et al.* (1999). It was found that even in conditions in which a significant fraction of the individuals are potential buyers, the product does not reach its potential market unless certain relations between the social connectivity within the buyers population and its relative density are fulfilled. The model predicts a dramatic jump in the sales once the system reaches the threshold where these relations are fulfilled.

6.9. MICROSCOPIC SIMULATION IN BIOLOGY

Biological simulations seem much more widespread than sociological ones. Suitable journals have existed for half a century, such as the *Journal of Theoretical Biology*. But the field lacks an obvious success, such as Kepler's laws of planetary motion from around 1610 or Einstein's relativity from the beginning of the 20th century. Not all experimental biologists take these theoretical attempts seriously, whether they use microscopic simulations or analytic solutions.

One problem is, for example, why and how we age (La Recherche, 1999). An old hypothesis particularly suited to microscopic simulation is the accumulation of random inherited mutations over many generations. Typically, the negative effects of these mutations become noticable mostly at old age, after one has already procreated, because mutations that are killing the carrier before reproduction are not passed on. Most of the simulations along these lines used the Penna model (1995) as summarized in the book of Moss de Oliveira *et al*. (1999). The book also summarizes simulations on the question of why sexual reproduction is chosen by so many lifeforms in spite of the apparent waste of half of the population never bearing the offspring.

Another aspect of biology simulated with methods close to physics is immunology. An early application of microscopic techniques to immunology is shown in the paper by Kaufman *et al*. (1995), which represented the immune system dynamics in terms of cellular automata and compared it with the results from differential equations. The genetic signature of the antigens and antibodies is represented by strings of 0's and 1's, which mutate and reproduce according to microscopic rules in Castiglione (1999) and Zorzenon (1999). A comprehensive microscopic simulation model of the emergence of the macroscopic immune functions from the microscopic interactions of cells and molecules was developed by Louzon, Solomon, Atlan and Cohen (1999).

7

THE LLS MICROSCOPIC SIMULATION MODEL

7.1. INTRODUCTION

From the discussion in Chapters 2 through 4, we can safely conclude that actual investors have a different investment behavior than that of the idealized rational investor assumed in most economic and financial models. Investors differ one from the other in their preferences, their investment horizon, the information at their disposal, and their interpretation of this information. No financial economist seriously doubts these observations. However, modeling the empirically and experimentally documented investor behavior and the heterogeneity of investors is very difficult and in most cases practically impossible to do within an analytic framework. For instance, the empirical and experimental evidence suggests that most investors are characterized by constant relative risk aversion (CRRA), which implies a power (myopic) utility function (see Chapter 3). However, for a general distribution of returns it is impossible to obtain an analytic solution for the portfolio optimization problem of investors with these preferences[1]. Extrapolation of future returns from past returns, biased probability weighting, and partial deviations from rationality are also all experimentally documented but difficult to incorporate in an analytical setting. One is then usually forced to make the assumptions of rationality and homogeneity (at least in some dimension) and to make unrealistic assumptions regarding investors' preferences in order to obtain a model with a tractable solution. The hope in these circumstances is that the model will capture the essence of the system under investigation and will serve as a useful benchmark, even though some of the underlying assumptions are admittedly false.

Most homogeneous rational agent models lead to the following predictions: no trading volume, zero autocorrelation of returns, and price volatility, which is equal to or lower than the volatility of the "fundamental value" of the stock (defined as the present value of all future dividends; see Shiller, 1981). However, the empirical evidence is very different:

- Trading volume can be extremely heavy (Karpoff, 1987; Admati and Pfleiderer, 1988).
- Stock returns exhibit short-run momentum (positive autocorrelation) and long-run mean reversion (negative autocorrelation) (Fama and French, 1988; Poterba and Summers, 1988; Jegadeesh and Titman, 1993; Levy and Lim, 1998).
- Stock returns are excessively volatile relative to the dividends (Shiller, 1981).

Since most standard rational-representative-agent models cannot explain these empirical findings; these phenomena are known as "anomalies" or "puzzles." Are these anomalies due to elements of investors' behavior that are unmodeled in the standard rational-representative-agent models, such as the experimentally documented deviations of investors' behavior from rationality and/or to the heterogeneity of investors? Until recently it has been very difficult to investigate stock market models incorporating

[1] Even if there is a single risky stock that has a discrete return distribution with only four possible outcomes, for CRRA preferences (power utility) it is impossible to find the optimal portfolio diversification between the stock and the riskless asset analytically. One of the single situations in which the optimization can be solved analytically is the case of a negative exponential utility function and a normal return distribution. This is why many models assume exponential utility functions even though they imply CARA, which implies that a person does not change his or her dollar investment in the stock, even if the person wins a million dollars in a lottery.

these elements of investor behavior, and thus this question is yet to be answered[2].

Microscopic simulation (MS) allows the investigator to study realistic models that are based on the empirical and experimental findings regarding investors' behavior, and incorporate the many ways in which investors differ from one another. The strength of the MS method is that since it is not restricted to the scope of analytical methods, one is able to investigate virtually any imaginable investor behavior and market structure. Thus, one can study models that incorporate the experimental findings regarding the behavior of investors and evaluate the effects of various behavioral elements on market dynamics and asset pricing.

In this chapter we present an MS model of the stock market. This model is based on the principles employed previously by Levy, Levy, and Solomon (LLS) (1994, 1995), Solomon (1995), and Levy and Levy (1996). The LLS model incorporates some of the main empirical findings regarding investor behavior, and we employ this model to study the effect of each element of investor behavior on asset pricing and market dynamics. We start out with a benchmark model in which all of the investors are rational, informed, and identical, and then, one by one, we add elements of heterogeneity and deviations from rationality to the model in order to study their effects on the market dynamics.

In the benchmark model all investors are rational, informed, and identical (RII investors). The RII investors are informed about the dividend process, and they rationally act to maximize their expected utility. The RII investors make investment decisions based on the present value of future cash flows. They are essentially fundamentalists who evaluate the stock's fundamental value and try to find bargains in the market. The benchmark model in which all investors are RII yields results that are typical of most rational-representative-agent models: in this model prices follow a random walk, there is no excess volatility of the prices relative to the volatility of the dividend process, and since all agents are identical, there is no trading volume.

After describing the properties of the benchmark model, we investigate the effects of introducing various elements of investor behavior that

[2] Several studies have suggested that "behavioral" effects may account for these market anomalies. De Bondt and Thaler (1985) model the tendency of investors to overreact to news and show that this overreaction can explain the mean-reversion of returns. Shefrin and Statman (1985) show that investors' loss aversion, regret aversion, and mental accounting can explain return autocorrelations. Barberis, Shleifer, and Vishny (1998) present a model that captures the psychological phenomenon of conservatism (slow updating) and the representativeness heuristic. Daniel, Hirshleifer, and Subrahmanyan (1999) show that some anomalies can be explained in terms of investors' overconfidence and biased self-attribution (which means that investors update their self-confidence in a biased way).

are found in laboratory experiments but are absent in most standard models. We do so by adding to the model a minority of investors who do not operate like the RII investors. These investors are efficient market believers (EMB investors). The EMBs are investors who believe that the price of the stock reflects all of the currently available information about the stock. As a consequence, they do not try to time the market or to buy bargain stocks. Rather, their investment decision is reduced to the optimal diversification problem. For this portfolio optimization, the *ex ante* return distribution is required. However, since the *ex ante* distribution is unknown, the EMB investors use the *ex post* distribution to estimate the *ex ante* distribution. It has been documented that, in fact, many investors form their expectations regarding the future return distribution based on past performance (see Chapter 4). This behavior is rational if the process generating the returns is fairly stable over time and a long enough *ex post* period is taken in order to obtain a reliable estimate of the future distribution. However, it is well known that the human tendency to look for continuity and to extrapolate from the past into the future persists even when it is not necessarily rational or justified to do so (see, for example, Gilovich, Vallone, and Tversky, 1985). Research has shown that investors look for clues in the historical returns, even in circumstances when they *know* that the *ex post* returns are irrelevant. For example, in an experimental study that involved an investment in risky and riskless assets, Kroll, Levy, and Rapoport (1988a) told the subjects (investors) that rates of return are drawn randomly from normal distributions with parameters that where given to the subjects. This implies that historical prices and returns are not useful for future decision making. Yet each of the subjects required information regarding the historical prices at least once (see p. 400). The authors conclude the following:

> Subjective beliefs (which probably vary from one investor to another) seem to be based on the assumption that past stock price changes are relevant for future decisions. Models based on these beliefs presumably reflect common, though not necessarily accurate, knowledge about the behavior of risky assets in actual rather than experimental markets. The terms "bear market" and "bull market" have become part of our language. When interrogated informally, most of our subjects expressed their belief that 'stocks follow trends.'

Thus, in this experimental study historical rates of return were used to form beliefs regarding future rates of return.

There are various ways to incorporate the investment decisions of the EMBs. This stems from the fact that there are different ways to estimate the *ex ante* distribution from the *ex post* distribution. How far back should one look at the historical returns? Should more emphasis be given to more recent returns? Should some "outlier" observations be filtered out? Of course, there are no clear answers to these questions, and different investors may have different ways of forming their estimation of the *ex* *ante* return distribution (even though they are looking at the same series of historical returns). Moreover, some investors may use the objective *ex post* probabilities when constructing their estimation of the *ex ante* distribution, whereas others may use biased subjective probability weights.[3] To build the analysis step by step, we start by analyzing the case in which the EMB population is homogeneous, and then introduce various forms of heterogeneity into this population.

An important issue in market modeling is that of the degree of investors' rationality. Most models in economics and finance assume that people are fully rational. This assumption usually manifests itself as the maximization of an expected utility function by the individual. However, numerous experimental studies have shown that people deviate from rational decision making.[4] Some studies model deviations from the behavior of the rational agent by introducing a subgroup of liquidity, or "noise", traders. These are traders that buy and sell stocks for reasons that are not directly related to the future payoffs of the financial asset—their motivation to trade arises from outside of the market (for example, a noise trader's daughter unexpectedly announces her plans to marry, and the trader sells stocks because of this unexpected need for cash). The exogenous reasons for trading are assumed random, and thus lead to random or noise trading (see Glosten and Milgrom, 1985; Grossman and Stiglitz, 1980; Hellwig, 1980; and Kyle, 1985a). The LLS model takes a different approach to the modeling of noise trading. Rather than dividing investors into the extreme categories of "fully rational" and "noise traders," the LLS model assumes that most investors try to act as rationally as they can, but are influenced by a multitude of factors causing them to deviate to some extent from the behavior that would have been optimal from their point of view. Namely, all investors are characterized by a utility function and act to maximize their expected utility; however, some investors may deviate to some extent from the optimal choice that maximizes their expected utility. These deviations from the optimal choice may be due to irrationality, inefficiency, liquidity constraints, or a combination of all of the above.

In the framework of the LLS model, we examine the effects of the EMBs' deviations from rationality and their heterogeneity, relative to the

[3] For a description of the empirical findings regarding the use of subjective versus objective probabilities, see Edwards (1953), Edwards (1954), Kahneman and Tversky (1979), and Tversky and Kahneman (1992) and the review in Chapter 2.

[4] A famous example of people's irrationality is the lost ticket result: most people would buy a $10 ticket for a play even if they discover that they lost a $10 bill on the way to the theater. However, if they had the $10 ticket and lost it on the way to the theater, most people will not buy a new ticket. For a description of this and other empirical findings regarding the irrationality of people, see Tversky and Kahneman (1981). For a review of irrationality in the context of financial investments, see Chapters 2 and 4.

benchmark model in which all investors are informed, rational, and homogeneous. We find that the behavioral elements that are empirically documented—namely, extrapolation from past returns, deviation from rationality, and heterogeneity among investors—lead to all of the following empirically documented "puzzles":

- Excess volatility
- Short-term momentum
- Longer term return mean reversion
- Heavy trading volume
- Positive correlation between volume and contemporaneous absolute returns
- Positive correlation between volume and lagged absolute returns

The fact that all these anomalies or "puzzles," which are hard to explain with standard rational-representative-agent models, are generated naturally by a simple model which incorporates the experimental findings regarding investor behavior and the heterogeneity of investors, leads one to suspect that these behavioral elements and the diversity of investors are a crucial part of the workings of the market, and as such they cannot be "assumed away." As the experimentally documented bounded-rational behavior and heterogeneity are in many cases impossible to analyze analytically, MS presents a very promising tool for investigating market models incorporating these elements.

This chapter is organized as follows: in the next section we introduce the LLS model. Section 7.3 presents the results of the benchmark model in which all investors are rational, informed, and identical. In Section 7.4 we introduce a small proportion of EMB investors into the model. We first study the case in which the EMB population is homogeneous, and then we investigate the effects of heterogeneity of the EMB population. Section 7.5 discusses the survivability of the EMB population. A summary of this chapter is given in section 7.6.

7.2. THE MODEL

The stock market consists of two investment alternatives: a stock (or index of stocks) and a bond. The bond is assumed to be a riskless asset, and the stock is a risky asset. The stock serves as a proxy for the market portfolio (e.g., the Standard & Poors 500 index). The extension from one risky asset to many risky assets is possible; however, one stock (the index) is sufficient for our present analysis because we restrict ourselves to global market phenomena and do not wish to deal with asset allocation across several risky assets. Investors are allowed to revise their portfolio at given time points (i.e. we discuss a discrete time model).

The bond is assumed to be a riskless investment yielding a constant return at the end of each time period. The bond is in infinite supply and investors can buy from it as much as they wish at a given rate of r_f. The stock is in finite supply. There are N outstanding shares of the stock. The return on the stock is composed of two elements:

1. *Capital gain*. If an investor holds a stock, any rise (fall) in the price of the stock contributes to an increase (decrease) in the investor's wealth.
2. *Dividends*. The company earns income and distributes dividends at the end of each time period. We denote the dividend per share paid at time t by D_t. We assume that the dividend is a stochastic variable following a multiplicative random walk—that is, $\tilde{D}_t = D_{t-1}(1 + \tilde{z})$, where $\tilde{z}$ is a random variable with some probability density function $f(z)$ in the range $[z_1, z_2]$. (In order to allow for a dividend cut as well as a dividend increase, we typically choose: $z_1 < 0$, $z_2 > 0$).

The total return on the stock in period t, which we denote by R_t is given by

$$\tilde{R}_t = \frac{\tilde{P}_t + \tilde{D}_t}{P_{t-1}} \tag{7.1}$$

where $\tilde{P}_t$ is the stock price at time t.

All investors in the model are characterized by a von Neuman-Morgenstern utility function. We assume that all investors have a power utility function of the form

$$U(W) = \frac{W^{1-\alpha}}{1-\alpha} \tag{7.2}$$

where α is the risk aversion parameter.[5] This form of utility function implies constant relative risk aversion (CRRA) (which is a special case of decreasing absolute risk aversion, DARA[6]). We employ the power utility function (Eq. 7.2) because the empirical evidence suggests that relative risk aversion is approximately constant (see, for example, Friend and Blume, 1975; Gordon, Paradis, and Rorke, 1972; Kroll, Levy, and Rapoport, 1988a; Levy, 1994, and the discussion in Chapter 3), and the power utility function is the unique utility function that satisfies the CRRA condition (see Chapter 3). Another implication of CRRA is that the optimal investment choice is independent of the investment horizon (Samuelson, 1989, 1994). In other words, regardless of investors' actual investment horizon,

[5] The model can be analyzed with homogeneous or heterogeneous risk aversion. Since heterogeneous risk aversion does not seem to play a key role (see Levy, Levy, and Solomon, 1995), for the sake of simplicity, we assume here homogeneous risk aversion.

[6] For a formal definition of DARA and CRRA, see Pratt (1964) and Arrow (1965).

they choose their optimal portfolio as though they are investing for a single period. This property is also known as myopia, or "short vision," and hence the power utility function given in Eq. (7.2) is also known as the myopic utility function. The myopia property of the power utility function simplifies our analysis, as it allows us to assume that investors maximize their one-period-ahead expected utility.[7]

We model two different types of investors: rational, informed, identical (RII) investors and efficient market believers (EMB) investors. These two investor types are described next.

7.2.1 Rational Informed Identical (RII) Investors

RII investors evaluate the "fundamental value" of the stock as the discounted stream of all future dividends. They believe that the stock price may deviate from the fundamental value in the short run, but if it does, it will eventually converge to the fundamental value. This is similar to the convergence to an "asymptotic 'normal' price-level" modeled by Merton (1971). The RII investors act according to the assumption of asymptotic convergence: if the stock price is low relative to the fundamental value, they buy in anticipation that the underpricing will be corrected, and vice versa. We make the simplifying assumption that the RII investors believe that the convergence of the price to the fundamental value will occur in the next period; however, our results hold for the more general case where the convergence is assumed to occur some T periods ahead, with $T > 1$.

To estimate next period's return distribution, the RII investors need to estimate the distribution of next period's price, $\tilde{P}_{t+1}$, and of next period's dividend, $\tilde{D}_{t+1}$. Since they know the dividend process, the RII investors know that $\tilde{D}_{t+1} = D_t(1 + \tilde{z})$, where $\tilde{z}$ is distributed according to $f(z)$ in the range $[z_1, z_2]$. The RII investors employ Gordon's dividend stream model[8] to calculate the fundamental value of the stock:

$$P_{t+1}^{f} = \frac{E_{t+1}\left[\tilde{D}_{t+2}\right]}{k - g} \tag{7.3}$$

where the superscript f stands for the *fundamental* value, $E_{t+1}[\tilde{D}_{t+2}]$ is the dividend corresponding to time $t + 2$ as expected at time $t + 1$, k is the discount factor or the expected rate of return demanded by the market

[7] For utility functions other than the power function, the investment horizon of the investor *does* influence the portfolio choice, and dynamic programming issues must be considered.

[8] The crucial ingredient here is that the fundamentalists employ a valuation model based on future cash flows. The Gordon model is chosen for simplicity; other choices do not change the results.

for the stock,[9] and g is the *expected* growth rate of the dividend—that is, $g = E(\tilde{z}) = \int_{z_1}^{z_2} f(z)z\,dz$.

The RII investors believe that the stock price may temporarily deviate from the fundamental value; however, they also believe that the price will eventually converge to the fundamental value. For simplification we assume that the RII investors believe that the convergence to the fundamental value will take place next period. Thus, the RII investors estimate P_{t+1} as follows:

$$P_{t+1} = P_{t+1}^f$$

The expectation at time $t + 1$ of $\tilde{D}_{t+2}$ depends on the realized dividend observed at $t + 1$:

$$E_{t+1}\left[\tilde{D}_{t+2}\right] = D_{t+1}(1 + g)$$

Thus, the RII investors believe that the price at $t + 1$ will be given by

$$P_{t+1} = P_{t+1}^f = \frac{D_{t+1}(1 + g)}{k - g}$$

At time t, D_t is known, but D_{t+1} is not; therefore P_{t+1}^f is also not known with certainty at time t. However, given D_t, the RII investors know the distribution of $\tilde{D}_{t+1}$:

$$\tilde{D}_{t+1} = D_t(1 + \tilde{z})$$

where $\tilde{z}$ is distributed according to the known $f(z)$. The realization of $\tilde{D}_{t+1}$ determines P_{t+1}^f. Thus, at time t, RII investors believe that P_{t+1} is a random variable given by

$$\tilde{P}_{t+1} = \tilde{P}_{t+1}^f = \frac{D_t(1 + \tilde{z})(1 + g)}{k - g}$$

Notice that the RII investors face uncertainty regarding next period's price. In our model we assume that the RII investors are certain about the dividend growth rate g, the discount factor k, and the fact that the price will converge to the fundamental value next period. In this framework the only source of uncertainty regarding next period's price stems from the uncertainty regarding next period's dividend realization. More generally, the RII investors' uncertainty can result from uncertainty regarding any

[9] The discount factor k can be estimated in various ways (e.g., the CAPM or factor models). Independent of how k is estimated, in a fundamentalist market, under general parameterizations of the model, k is self-consistent and therefore justified—that is, if investors estimate the appropriate discount factor as k, the model will generate an average rate of return, which is close to this value (see Appendix 7.1).

one of the previous factors or a combination of several of these factors. Any mix of these uncertainties is possible to investigate in the MS framework but very hard, if not impossible, to incorporate in an analytic framework. As a consequence of the uncertainty regarding next period's price and of their risk aversion, the RII investors do not buy an infinite number of shares even if they perceive the stock as underpriced. Rather, they estimate the stock's next period's return distribution and find the optimal mix of the stock and the bond that maximizes their expected utility. The RII investors estimate next period's return on the stock as follows:

$$\tilde{R}_{t+1} = \frac{\tilde{P}_{t+1} + \tilde{D}_{t+1}}{P_t} = \frac{\dfrac{D_t(1+\tilde{z})(1+g)}{k-g} + D_t(1+\tilde{z})}{P_t} \tag{7.4}$$

where $\tilde{z}$, the next year growth in the dividend, is the source of uncertainty. The demands of the RII investors for the stock depend on the price of the stock. For any *hypothetical* price P_h, investors calculate the proportion of their wealth x they should invest in the stock in order to maximize their expected utility. The RII investor i believes that if she invests a proportion x of her wealth in the stock at time t, then at time $t+1$ her wealth will be

$$\tilde{W}^i_{t+1} = W^i_h\left[(1-x)(1+r_f) + x\tilde{R}_{t+1}\right] \tag{7.5}$$

where $\tilde{R}_{t+1}$ is the return on the stock, as given by Eq (7.4), and W^i_h is the wealth of investor i at time t given that the stock price at time t is P_h.[10]

If the price in period t is the hypothetical price P_h, the $t+1$ expected utility of investor i is the following function of her investment proportion in the stock, x:

$$EU\left(\tilde{W}^i_{t+1}\right) = EU\left(W^i_h\left[(1-x)(1+r_f) + x\tilde{R}_{t+1}\right]\right). \tag{7.6}$$

Substituting $\tilde{R}_{t+1}$ from Eq (7.4), using the power utility function (eq. (7.2)), and substituting the hypothetical price P_h for P_t, the expected utility

[10] W^i_h is given by the wealth in the last period, plus dividends and interest accumulated during the time period $(t-1, t)$, plus capital gains or losses that result from changes in the stock price at the time t trade:

$$W^i_h = W^i_{t-1} + N^i_{t-1}D_t + (W^i_{t-1} - N^i_{t-1}P_{t-1})r_f + N^i_{t-1}(P_h - P_{t-1})$$

where N^i_{t-1} is the number of shares held by investor i at time $t-1$.

becomes the following function of x:

$$EU\left(\tilde{W}_{t+1}^{i}\right) = \frac{(W_h^i)^{1-\alpha}}{1-\alpha} \int_{z_1}^{z_2} \left[(1-x)(1+r_f) + x \left(\frac{\frac{D_t(1+z)(1+g)}{k-g} + D_t(1+z)}{P_h} \right) \right]^{1-\alpha} f(z)\, dz \tag{7.7}$$

where the integration is over all possible values of z. In the MS framework, this expression for the expected utility, and the optimal investment proportion, x, can be solved numerically for any general choice of distribution $f(z)$. For the sake of simplicity we restrict the present analysis to the case where $\tilde{z}$ is distributed uniformly in the range $[z_1, z_2]$. This simplification leads to the following expression for the expected utility:

$$\begin{aligned} EU\left(\tilde{W}_{t+1}^{i}\right) = {} & \frac{(W_h^i)^{1-\alpha}}{(1-\alpha)(2-\alpha)} \frac{1}{(z_2 - z_1)} \left(\frac{k-g}{k+1} \right) \frac{P_h}{xD_t} \\ & \times \left\{ \left[(1-x)(1+r_f) + \frac{x}{P_h} \left(\frac{k+1}{k-g} \right) D_t(1+z_2) \right]^{(2-\alpha)} \right. \\ & \left. - \left[(1-x)(1+r_f) + \frac{x}{P_h} \left(\frac{k+1}{k-g} \right) D_t(1+z_1) \right]^{(2-\alpha)} \right\} \end{aligned} \tag{7.8}$$

(For a derivation of Eq. (7.8) see Appendix 7.2.) For any hypothetical price P_h, each investor (numerically) finds the optimal proportion x_h, which maximizes his or her expected utility given by Eq. (7.8).[11] Notice that the optimal proportion, x_h, is independent of the wealth, W_h^i. Thus, if all RII investors have the same degree of risk aversion, α, they will have the same optimal investment proportion in the stock, regardless of their wealth. The number of shares demanded by investor i at the hypothetical price P_h is

[11] In order to avoid problems of negative wealth, bankruptcy, and negative prices, we impose the constraint of no borrowing and no short-selling (i.e., $0 \le x_h \le 1$). In the more realistic case with limited borrowing and short selling, we would have some other restrictions on x_h of the form $a \le x_h \le b$.

given by the following:

$$N_h^i(P_h) = \frac{x_h^i(P_h)W_h^i(P_h)}{P_h} \tag{7.9}$$

7.2.2 Efficient Market Believers (EMB)

The second type of investors in the LLS model are EMBs. The EMBs believe in market efficiency—they believe that the stock price accurately reflects the stock's fundamental value. Thus, they do not try to time the market or to look for "bargain" stocks. Rather, their investment decision is reduced to the optimal diversification between the stock and the bond. This diversification decision requires the *ex ante* return distribution for the stock, but as the *ex ante* distribution is not available, the EMBs assume that the process generating the returns is fairly stable, and they employ the *ex post* distribution of stock returns in order to estimate the *ex ante* return distribution.

Different EMB investors may disagree on the optimal number of *ex post* return observations that should be employed in order to estimate the *ex ante* return distribution. There is a trade-off between using more observations for better statistical inference and using a smaller number of only more recent observations, which are probably more representative of the *ex ante* distribution. As in reality, there is no recipe for the optimal number of observations to use. EMB investor i believes that the m^i most recent returns on the stock are the best estimate of the *ex ante* distribution. In absence of additional information, *unbiased* investors create an estimation of the *ex ante* return distribution by assigning an equal probability to each of the m^i most recent return observations:

$$\text{Prob}^i\left(\tilde{R}_{t+1} = R_{t-j}\right) = \frac{1}{m^i} \qquad \text{for } j = 1, \ldots, m^i \tag{7.10}$$

In this chapter we deal only with *unbiased* EMB investors. In Chapter 9 we examine the effects of biases investors may have in estimating the *ex ante* distribution. Namely, we introduce the probability distortion effects empirically observed by Edwards (1953), Tversky and Kahneman (1992), and others (see Chapter 2), and we study the impact of these distortions on the market dynamics.

The expected utility of EMB investor i is given by the following:

$$EU(W_{t+1}^i) = \frac{(W_h^i)^{1-\alpha}}{(1-\alpha)} \frac{1}{m^i} \sum_{j=1}^{m^i} \left[(1-x)(1+r_f) + xR_{t-j}\right]^{1-\alpha} \tag{7.11}$$

where the summation is over the set of m^i most recent *ex post* returns, x

is the proportion of wealth invested in the stock, and, as before, W_h^i is the wealth of investor i at time t given that the stock price at time t is P_h (see footnote 10). Notice that W_h^i does not change the optimal diversification policy (i.e., x). Given a set of m^i past returns, the optimal portfolio for the EMB investor i is an investment of a proportion x^{*i} in the stock and $(1 - x^{*i})$ in the bond, where x^{*i} is the proportion that maximizes the previous expected utility (Eq. (7.11)) for investor i. Notice that x^{*i} generally cannot be solved for analytically. However, in the MS framework this does not constitute a problem, as one can find x^{*i} numerically.

7.2.3 Deviations from Rationality

Investors who are efficient market believers, and are rational, choose the proportion of investment x^* that maximizes their expected utility. However, many empirical studies have shown that the behavior of investors is driven not only by rational expected utility maximization but by a multitude of other factors (see, for example, Edgar, 1991; Tversky and Kahneman, 1981). Deviations from the optimal rational investment proportion can be due to the cost of the resources that are required for the portfolio optimization: time, access to information, computational power, and so on, or due to exogenous events (for example, an investor plans to revise his portfolio but gets distracted because his car breaks down). We assume that the different factors causing the investor to deviate from the optimal investment proportion x^* are random and uncorrelated with each other. By the central limit theorem, the aggregate effect of a large number of random uncorrelated influences is a normally distributed random influence, or "noise." Hence, we model the effect of all the factors causing the investor to deviate from his optimal portfolio by adding a normally distributed random variable to the optimal investment proportion. To be more specific, we assume

$$x^i = x^{*i} + \tilde{\varepsilon}^i \tag{7.12}$$

where $\tilde{\varepsilon}^i$ is a random variable drawn from a truncated normal distribution with mean zero and standard deviation σ[12]. Notice that noise is investor specific, thus, $\tilde{\varepsilon}^i$ is drawn separately and independently for each investor.

The noise can be added to the decision making of the RII investors, the EMB investors, or to both. The results are not much different with these various approaches. Since the RII investors are taken as the bench-

[12] To avoid bankruptcies and possible negative prices, the distribution of $\tilde{\varepsilon}^i$ is truncated such that $0 \le x^i \le 1$ (see footnote 11).

mark of rationality, in this chapter we add the noise only to the decision-making practices of the EMB investors.

7.2.4 Market Clearance

The number of shares demanded by each investor is a monotonically decreasing function of the hypothetical price P_h (see Appendix 7.3). As the total number of outstanding shares is N, the price of the stock at time t is given by the market clearance condition: P_t is the unique price at which the total demand for shares is equal to the total supply, N:

$$\sum_i N_h^i(P_t) = \sum_i \frac{x_h(P_t)W_h^i(P_t)}{P_t} = N \tag{7.13}$$

where the summation is over all the investors in the market, RII investors as well as EMB investors.

7.2.5 Dynamics

The market dynamics begin with a set of initial conditions that consist of an initial stock price P_0, an initial dividend D_0, and the wealth and number of shares held by each investor at time $t = 0$. At the first period ($t = 1$), interest is paid on the bond, and the time 1 dividend $\tilde{D}_1 = D_0(1 + \tilde{z})$ is realized and paid out. Then investors submit their demand orders, $N_h^i(P_h)$, and the market clearing price P_1 is determined. After the clearing price is set, the new wealth and number of shares held by each investor are calculated. This completes one time period. This process is repeated over and over as the market dynamics develop.

We would like to stress that even the simplified benchmark model, with only RII investors, is impossible to solve analytically. The reason for this is that the optimal investment proportion, $x_h(P_h)$, cannot be calculated analytically. This problem is very general and it is encountered with almost any choice of utility function and distribution of returns. One important exception is the case of a negative exponential utility function and normally distributed returns. Indeed, many models make these two assumptions for the sake of tractability.[13] The problem with the assumption of negative exponential utility is that it implies constant absolute risk aversion (CARA), which is very unrealistic, as it implies that investors choose to invest the same dollar amount in a risky prospect *independent of their wealth*. This is not only in sharp contradiction to the empirical evidence (see Chapter 3), but it also excludes the investigation of the two-way

[13] See, for example, Grossman and Stiglitz (1980), Kyle (1985a), and Glosten and Milgrom (1985).

interaction between wealth dynamics and price dynamics, which is crucial to the understanding of the market (see Section 7.5).

Thus, one contribution of the MS method is that it allows investigation of models with realistic assumptions regarding investors' preferences. However, the main contribution of this method is that it permits us to investigate models that are much more complex (and realistic) than the benchmark model, in which all investors are RII. As will be demonstrated later on in the chapter, with the MS method one can study models incorporating the empirically and experimentally documented investors' behavior and the heterogeneity of investors.

In the next section we describe the properties of the benchmark model in which all investors are rational, informed, and identical. In latter sections we introduce EMB investors into the market and study their effect on asset pricing and market dynamics.

7.3. RESULTS OF THE BENCHMARK MODEL

In the benchmark model, all investors are RII: rational, informed, and identical. Thus, it is not surprising that the benchmark model generates market dynamics that are typical of homogeneous rational agent models:

7.3.1 No Volume

All investors in the model are identical; they therefore always agree on the optimal proportion to invest in the stock. As a consequence, all the investors always achieve the same return on their portfolio. This means that at any time period the ratio between the wealth of any two investors is equal to the ratio of their initial wealths—that is,

$$\frac{W_t^i}{W_t^j} = \frac{W_0^i}{W_0^j} \tag{7.14}$$

As the wealth of investors is always in the same proportion, and as they always invest the same fraction of their wealth in the stock, the number of shares held by different investors is also always in the same proportion:

$$\frac{N_t^i}{N_t^j} = \frac{\dfrac{x_t W_t^i}{P_t}}{\dfrac{x_t W_t^j}{P_t}} = \frac{W_t^i}{W_t^j} = \frac{W_0^i}{W_0^j} \tag{7.15}$$

Since the total supply of shares is constant, this implies that each investor always holds the same number of shares, and there is *no trading volume* (the number of shares held may vary from one investor to the other as a consequence of different initial endowments).

7.3.2 Log-Prices follow a Random Walk

In the benchmark model, all investors believe that next period's price will converge to the fundamental value given by the discounted dividend model (Eq. (7.3)). Therefore, the actual stock price is always close to the fundamental value. The fluctuations in the stock price are driven by fluctuations in the fundamental value, which in turn are driven by the fluctuating dividend realizations. As the dividend fluctuations are (by assumption) uncorrelated over time, one would expect that the price fluctuations will also be uncorrelated. To verify this intuitive result, we examine the return autocorrelations in simulations of the benchmark model.

Let us turn to the simulation of the model. We first describe the parameters and initial conditions used in the simulation, and then we report the results. We simulate the benchmark model with the following parameters:

- Number of investors = 1000
- Risk aversion parameter $\alpha = 1.5$. This value conforms with the estimate of the risk aversion parameter found empirically and experimentally, as described in Chapter 3.
- Number of shares = 10000

  We take the time period to be a quarter, and accordingly we choose:
- Riskless interest rate $r_f = 0.01$.
- Required rate of return on stock $k = 0.04$
- Maximal one-period dividend decrease $z_1 = -0.07$
- Maximal one-period dividend growth $z_2 = 0.10$

  $\tilde{z}$ is uniformly distributed between these values. Thus, the average dividend growth rate is $g = \frac{z_1 + z_2}{2} = 0.015$.

Initial Conditions: Each investor is endowed at time $t = 0$ with a total wealth of \$1000, which is composed of 10 shares worth an initial price of \$50 per share, and \$500 in cash. The initial quarterly dividend is set at \$0.5 (for an annual dividend yield of about 4%). As will soon become evident, the dynamics are not sensitive to the particular choice of initial conditions.

Figure 7.1 shows the price dynamics in a typical simulation with these parameters (simulations with the same parameters differ one from the

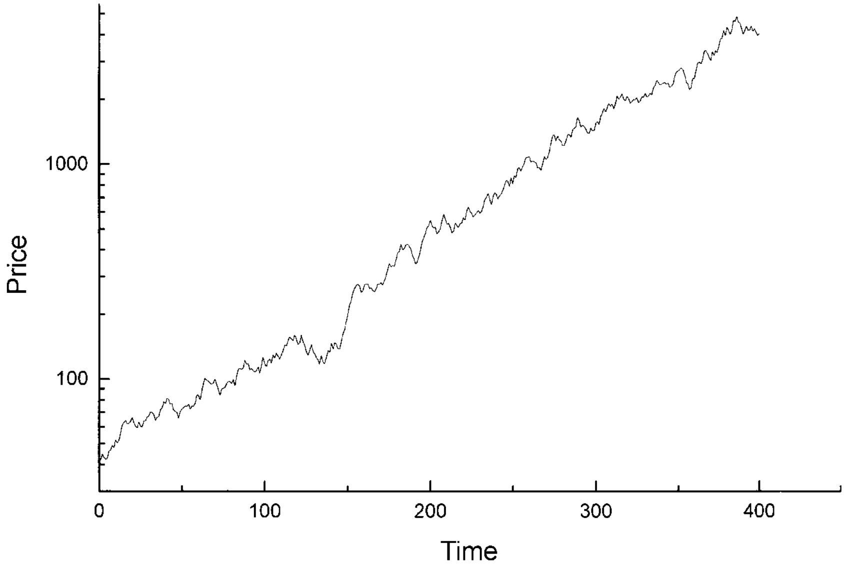


FIGURE 7.1 Price dynamics in the benchmark model.

other because of the different random dividend realizations).[14] Notice that the vertical axis in this figure is logarithmic. Thus, the roughly constant slope implies an approximately exponential price growth, or an approximately constant *average* return. As explained in Appendix 7.1, this average return is consistent with investors' beliefs regarding the appropriate discount factor for the stock (the average rate of return on the stock in this simulation was 0.037, which should be compared with the discount factor which was taken as 0.040).

The prices in this simulation seem to fluctuate randomly around the trend. However, Figure 7.1 shows only one simulation. To have a more rigorous analysis we perform many independent simulations and employ statistical tools. Namely, for each simulation we calculate the autocorrelation of returns. Following Campbell, Lo, and Mackinlay (1997), we perform a univariate regression of the return in time t on the return in time $t - j$:

$$R_t = \alpha_j + \beta_j R_{t-j} + \varepsilon$$

[14] In many markets the price does not change unless there is some trading volume. Here we assume that if the demand at a given price increases (or decreases), the price adjusts such that investors optimize their portfolios and the market clears, even though no shares change hands (because all investors are identical). This serves as a benchmark with which to compare the results of the market with heterogeneous investors, in which there is nonzero volume.

where R_t is the return in period t, and j is the lag. The autocorrelation of returns for lag j is defined as

$$\rho_j = \frac{\text{cov}(R_t, R_{t-j})}{\hat{\sigma}^2(R)}$$

and it is estimated by $\hat{\beta}$. We calculate the autocorrelation for different lags, $j = 1, \ldots 40$. Table 7.1 summarizes the average autocorrelation and t-values over 100 independent simulations. Notice that none of the average t-values is large. We also calculate the number of t-values that are positive and significant or negative and significant out of the total of 100 t-values calculated for each lag. For each lag, we find a small number of significant positive t-values, and a roughly similar small number of significant negative t-values, as would be expected if the autocorrelation is not significant. Figure 7.2 shows the average autocorrelation as a function of the lag. It is evident both from the table and the figure that the returns are uncorrelated in the benchmark model, conforming with the random-walk hypothesis.

7.3.3 No Excess Volatility

Since the RII investors believe that the stock price will converge to the fundamental value next period, in the benchmark model prices are always close to the fundamental value given by the discounted dividend stream. Thus, we do not expect prices to be more volatile than the value of the discounted dividend stream. For a formal test of excess volatility, we follow the technique in Shiller (1981). For each time period we calculate the actual price, P_t, and the fundamental value of discounted dividend stream, P_t^f, as in eq. (7.3). Since prices follow an upward trend, in order to have a meaningful measure of the volatility, we must detrend these price series. Following Shiller, we run the regression:

$$\ln P_t = bt + c + \varepsilon_t \tag{7.16}$$

in order to find the average exponential price growth rate (where b and c are constants). Then, we define the detrended price as: $p_t = \frac{P_t}{e^{\hat{b}t}}$. Similarly, we define the detrended value of the discounted dividend stream p_t^f, and compare $\sigma(p_t)$ with $\sigma(p_t^f)$. For 100 1000-period simulations, we find an average $\sigma(p_t)$ of 22.4, and an average $\sigma(p_t^f)$ of 22.9. As expected, the actual price and the fundamental value have almost the same volatility.

To summarize the results obtained for the benchmark model, we find that when all investors are assumed to be rational, informed, and identical, we obtain results that are typical of rational-representative-agent models: no volume, no return autocorrelations, and no excess volatility. We next

TABLE 7.1 Return Autocorrelations in the Benchmark Model

Autocorrelation of returns in the benchmark model where all investors are RII. We run 100 independent 1000-period-long simulations. For each simulation, we run a univariate regression of R_t on R_{t-j} for different lags $j = 1 \ldots 40$. For each lag, we report the average autocorrelation and average t-value (averaged over all 100 simulations), as well as the number of significant positive and negative t-values obtained (out of the total of 100 t-values calculated). There are no significant return autocorrelations.

Lag	Autocorrelation	Average *t*-value	Significant Positive *t*-values*	Significant Negative *t*-values*
1	0.006	0.177	5	1
2	−0.001	−0.018	4	3
3	−0.005	−0.155	1	3
4	−0.002	−0.069	2	4
5	0.001	0.017	5	1
6	0.002	0.049	3	2
7	0.001	0.039	0	0
8	0.002	0.058	3	3
9	−0.002	−0.063	3	6
10	0.004	0.124	3	0
11	0.001	0.032	4	3
12	−0.004	−0.128	0	4
13	−0.002	−0.050	1	6
14	0.002	0.048	3	4
15	0.003	0.094	3	3
16	−0.005	−0.164	4	4
17	0.003	0.102	2	2
18	−0.006	−0.196	0	2
19	−0.003	−0.087	2	5
20	−0.002	−0.062	3	3
21	−0.006	−0.198	4	5
22	0.002	0.065	4	2
23	−0.005	−0.165	2	3
24	−0.004	−0.122	1	3
25	−0.005	−0.149	2	4
26	0.001	0.034	2	2
27	0.001	0.020	1	2
28	−0.003	−0.087	6	7
29	−0.002	−0.045	4	3
30	−0.002	−0.047	3	1
31	0.001	0.034	2	1
32	−0.004	−0.108	0	0
33	0.001	0.018	2	2
34	0.000	−0.013	4	3
35	−0.003	−0.087	1	3
36	−0.005	−0.143	2	4
37	−0.003	−0.106	2	5
38	−0.001	−0.022	3	2
39	0.002	0.052	3	4
40	−0.001	−0.036	1	2

* Significant at 95%.

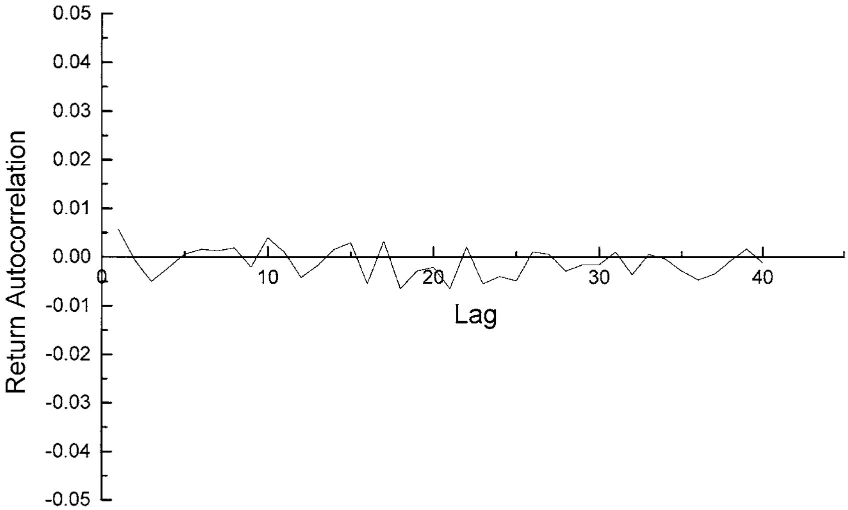


FIGURE 7.2 Return autocorrelation in benchmark model.

examine the effect of introducing into the market EMB investors, which model empirically and experimentally documented elements of investors' behavior.

7.4. RESULTS OF THE LLS MODEL WITH A SMALL MINORITY OF EMBS

In this section we will show that the introduction of a small minority of heterogeneous EMB investors who believe that the market is efficient and who estimate the *ex ante* return distribution from the *ex post* distribution, generates many of the empirically observed market "anomalies" that are absent in the benchmark model and, indeed, in most other rational-representative-agent models. We take this as strong evidence that the "nonrational" elements of investor behavior that are documented in experimental studies and the heterogeneity of investors, both of which are incorporated in the LLS model, are crucial to understanding the dynamics of the market.

In presenting the results of the LLS model with EMB investors, we take an incremental approach. We begin by describing the results of a model with a small subpopulation of *homogeneous* EMB investors. This model produces the market anomalies mentioned earlier; however, it produces unrealistic cyclic market dynamics. Thus, this model is presented both for analyzing the source of the anomalies in a simplified setting and as a reference point with which to compare the dynamics of the model with a *heterogeneous* EMB believer population.

We investigate the effects of investors' heterogeneity by first analyzing the case in which there are two types of EMBs. The two types differ in the method they use to estimate the *ex ante* return distribution. Namely, the first type looks at the set of the last m_1 *ex post* returns, whereas the second type looks at the set of the last m_2 *ex post* returns. It turns out that the dynamics in this case are much more complicated than a simple "average" between the case where all EMB investors have m_1 and the case where all EMB investors have m_2. Rather, there is a complex nonlinear interaction between the two EMB subpopulations. This implies that the heterogeneity of investors is a very important element determining the market dynamics, an element that is completely absent in representative-agent models.

Finally, we present the case where there is an entire spectrum of EMB investors differing in the number of *ex post* observations they take into account when estimating the *ex ante* distribution. This general case generates very realistic-looking market dynamics with all of the previously mentioned market anomalies.

7.4.1 Homogeneous Subpopulation of EMBs

When a very small subpopulation of EMB investors is introduced to the benchmark LLS model, the market dynamics change dramatically. Figure 7.3 depicts a typical price path in a simulation of a market with 95% RII investors and 5% EMB investors. The EMB investors have $m = 10$ (i.e., they estimate the *ex ante* return distribution by observing the set of the last 10 *ex post* returns). σ, the standard deviation of the random noise affecting the EMBs' decision making, is taken as 0.2. All investors, RII and EMB alike, have the same risk aversion parameter $\alpha = 1.5$ (as before). In the first 150 trading periods, the price dynamics look very similar to the typical dynamics of the benchmark model. However, after the first 150 or so periods, the price dynamics change. From this point onward the market is characterized by periodic booms and crashes. Of course, Figure 7.3 describes only one simulation. However, as will become evident shortly, different simulations with the same parameters may differ in detail, but the pattern is general: at some stage (not necessarily after 150 periods) the EMB investors induce cyclic price behavior. It is quite astonishing that such a small minority of only 5% of the investors can have such a dramatic impact on the market.

To understand the periodic booms and crashes, let us focus on the behavior of the EMB investors. After every trade, the EMB investors revise their estimation of the *ex ante* return distribution, because the set of *ex post* returns they employ to estimate the *ex ante* distribution changes. Namely, investors add the latest return generated by the stock to this set

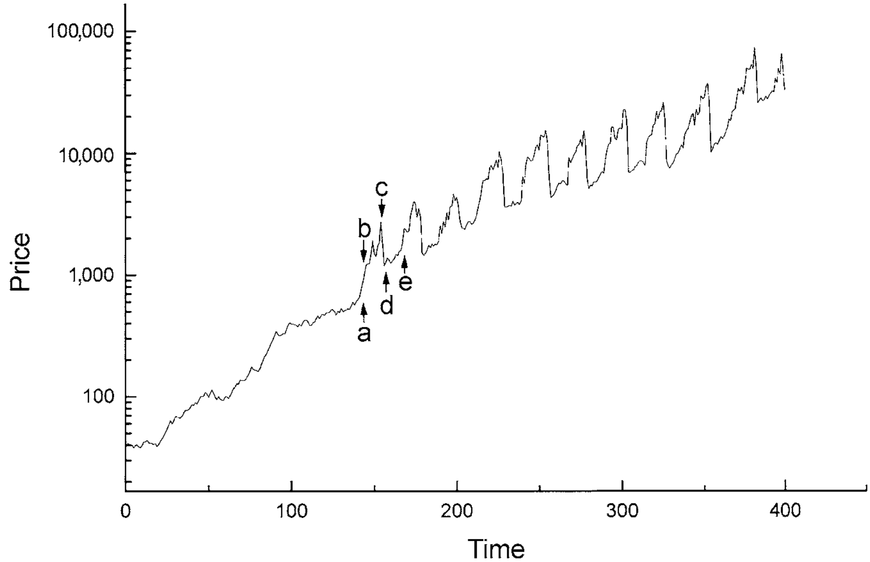


FIGURE 7.3 5% of investors are efficient market believers; 95% are rational informed investors.

and delete the oldest return from this set. As a result of this update in the estimation of the *ex ante* distribution, the optimal investment proportion x^* changes, and EMB investors revise their portfolios at next period's trade. During the first 150 or so periods, the informed investors control the dynamics and the returns fluctuate randomly (as in the benchmark model). As a consequence, the investment proportion of the EMB investors also fluctuates irregularly. Thus, during the first 150 periods the EMB investors do not affect the dynamics much. However, at point **a** the dynamics change qualitatively (see Figure 7.3). At this point, a relatively high dividend is realized and, as a consequence, a relatively high return is generated. This high return leads the EMB investors to increase their investment proportion in the stock at the next trading period. This increased demand of the EMB investors is large enough to affect next period's price, and thus a second high return is generated. Now the EMB investors look at a set of *ex post* returns with two high returns, and they increase their investment proportion even further. Thus, a positive feedback loop is created.

Notice that as the price goes up, the informed investors realize that the stock is overvalued relative to the fundamental value P^f and they decrease their holdings in the stock. However, this effect does not stop the price increase and break the feedback loop because the EMB investors continue to buy shares aggressively. The positive feedback loop pushes the stock price further and further up to point **b**, at which the EMBs are invested 100% in the stock. At point **b** the positive feedback loop "runs out of gas." However, the stock price remains at the high level because the EMB investors remain fully invested in the stock (the set of past $m = 10$ returns includes at this stage the very high returns generated during the "boom"—segment **a-b** in Figure 7.3).

When the price is at the high level (segment **b-c**), the dividend yield is low, and as a consequence, the returns are generally low. As time goes by and we move from point **b** towards point **c**, the set of $m = 10$ last returns gets filled with low returns. Despite this fact, the extremely high returns generated in the boom are also still in this set, and they are high enough to keep the EMB investors fully invested. However, 10 periods after the boom, these extremely high returns are pushed out of the set of relevant *ex post* returns. When this occurs, at point **c**, the EMB investors face a set of low returns, and they cut their investment proportion in the stock sharply. This causes a dramatic crash (segment **c-d**). Once the stock price goes back down to the "fundamental" value, the informed investors come back into the picture. They buy back the stock and stop the crash.

The EMB investors stay away from the stock as long as the *ex post* return set includes the terrible return of the crash. At this stage the informed investors regain control of the dynamics and the stock price remains close to its fundamental value. Ten periods after the crash, the extremely negative return of the crash is excluded from the *ex post* return set, and the EMB investors start increasing their investment proportion in the stock (point **e**). This drives the stock price up, and a new boom-crash cycle is initiated. This cycle repeats itself over and over almost periodically.

Figure 7.3 depicts the price dynamics of a single simulation. One may therefore wonder how general the results discussed here are. Figure 7.4 shows two more simulations with the same parameters but different dividend realizations. It is evident from this figure that although the simulations vary in detail (because of the different dividend realizations), the overall price pattern with periodic boom-crash cycles is robust.

Although these dynamics are very unrealistic in terms of the periodicity, and therefore the predictability of the price, they do shed light on the mechanism generating many of the empirically observed market phenomena. In the next section, when we relax the assumption that the EMB population is homogeneous with respect to m, the price is no longer cyclic or predictable, yet the mechanisms generating the market phenomena are the same as in this homogeneous EMB population case. The homogeneous EMB population case generates the following market phenomena:

Heavy Trading Volume

As explained previously, shares change hands continuously between the RII investors and the EMB investors. When a "boom" starts the EMB

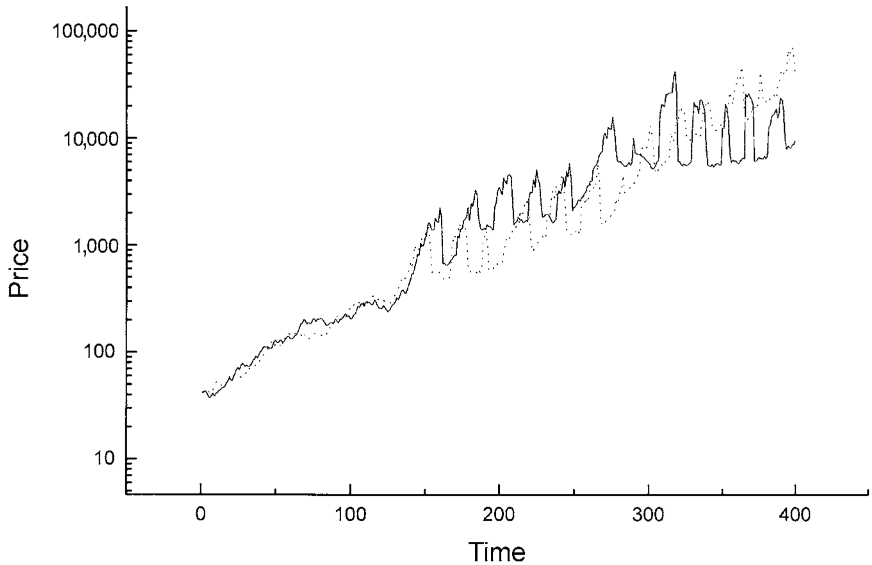


FIGURE 7.4 Two more simulations—same parameters as Figure 7.3, with different divident realizations.

investors observe higher *ex post* returns and become more optimistic, while the RII investor view the stock as becoming overpriced and become more pessimistic. Thus, at this stage the EMBs buy most of the shares from the RIIs. When the stock crashes, the opposite is true: the EMBs are very pessimistic, but the RII investors buy the stock once it falls back to the fundamental value. Thus, there is substantial trading volume in this market. The average trading volume in a typical simulation is about 1000 shares per period, or 10% of the total number of outstanding shares.

Autocorrelation of Returns

The cyclic behavior of the price yields a very definite return autocorrelation pattern. Table 7.2 shows the return autocorrelation and t-values for lags 1–40, averaged over 100 independent simulations. The autocorrelation pattern is depicted graphically in Figure 7.5. The autocorrelation pattern is directly linked to the length of the price cycle, which in turn are determined by m. Since the moving window of *ex post* returns used to estimate the *ex ante* distribution is $m = 10$ periods long, the price cycles are typically a little longer than 20 periods long: a cycle consists of the positive feedback loop (segment **a–b** in Figure 7.3), which is about 2 to 3 periods long, the upper plateau (segment **b–c** in Figure 7.3), which is about 10 periods long, the crash that occurs during one or two periods, and the lower plateau (segment **d–e** in Figure 7.3), which is again about 10 periods

TABLE 7.2 Return Autocorrelations in a Model with a Minority of Homogeneous EMB Investors

Autocorrelation of returns in a model in which 95% of the investors are RII and 5% of the investors are EMBs with $m = 10$. We run 100 independent 1000-period-long simulations. For each simulation, we run a univariate regression of R_t on R_{t-j} for different lags $j = 1 \ldots 40$. For each lag, we report the average autocorrelation and average t-value (averaged over all 100 simulations), as well as the number of significant positive and negative t-values obtained (out of the total of 100 t-values calculated).

Lag	Autocorrelation	Average *t*-value	Significant Positive *t*-values*	Significant Negative *t*-values*
1	−0.012	−0.368	0	4
2	0.021	0.646	4	0
3	0.001	0.041	0	0
4	0.011	0.326	0	0
5	0.003	0.098	3	0
6	0.003	0.099	2	0
7	−0.002	−0.060	0	0
8	−0.005	−0.140	0	2
9	−0.001	−0.041	0	0
10	0.024	0.731	8	0
11	−0.264	−8.429	0	100
12	−0.212	−6.691	0	100
13	−0.104	−3.228	0	94
14	−0.046	−1.420	0	27
15	−0.043	−1.336	0	17
16	−0.037	−1.126	0	17
17	−0.029	−0.885	0	10
18	−0.031	−0.970	0	14
19	−0.024	−0.725	2	8
20	−0.024	−0.731	0	6
21	−0.038	−1.175	0	13
22	0.018	0.556	6	0
23	0.066	2.057	49	0
24	0.058	1.783	37	0
25	0.051	1.583	34	0
26	0.051	1.582	34	0
27	0.042	1.310	28	0
28	0.033	1.008	17	0
29	0.038	1.176	23	0
30	0.046	1.410	21	0
31	0.038	1.175	23	0
32	0.038	1.181	25	0
33	0.032	0.980	22	0
34	0.015	0.469	11	1
35	−0.003	−0.092	5	7
36	−0.015	−0.471	2	10
37	−0.016	−0.495	2	16
38	−0.016	−0.483	3	7
39	−0.015	−0.455	3	10
40	−0.024	−0.742	0	17

* Significant at 95%.

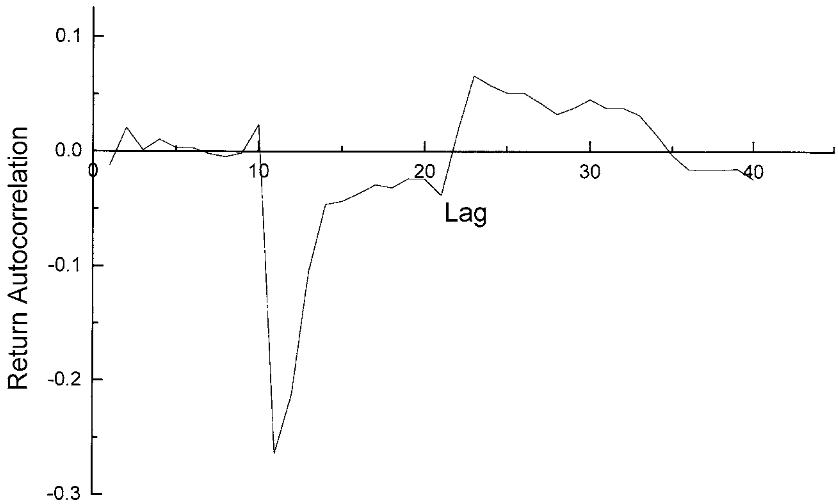


FIGURE 7.5 Return autocorrelation, 5% efficient market believers, $m = 10$.

long, for a total of about 23 to 25 periods. Thus, we expect positive autocorrelation for lags of about 23 to 25 periods, because this is the lag between one point and the corresponding point in the next (or previous) cycle. We also expect negative autocorrelation for lags of about 10 to 12 periods, because this is the lag between a boom and the following (or previous) crash, and vice versa. This is precisely the pattern we observe in Figure 7.5. Table 7.2 confirms that this autocorrelation pattern is statistically significant.

Excess Volatility

The EMB investors induce large deviations of the price from the fundamental value. Thus, price fluctuations are caused not only by dividend fluctuations (as the standard theory suggests) but also by the endogenous market dynamics driven by the EMB investors. This "extra" source of fluctuations causes the price to be more volatile than the fundamental value P^f. Indeed, for 100 1000-period independent simulations with 5% EMB investors we find an average $\sigma(p_t)$ of 46.4, and an average $\sigma(p_t^f)$ of 23.5—that is, we have excess volatility of about 100%.

As a first step in analyzing the effects of heterogeneity of the EMB population, in the next section we examine a market with two types of EMB investors. In Section 7.4.3 we analyze a model in which there is a full spectrum of EMB investors.

7.4.2 Two Types of EMBs

One justification for using a representative agent in economic modeling is that although investors are heterogeneous in reality, one can model their collective behavior with one representative or "average" investor. In this section we show that this is generally not true. Many aspects of the dynamics result from the nonlinear interaction between different investor types. To illustrate this point, in this section we analyze a very simple case in which there are only two types of EMB investors: one with $m = 5$ and the other with $m = 15$. Each of these two types consists of 2% of the investor population, and the remaining 96% are informed investors. The representative agent logic may tempt us to think that the resulting market dynamics would be similar to that of one "average" investor (i.e., an investor with $m = 10$). Figure 7.6 shows that this is clearly not the case. Rather than seeing periodic cycles of about 23 to 25 periods (which correspond to the average m of 10, as in Figure 7.3), we see an irregular pattern. As before, the dynamics are first dictated by the informed investors. Then, at point **a**, the EMB investors with $m = 15$ induce cycles that are about 30 periods long. At point **b** there is a transition to shorter cycles induced by the $m = 5$ population, and at point **c** there is another transition back to longer cycles. What is going on?

These complex dynamics result from the nonlinear interaction between the different subpopulations. The transitions from one price pattern

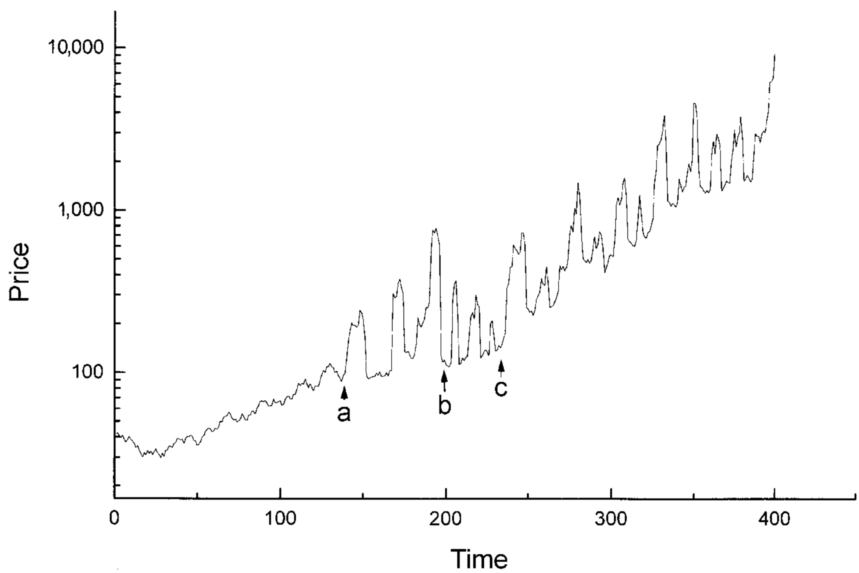


FIGURE 7.6 2% EMB $m = 5$, 2% EMB $m = 15$.

to another can be partly understood by looking at the wealth of each subpopulation. Figure 7.7 shows the proportion of the total wealth held by each of the two EMB populations (the remaining proportion is held by the informed investors). As seen in Figure 7.7, the cycles that start at point **a** are dictated by the $m = 15$ rather than the $m = 5$ population, because at this stage the $m = 15$ population controls more of the wealth than the $m = 5$ population. However, after three cycles (at point **b**) the picture is reversed. At this point the $m = 5$ population is more powerful than the $m = 15$ population, and there is a transition to shorter boom-crash cycles. At point **c** the wealth of the two subpopulations is again almost equal, and there is another transition to longer cycles. Thus, the complex price dynamics can be partly understood from the wealth dynamics. But how are the wealth dynamics determined? Why does the $m = 5$ population become wealthier at point **b**, and why does it lose most of this advantage at point **c**? It is obvious that the wealth dynamics are influenced by the price dynamics, thus there is a complicated two-way interaction between the two. Although this interaction is generally very complex, some principle ideas about the mutual influence between the wealth and price patterns can be formulated. For example, a population that becomes dominant and dictates the price dynamics typically starts underperforming, because it affects the price with its actions: this means pushing the price up when buying, and therefore buying high, and pushing the price down when selling. However, a more detailed analysis must consider the specific

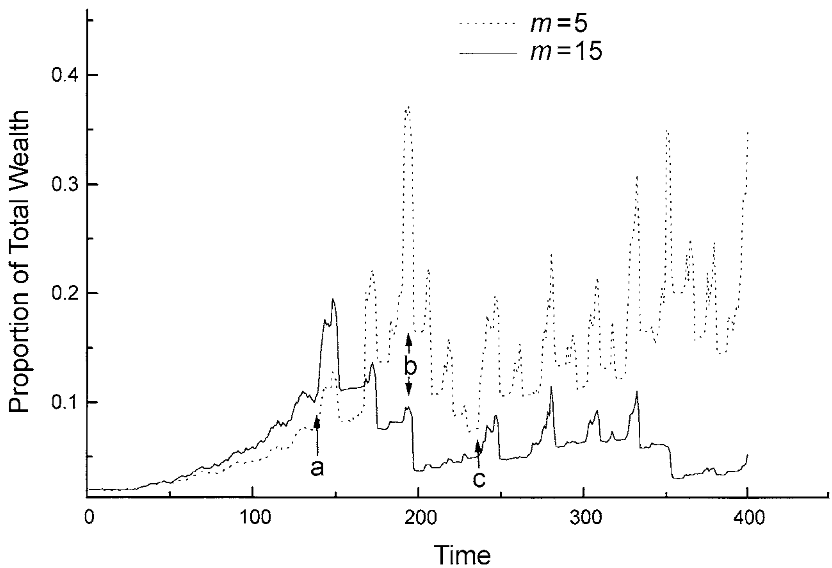


FIGURE 7.7 Proportion of the total weath held by the two EMB populations.

investment strategy employed by each population. For a more comprehensive analysis of the interaction between heterogeneous EMB populations, see Levy, Persky, and Solomon (1996).

The two-EMB-population model generates the same market phenomena as did the homogeneous population case: heavy trading volume, return autocorrelations, and excess volatility. Although the price pattern is much less regular in the two-EMB-population case, there still seems to be a great deal of predictability about the prices. Moreover, the booms and crashes generated by this model are unrealistically dramatic and frequent. In the next section, we analyze a model with a continuous spectrum of EMB investors. We show that this fuller heterogeneity of investors leads to very realistic price and volume patterns.

7.4.3 Full Spectrum of EMB Investors

Up to this point we have analyzed markets with at most three different subpopulations (one RII population and two EMB populations). The market dynamics we found displayed the empirically observed market anomalies, but they were unrealistic in the magnitude, frequency, and semipredictability of booms and crashes. In reality, we would expect not only two or three investor types, but rather an entire spectrum of investors. In this section we consider a model with a full spectrum of different EMB investors. It turns out that *more is different*. When there is an entire range of investors, the price dynamics become realistic: booms and crashes are not periodic or predictable, and they are also less frequent and dramatic. At the same time, we still obtain all of the market anomalies described before.

In this model each investor has a different number of *ex post* observations which he utilizes to estimate the *ex ante* distribution. Namely, investor i looks at the set of the m^i most recent returns on the stock, and we assume that m^i is distributed in the population according to a truncated normal distribution with average $\overline{m}$ and standard deviation σ_m (as $m \leq 0$ is meaningless, the distribution is truncated at $m = 0$).

Figure 7.8 shows the price pattern of a typical simulation of this model. In this simulation, 90% of the investors are informed, and the remaining 10% are heterogeneous EMB investors with $\overline{m} = 40$, and $\sigma_m = 10$. The price pattern seems very realistic with "smoother" and more irregular cycles. Crashes are dramatic, but infrequent and unpredictable. The heterogeneous EMB population model generates the following empirically observed market phenomena.

Return Autocorrelation: Momentum and Mean-Reversion

In the heterogeneous EMB population model trends are generated by the same positive feedback mechanism that generated cycles in the homo-

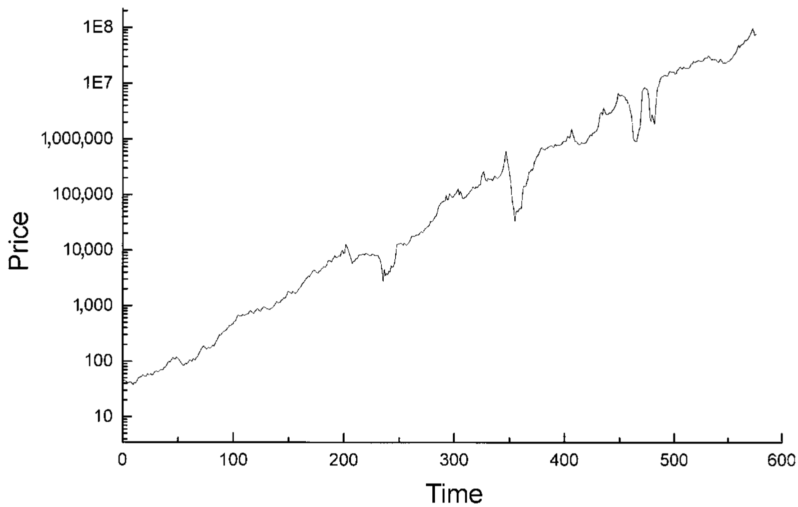


FIGURE 7.8 Spectrum of heterogeneous EMB investors (10% EMB investors, 90% RII investors).

geneous case: high (low) returns tend to make the EMB investors more (less) aggressive, this generates more high (low) returns, and so on. The difference between the two cases is that in the heterogeneous case there is a very complicated interaction between all the different investor subpopulations and as a result there are no distinct regular cycles but rather, smoother and more irregular trends. There is no single cycle length—the dynamics are a combination of many different cycles. This makes the autocorrelation pattern also smoother and more continuous. The return autocorrelations in the heterogeneous model are shown in Table 7.3 and in Figure 7.9. This autocorrelation pattern conforms with the empirical findings. In the short run (lags 1 through 4), the autocorrelation is positive —this is the empirically documented phenomena known as momentum: in the short run, high returns tend to be followed by more high returns, and low returns tend to be followed by more low returns. In the longer run (lags 5 through 13), the autocorrelation is negative, which is known as mean reversion. For even longer lags the autocorrelation eventually tends to zero. The short-run momentum, longer run mean reversion, and eventual diminishing autocorrelation creates the general U-shape that is found in empirical studies (Fama and French, 1988; Jegadeesh and Titman, 1993; Poterba and Summers, 1988), and is shown in Figure 7.9.

Excess Volatility

The price level is generally determined by the fundamental value of the stock. However, as in the homogeneous EMB population case, the

TABLE 7.3 Return Autocorrelations in a Model with a Minority of Heterogeneous EMB Investors

Autocorrelation of returns in a model in which 90% of the investors are RII and 10% of the investors are heterogeneous EMBs. m is distributed in the EMB population according to a truncated normal distribution with $\overline{m} = 40$ and $\sigma_m = 20$. We run 100 independent 1000-period-long simulations. For each simulation we run a univariate regression of R_t on R_{t-j} for different lags $j = 1 \ldots 40$. For each lag we report the average autocorrelation and average t-value (averaged over all 100 simulations), as well as the number of significant positive and negative t-values obtained (out of the total of 100 t-values calculated).

Lag	Autocorrelation	Average *t*-value	Significant Positive *t*-values*	Significant Negative *t*-values*
1	0.2609	8.3455	100	0
2	0.1918	6.0355	99	0
3	0.124	3.8538	82	0
4	0.0541	1.6726	40	1
5	− 0.0637	− 1.9706	3	12
6	− 0.1254	− 3.8974	1	58
7	− 0.1771	− 5.5491	0	65
8	− 0.1889	− 5.9354	0	99
9	− 0.186	− 5.8391	0	96
10	− 0.1669	− 5.2303	0	99
11	− 0.1365	− 4.2657	0	70
12	− 0.0797	− 2.4733	9	37
13	− 0.0217	− 0.6709	19	16
14	0.0528	1.6488	34	4
15	0.0695	2.1558	41	7
16	0.1062	3.2979	49	7
17	0.1261	3.9333	56	4
18	0.0982	3.057	68	6
19	0.0718	2.2258	60	6
20	0.0547	1.6953	41	3
21	0.0114	0.3448	18	3
22	− 0.0325	− 1.0093	9	20
23	− 0.0411	− 1.2874	8	19
24	− 0.0668	− 2.0784	7	55
25	− 0.0581	− 1.8055	8	48
26	− 0.0571	− 1.7686	12	55
27	− 0.0299	− 0.923	4	15
28	− 0.0071	− 0.2197	4	12
29	− 0.0121	− 0.3705	2	29
30	0.0179	0.5607	3	5
31	0.0197	0.6159	4	15
32	0.0324	1.016	15	10
33	0.0261	0.8083	14	6
34	0.0535	1.6595	37	6
35	0.028	0.8643	23	7
36	0.0071	0.2168	11	7
37	− 0.0098	− 0.3067	7	6
38	0.0077	0.2349	9	11
39	− 0.0057	− 0.1797	5	9
40	− 0.0245	− 0.7643	4	11

* Significant at 95%.

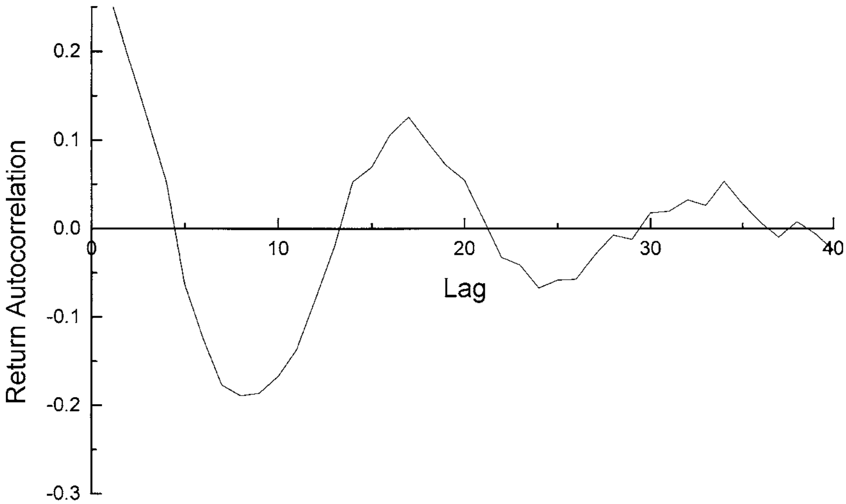


FIGURE 7.9 Return autocorrelation—heterogeneous EMB population.

EMB investors occasionally induce temporary departures of the price away from the fundamental value. These temporary departures from the fundamental value make the price more volatile than the fundamental value. Following Shiller's methodology (see Section 7.3), we define the detrended price, p, and fundamental value, p^f. Averaging 100 independent simulations we find $\sigma(p) = 27.1$ and $\sigma(p^f) = 19.2$, which is an excess volatility of 41%.

Heavy Volume

As investors in our model have different information (the informed investors know the dividend process, while the EMB investors do not) and different ways of interpreting the information (EMB investors with different memory spans have different estimations regarding the *ex ante* return distribution), there is a high level of trading volume in this model. The average trading volume in this model is about 1700 shares per period (17% of the total outstanding shares). As explained next, the volume is positively correlated with contemporaneous and lagged absolute returns.

Volume Is Positively Correlated with Contemporaneous and Lagged Absolute Returns

Investors revise their portfolios as a result of changes in their beliefs regarding the future return distribution. The changes in the beliefs can be due to a change in the current price, to a new dividend realization (in the case of the informed investors), or to a new observation of an *ex post* return (in the case of the EMB investors). If all investors change their beliefs in the same direction (for example, if everybody becomes more optimistic), the stock price can change substantially with almost no volume —everybody would like to increase the proportion of the stock in his or her portfolio and this will push the price up, but a very small number of shares will change hands. This scenario would lead to zero or perhaps even a negative correlation between the magnitude of the price change (or return) and the volume. However, the typical scenario in the LLS model is different. Typically, when a positive feedback trend is induced by the EMB investors, the opinions of the informed investors and the EMB investors change in opposite directions. The EMB investors see a trend of rising prices as a positive indication about the *ex ante* return distribution, while the informed investors believe that the higher the price level is above the fundamental value, the more overpriced the stock is and the harder it will eventually fall. The exact opposite holds for a trend of falling prices. Thus, price trends are typically interpreted differently by the two investor types, and therefore induce heavy trading volume. The more pronounced the trend, the more likely it is to lead to heavy volume, and at the same time, to large price changes, which are due to the positive feedback trading on behalf of the EMB investors.

This explains not only the positive correlation between volume and contemporaneous absolute rates of return, but also the positive correlation between volume and lagged absolute rates of return. The reason is that the behavior of the EMB investors induces short-term positive return autocorrelation, or momentum (as discussed earlier). That is, a large absolute return this period is associated not only with high volume this period, but also with a large absolute return next period, and therefore with high volume next period. In other words, when there is a substantial price increase (decrease), EMB investors become more (less) aggressive and the opposite happens to the informed traders. As we have seen before, when a positive feedback loop is started, the EMB investors are more dominant in determining the price, and therefore another large price increase (decrease) is expected next period. This large price change is likely to be associated with heavy trading volume as the opinions of the two populations diverge. Furthermore, this large increase (decrease) is expected to make the EMB investors even more optimistic (pessimistic) leading to another large price increase (decrease) and heavy volume next period.

To verify this relationship quantitatively, we regress volume on contemporaneous and lagged absolute rates of return for 100 independent simulations. We run the regressions:

$$V_t = \alpha + \beta_C |R_t - 1| + \varepsilon_t \text{ and } V_t = \alpha + \beta_L |R_{t-1} - 1| + \varepsilon_t, \quad (7.17)$$

where V_t is the volume at time t, R_t is the total return on the stock at time t, and the subscripts C and L stand for contemporaneous and lagged. We find an average value of 870 for $\hat{\beta}_C$ with an average t-value of 5.0 and an average value of 886 for $\hat{\beta}_L$ with an average t-value of 5.1.

7.5. SURVIVABILITY OF THE EMB INVESTORS

One of the main arguments against models of semirational agents is that since these agents act suboptimally they will eventually be "wiped out" by the rational population. Thus, the argument goes, even though we find experimentally that most people deviate from rationality, we should not be concerned with modeling these deviations from rationality, as they should not have a substantial effect on the market.

In contrast, in the LLS model we see that the existence of a small population of uninformed and semirational agents has a dramatic effect on the market dynamics and is responsible for many of the empirically observed market "anomalies." How can a small minority of uninformed semirational investors have such a large effect? And why isn't this minority wiped out?

It is instructive to analyze closely what happens to the wealth of investors through a typical cycle in which the price moves away from the fundamental value and eventually comes back to this level. Figure 7.10 is a simplified schematic representation of a typical cycle (detrended). The dotted line is the fundamental value, and the solid line is the price. When the price becomes higher than the fundamental value (at point **a**), the informed investors believe that the stock is overpriced and sell their shares. They buy the shares again only when the price goes back down to the fundamental value (at point **d**). Thus, the informed investors sell at a price that is slightly higher than the fundamental value and buy at a price that is slightly lower than the fundamental value, and they net this small difference.

A "typical" EMB investor buys aggressively at point **a**, increases his investment up to point **b**, at which he is fully invested in the stock, and remains fully invested up to point **c**, at which he sells. Thus, a "typical" EMB investor buys most of his shares somewhere close to point **a** and sells at point **d**. Hence, the "typical" EMB investor does not gain or lose much during the cycle. However, a crucial point is that the decision making of the EMB investors is "noisy" (i.e., $x^i = x^{*i} + \tilde{\varepsilon}^i$). For example, if the optimal investment proportion in the stock for two identical investors is 90%, one investor may hold 85% in the stock and the other may hold 95%. This means that some EMB investors will start buying before or after point **a**, and others will start selling before or after point **c**. The randomness of the noisy decision making makes some investors luckier than the typical or

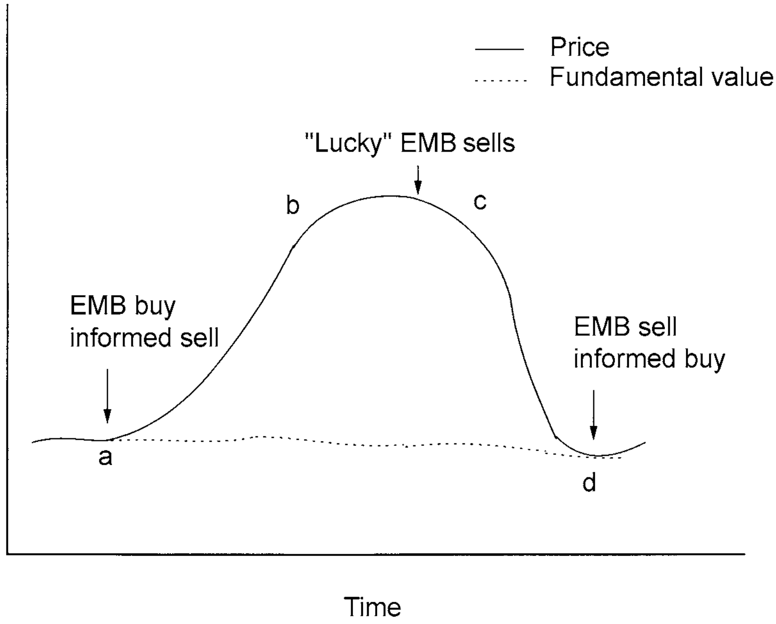


FIGURE 7.10 A schematic cycle.

average investor, and others less lucky. Most of the deviations from the typical behavior are not economically significant. For example, if an investor starts buying a few periods before **a**, this will not make her substantially more or less wealthy than the typical EMB investor who buys at **a**. However, there is one type of deviation from the typical EMB behavior, which is very profitable. Namely, if an investor buys at **a** but starts selling before **c**, she will accumulate a great deal of wealth during the cycle.

The bottom line is that even though the typical EMB investor does slightly worse than the informed investor during the cycle, there are a few "lucky" EMB investors who do much better. The overall effect of such a cycle on the relative wealth of the entire EMB population compared with the wealth of the entire informed population can go both ways. Which population gains more from the cycle may depend on many factors, one of which may be the initial wealth of each population *before* going into the cycle. Figure 7.11 shows the proportion of the total wealth held by the informed population in a simulation with 90% informed investors and 10% heterogeneous EMB investors. In this simulation the initial wealth of each investor was identical, thus the informed population initially holds 90% of the wealth. This figure reveals that there is a constant "tug-of-war" between the two populations. It is clear that the EMB population defi-

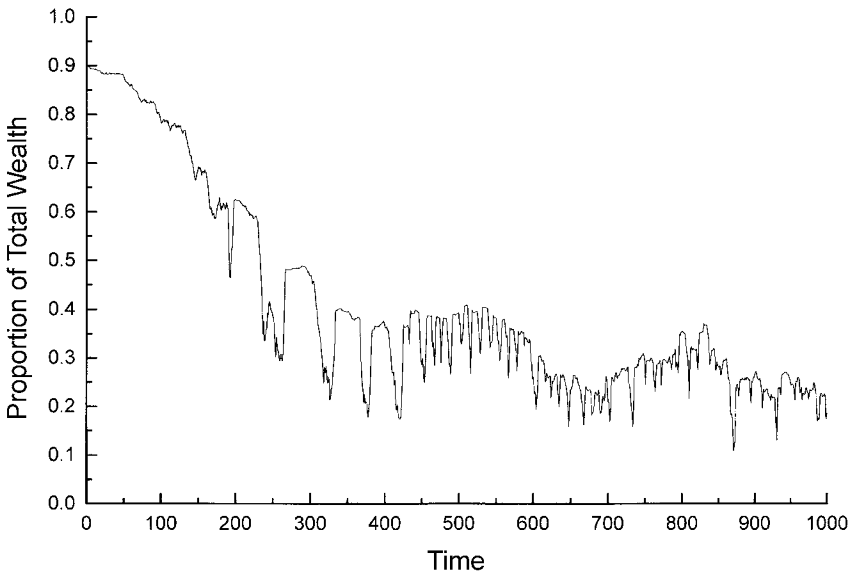


FIGURE 7.11 Proportion of the total wealth held by the rational informed investors.

nitely survives this war and is anything but wiped out by the informed population.

The surprising survivability, and even prosperity, of the EMB investors raises two questions:

1. Given that the EMB population benefits on average from the noisy decision making, then maybe if the RII investors also made noisy decisions, the EMB population would lose its edge (and get wiped out)?
2. Is the performance of the EMB strategy superior to the performance of the RII strategy? If this is the case, we would expect the RII investors to realize this and change strategies.

The answer to the first question is negative. When we run a simulation with the same parameters as in the simulation depicted in Figure 7.11 but with noise added to the informed investors' decision making, the results remain qualitatively the same (see Fig. 7.12). The reason for this is that the noise does not induce significant gains (or losses) to the informed investors. Even if a lucky informed investor has an investment proportion in the stock that is high relative to the rational proportion x^* (because of the noise) when a boom starts, he is not likely to maintain his increased position in the stock in the next period (because the optimal investment proportion in the stock for the RII investors decreases as the price increases). Hence, even if a lucky RII investor increases the number of his

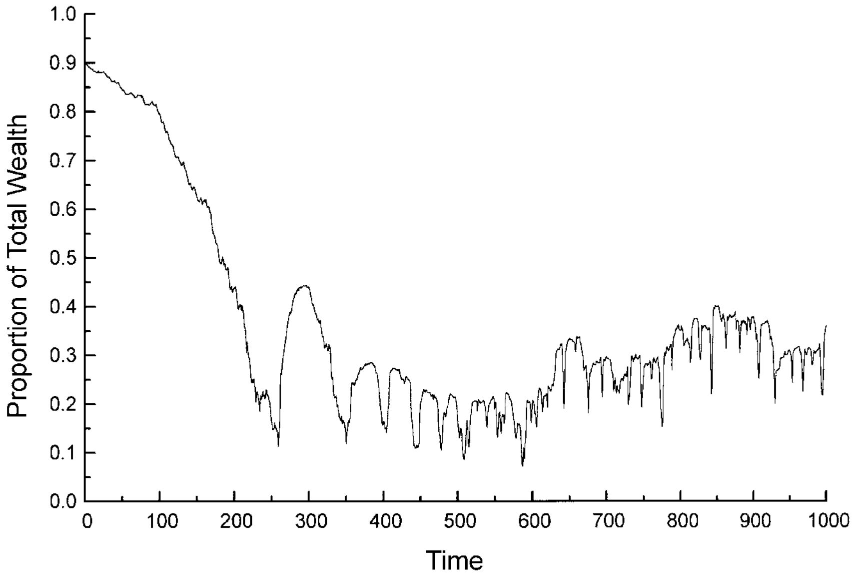


FIGURE 7.12 Proportion of the total wealth held by the informed investors who make "noisy" decisions.

shares just when a boom starts, he is likely to sell most of the extra shares he bought in the next period, and the gains from such a lucky "error" are not large. In contrast, a lucky EMB investor, who happens to increase her stock position before a boom is likely to maintain her increased position throughout the boom and reap large capital gains. Thus, the noise term does not have the same effect for the RII population as it does on the EMB population. Adding a noise term to the decision making of the RII investors does not change the fact that the EMB population survives and prospers.

Figures 7.11 and 7.12 may suggest that the EMB strategy outperforms the RII strategy. However, this is not the case, and the answer to the second question posed earlier is also negative. Although the EMB population as a whole increases its proportion of the total wealth, from the point of view of the individual investor the EMB strategy is much more risky than the RII strategy. Figure 7.13 depicts the distribution of one-period rates of return to the RII and EMB populations in the simulation described in Figure 7.11. While the mean one-period rate of return to the EMB investors is greater than that of the RII investors ($\bar{r}_{\mathrm{EMB}} = 6.0\%$, $\bar{r}_{\mathrm{RII}} = 2.9\%$), the returns to the EMB investors are also more risky ($\sigma_{\mathrm{EMB}} = 8.9\%$, $\sigma_{\mathrm{RII}} = 1.5\%$). Analysis of the cumulative probability distri-

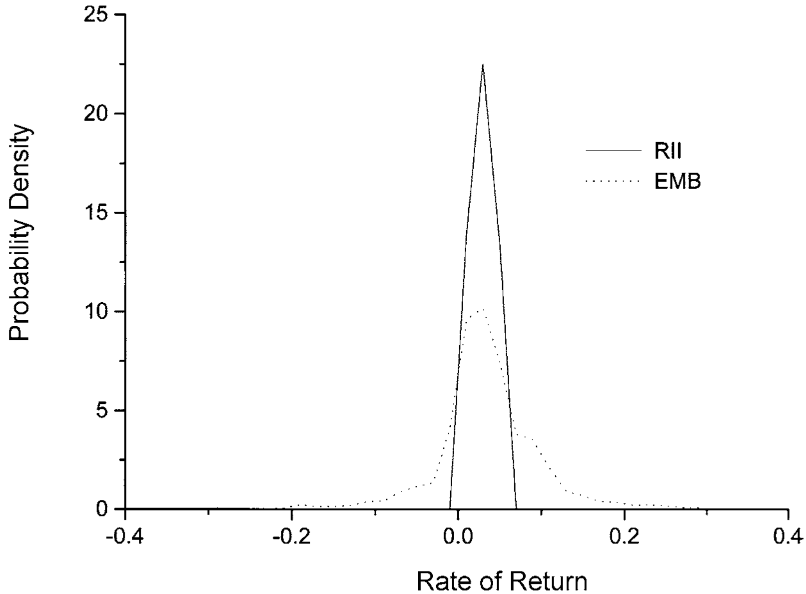


FIGURE 7.13 Distribution of one-period rates of return to the RII and EMB investors.

butions (Figure 7.14) reveals that there is no first- or second-degree stochastic dominance of one strategy over the other (the "+" area is 1.44, and the "−" area is 2.30). (For a review of stochastic dominance rules see Levy, 1998b.) Hence, it is not possible to assert that one of the strategies is better than the other. This justifies the coexistence of the various investor types in the market.

7.6. SUMMARY

The LLS model is a microscopic simulation model that incorporates some of the fundamental experimental findings regarding the behavior of investors. The main nonstandard assumption of the model is that there is a small minority of investors in the market who are uninformed about the dividend process and who believe in market efficiency. The investment decision of these investors is reduced to the optimal diversification between the stock and the bond. The three main characteristics of the EMB investors are as follows:

a) They employ the *ex post* return distribution in order to estimate the *ex ante* return distribution.

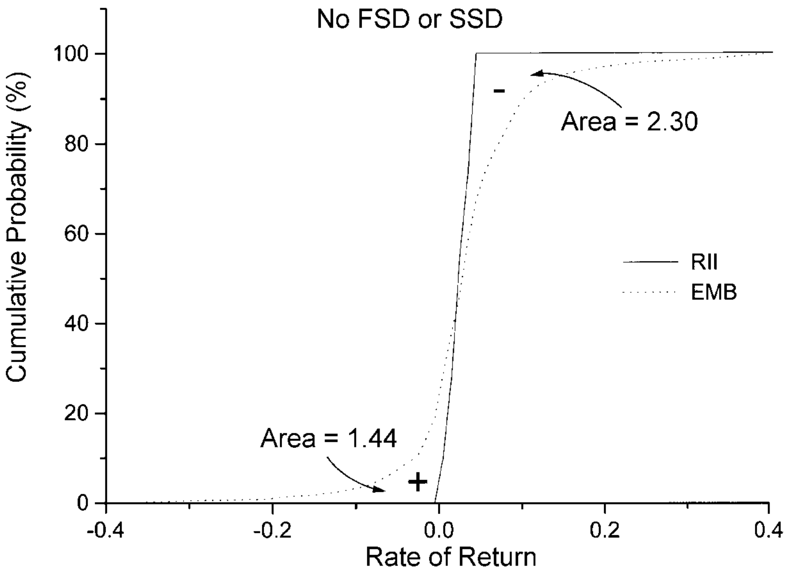


FIGURE 7.14 Cumulative rate of return distributions: no FSD or SSD.

b) They may be heterogeneous in the way they form their expectations (i.e., $m^i \neq m^j$).
c) Their investment decision is based on expected utility maximization; however, they may deviate to some extent from optimality (i.e., $x^i = x^{*i} + \tilde{\varepsilon}^i$).

The LLS model generates many of the empirically documented market phenomena that are hard to explain in the analytical rational-representative-agent framework. These phenomena are as follows:

- Short-term momentum
- Longer term mean reversion
- Excess volatility
- Heavy trading volume
- Positive correlation between volume and contemporaneous absolute returns
- Positive correlation between volume and lagged absolute returns
- Endogenous market crashes

The fact that so many "puzzles" are explained with a simple model built on a small number of empirically documented behavioral elements leads us to suspect that these behavioral elements are very important in understanding the workings of the market. This is especially true in light of the observations that a small minority of the nonstandard bounded-rational investors can have a dramatic influence on the market and that these investors are not wiped out by the majority of rational investors. As there is no dominance of one investor group over the other (by FSD or SSD), we expect to have a coexistence of the various groups in equilibrium. None of the groups is inferior to the other, and therefore none of the groups vanish.

In Chapter 9 we further pursue our investigation of the effects of various behavioral elements on the market dynamics and on asset pricing. Namely, we analyze the effects of the behavioral elements of prospect theory on the market dynamics in the LLS framework. We do so by extending the LLS model to include investors who are characterized by an S-shaped value function (rather than a utility function) and who base their decisions on subjective probability weights rather than on the objective probabilities.

APPENDIX 7.1

In the LLS model, RII investors assume an apparently arbitrary discount factor k. *A priori* the arbitrary k seems unjustified and disconnected from the actual average rate of return on the stock. In this appendix, we show that for any reasonable set of parameters in the benchmark model, the discount factor turns out to be self-consistent (i.e., the average actual rate of return on the stock $\bar{r}$ is approximately equal to the assumed discount factor).

The actual rate of return on the stock is given by

$$\tilde{r}_t = \tilde{R}_t - 1 = \frac{\tilde{P}_{t+1} - P_t + \tilde{D}_{t+1}}{P_t} \qquad (7.18)$$

In the benchmark model, the stock price is always close to the fundamental value (because the RII investors assume that the stock price will converge to the fundamental value next period); hence, $P_t \approx P_t^f$. Since the fundamental value grows on average at a rate of $1 + g$ (see Eq. (7.3)), the actual price also grows on average approximately at this rate—that is, $\overline{P}_{t+1} \approx P_t(1 + g)$. Thus, the average rate of return on the stock is

$$\bar{r}_t = \frac{\overline{P}_{t+1} - P_t + \overline{D}_{t+1}}{P_t} \approx g + \frac{\overline{D}_{t+1}}{P_t} \qquad (7.19)$$

Recalling that $\overline{D}_{t+1} = (1 + g)D_t$ and that $P_t \approx P_t^f = \frac{(1+g)D_t}{k-g}$, we obtain

$$\bar{r}_t \approx g + \frac{(1+g)D_t}{P_t} \approx g + \frac{(1+g)D_t}{\left(\dfrac{(1+g)D_t}{k-g}\right)} = k \qquad (7.20)$$

APPENDIX 7.2

The expected utility of the RII investors as a function of their investment proportion in the stock, x, is given by (see Eq. (7.7))

$$EU\left(\tilde{W}^i_{t+1}\right) = \frac{(W^i_h)^{1-\alpha}}{1-\alpha}\int_{z_1}^{z_2}\left[(1-x)(1+r_f) + x\left(\frac{\dfrac{D_t(1+z)(1+g)}{k-g} + D_t(1+z)}{P_h}\right)\right]^{1-\alpha} f(z)\,dz$$

If z is uniformly distributed in the range $[z_1, z_2]$, we have $f(z) = \frac{1}{z_2 - z_1}$. Substituting this expression for $f(z)$ and rearranging, we obtain

$$EU\left(\tilde{W}^i_{t+1}\right) = \frac{(W^i_h)^{1-\alpha}}{1-\alpha}\frac{1}{(z_2-z_1)} \times \int_{z_1}^{z_2}\left[(1-x)(1+r_f) + \left(\frac{x}{P_h}\frac{k+1}{k-g}D_t\right)(1+z)\right]^{1-\alpha} dz$$

Integrating over z we obtain

$$EU\left(\tilde{W}^i_{t+1}\right) = \frac{(W^i_h)^{1-\alpha}}{(1-\alpha)(2-\alpha)}\frac{1}{(z_2-z_1)}\left(\frac{k-g}{k+1}\right)\frac{P_h}{xD_t} \times \left\{\left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right)D_t(1+z_2)\right]^{(2-\alpha)} - \left[(1-x)(1+r_f) + \frac{x}{P_h}\left(\frac{k+1}{k-g}\right)D_t(1+z_1)\right]^{(2-\alpha)}\right\}$$

APPENDIX 7.3

In this appendix we show that the demand for shares is a monotonically decreasing function of the stock price. This is not obvious, because although the RII investors' optimal investment proportion in the stock decreases with the stock price, if investors hold shares, their wealth increases with the price. Thus, there are two opposite effects of the price on the demand for shares. However, it turns out that the net effect yields a demand function that is monotonically decreasing with the stock price. To see this, recall that the demand of investor i for shares at a hypothetical price P_h is given by

$$N_h^i(P_h) = \frac{x_h(P_h)W_h^i(P_h)}{P_h} \tag{7.21}$$

where $x_h(P_h)$ is the optimal investment proportion in the stock given the hypothetical price P_h, and $W_h^i(P_h)$ is the investors' wealth given P_h. $W_h^i(P_h)$ can be thought of as the investor's total wealth before the trade (which we denote by W_{bt}^i), plus the capital gains or losses due to the change in price:

$$W_h^i(P_h) = W_{bt}^i + N_{t-1}^i(P_h - P_{t-1}) \tag{7.22}$$

(Recall that the capital gains and losses at time t are due to the shares held *before* the time t trade, and not shares bought or sold at time t.) Substituting $W_h^i(P_h)$ in the demand function, we obtain

$$\begin{aligned} N_h^i(P_h) &= \frac{x_h(P_h)(W_{bt}^i + N_{t-1}^i(P_h - P_{t-1}))}{P_h} \\ &= x_h(P_h)\left[N_{t-1}^i + \frac{(W_{bt}^i - N_{t-1}^i P_{t-1})}{P_h}\right] \end{aligned} \tag{7.23}$$

Notice that W_{bt}^i is the investor's total wealth before the trade, and $N_{t-1}^i P_{t-1}$ is the wealth held in the stock, thus $W_{bt}^i - N_{t-1}^i P_{t-1}$ is positive, and the square brackets on the right are therefore a decreasing function of P_h. Since $x_h(P_h)$ is decreasing in P_h for the RII investors (see Eq. (7.8)) (and is constant for the EMB investors), we obtain an aggregate demand function that is monotonically decreasing in the price.

8

VARIOUS FINANCIAL MICROSCOPIC SIMULATIONS

8.1. INTRODUCTION

Although the microscopic simulation (MS) method is not yet well known in the financial research community, quite a number of leading researchers have used MS techniques to study and explain certain financial phenomena. Most of the models that resulted, rather than aiming to create a comprehensive market simulation framework, have addressed well-defined questions on the microscopic origins of certain macroscopic phenomena and have used MS to support their hypotheses. We cannot hope to give a comprehensive review of MS in finance here. However, in this chapter we attempt to describe some of the pioneering studies that have shaped this rapidly growing field. Several other microscopic simulations are reviewed in Chapter 4 of Moss de Oliveira, de Oliveira, and Stauffer (1999) (see also Mantegna and Stanley, 1999, and Bouchaud and Potters, 1997).

In the studies described here, MS was employed to investigate models that are otherwise analytically intractable. The sources of the intractability in these models are the nonstandard assumption regarding investors' behavior, the complex interactions between different types of investors, the coevolution of various trading strategies, or a combination of all of the above (Solomon, 1999). As the MS methodology is one of the standard tools in physics, and since it is a relatively new tool in financial research, it is not surprising that some of the studies described here are the product of collaboration between physicists and researchers in finance and in economics.

8.2. STIGLER'S RANDOM TENDER STREAM MODEL

Stigler's (1964) study appears to be the first MS of market dynamics. Although this is not a full-blown MS market model, in the sense that there is no explicit modeling of investors and their trading strategies, Stigler's study was one of the first steps in the direction of MS. The context in which Stigler's simulations were produced was the debate on the criteria for an efficient stock market and on the trading mechanisms necessary to ensure efficiency. The specific focus was on the New York Stock Exchange (NYSE). Instead of the concept (proposed at the time by other researchers and practitioners) that the criterion for market efficiency is the smoothness of the price movements, Stigler defined market efficiency in terms of the fulfillment of the following requirements:

- If a bid exceeds the lowest asking price, a transaction takes place (and similarly for offers).
- Higher bids are fulfilled before lower ones.
- Prices fluctuate only within the limits of speculators' costs of providing a market.

Stigler asked the question: assuming that these three efficiency requirements hold, what are the market dynamics induced by a random sequence of bids and asks ("tenders"), which are spread around a fixed equilibrium price? The answer he obtained (using MS) was that even for normally distributed random tender inputs, in the absence of specialists/arbitrageurs, the market clearing mechanism outputs significant price variations. In fact, Stigler found that the departures of the transaction price from the equilibrium price can be very significant. The trading process, far from being a smooth flow, is characterized by long time intervals in which no matches (and transactions) take place.

As Stigler presents the main procedure of his model, most of the characteristic MS key concepts are implicit:

> Demand and supply are flows, and erratic flows with sequences of bids and asks dependent upon the random circumstances of individual traders.
>
> A random sequence of bids and asks . . . two-digit numbers from a table of random numbers are drawn, and the first digit determines whether it is a bid or ask (even or odd, respectively) and the second digit determines the (price) level of the bid or ask. (pp. 126–127)

Stigler considered a stock with 710,000 outstanding shares and a *fixed* equilibrium price whose value is somewhere between \$$29\frac{3}{4}$ and \$30 (the equilibrium price cannot be expressed more precisely than this due to the fact that the tender levels and the transaction prices are multiples of \$$\frac{1}{4}$). Since at the time, standard on-line computer random generators did not exist, the input tender time series was generated according to the following random procedure. A list of two-digit random numbers was used:

- The first digit served to decide if the corresponding tender was a bid (even digit) or an ask (odd digit).
- The second digit determined the level of the bid or ask: 0 represented \$$28\frac{3}{4}$ while 9 represented \$31. The values in between are distributed linearly: 1 corresponds to \$29, 2 to \$$29\frac{1}{4}$, 3 to \$$29\frac{1}{2}$, 4 to \$$29\frac{3}{4}$, 5 to \$30, and so on. Notice that the bids and asks are distributed around the equilibrium price.

Stigler assumes that the tenders arrive at the market in the time order in which the corresponding two-digit numbers appear in the random list. A transaction takes place whenever a bid at a certain level is followed at some latter stage by an ask at a lower level, or vice versa. The price of the transaction is then the level of the former tender. For instance, if the list contains a random number 48, this would mean that the corresponding order is a bid (4 is even) at the level $\$28\frac{3}{4} + (8 \times \frac{1}{4}) = \$30\frac{3}{4}$. If later the list contains the number 70 (corresponding to an ask of \$$29\frac{3}{4}$) then a transaction takes place at \$$30\frac{3}{4}$ (the value of the initial bid). As the sequence of two-digit numbers unfolds, one can imagine that the market receives sequentially the corresponding tenders. During the simulation Stigler measures the following:

- The sequence of transaction prices
- The time until a bid or ask is met (and a transaction occurs)

The first crucial observation in the paper is that the transaction prices deviate substantially from the (fixed) equilibrium price (the distance is taken to $29\frac{3}{4}$ or to 30, whichever is closer). On the other hand, there are significant erratic delays in filling bids and asks. These findings anticipate the later, more sophisticated/complex simulations that obtain systematically similar effects. Stigler went on and performed a series of further experiments on the characteristic price patterns when the following occurred:

- The random numbers representing the level of the tenders were distributed normally.
- The equilibrium price dropped sharply at some moment.
- The equilibrium price started to increase linearly at some moment.

The operative conclusions of Stigler's paper are that large price fluctuations and delays in fulfilling/matching market orders can be induced by a flow of random tenders, even in conditions of a stable fixed equilibrium price. This, in turn, indicates the presence of intrinsic difficulties with realizing market efficiency and liquidity, especially in the absence of arbitrage, specialists and speculators. Stigler's paper opened the way to further studies of the return distributions and of the time evolution of the market dynamics. Space limitations force us to jump directly to recent work omitting the important work done in between.

8.3. THE PORTFOLIO INSURERS MODEL OF KIM AND MARKOWITZ

Harry Markowitz is very well known for being one of the founders of modern portfolio theory. It is less well known, however, that he is also one of the pioneers in employing MS in financial study. In fact, in the 1960s he authored and coauthored two books on the SIMSCRIPT simulation programming language (Markowitz, 1963; Kiviat, Villanueva, and Markowitz, 1968).

Following the Wall Street crash of 1987, people started to look for endogenous market features, rather than external forces, as sources of price variation. The Kim-Markowitz MS model explains the 1987 crash as resulting from investors' "Constant Proportion Portfolio Insurance" (CPPI) policy. Kim and Markowitz (1989) proposed that market instabilities arise as a consequence of the individual insurers' efforts to cut their losses by selling once the stock prices are going down.

Various simplifications of this model were studied lately intensively by Egenter, Lux, and Stauffer (1999). The Kim Markowitz MS model involves two groups of individual investors: rebalancers and insurers (CPPI investors). The rebalancers are aiming to keep a constant composition of their portfolio, while the insurers make the appropriate operations to ensure that their eventual losses will not exceed a certain fraction of the investment per time period.

The rebalancers act to keep a portfolio structure with half of their wealth in cash and half in stocks. If the stock price rises, then the stocks weight in the portfolio will increase and the rebalancers will sell shares until the shares again constitute 50% of the portfolio. If the stock price decreases, then the value of the shares in the portfolio decreases, and the rebalancers will buy shares until the stock again constitutes 50% of the portfolio. Thus, the rebalancers have a stabilizing influence on the market by selling when the market rises and buying when the market falls.

A typical CPPI investor has as her main objective not to lose more than 25% of her initial wealth during a quarter, which consists of 65 trading days. Thus, the investor has to ensure that at each cycle 75% of the initial wealth is out of a reasonable risk. To this effect, the investor assumes that the current value of the stock will not fall in one day by more than a factor of 2. The result is that she always keeps in stock twice the difference between the present wealth and 75% of the initial wealth (which she had at the beginning of the 65 days investing period).[1] This determines the amount the CPPI agent is bidding or offering at each stage. Obviously, after a price fall, the amount she wants to keep in stocks will fall and the CPPI investor will sell and further destabilize the market. After an increase in the prices (and personal wealth), the amount the CPPI agent wants to keep in shares will increase: she will buy and may support a price bubble.

The simulations reveal that even a fraction of CPPI investors (i.e., less than 50%) is enough to destabilize the market, and crashes and booms are observed. Hence, the claim of Kim and Markowitz that the CPPI policy may be responsible for the 1987 crash is supported by the MS.[2]

8.4. THE FINANCIAL LIFE OF ARTHUR, HOLLAND, LEBARON, PALMER, AND TAYLER

The focus of the MS market model of Palmer, Arthur, Holland, Lebaron and Tayler (1994) and Arthur, Holland, Lebaron, Palmer, and Tayler (1997) is the concept of coevolution. This concept can be described in short as follows. Each investor adapts his or her investment strategy such as to maximally exploit the market dynamics generated by the investment strategies of all others investors. This leads to an ever-evolving market, driven endogenously by the ever-changing strategies of the investors.

The main objective of the Arthur *et al.* model is to prove that market fluctuations may be induced by this endogenous coevolution, rather than by exogenous events. Moreover, Arthur *et al.* studied the various regimes of the system: the regime in which rational fundamentalist strategies are dominating versus the regime in which investors start developing strategies based on technical trading. In the technical trading regime, if some of the investors follow fundamentalist strategies, they will be punished rather than rewarded by the market. Arthur *et al.* also studied the relation between the various strategies (fundamentals versus technical) and the volatility properties of the market (clustering, excess volatility, volume-volatility correlations, etc.).

[1] If the initial wealth at the beginning of the quarter is denoted by W_0 and the current wealth is denoted by W, then the investor holds $2(W - 0.75W_0)$ in the stock and $W - 2(W - 0.75W_0) = 1.5W_0 - W$ in cash. If the worst-case scenario occurs, and the stock price is halved, the investor will be left with $(W - 0.75W_0) + 1.5W_0 - W = 0.75W_0$.

[2] Egenter *et al.* find that the price time evolution becomes unrealistically periodic for a large number of investors. The periodicity seems related to the fixed 65-day quarter and is significantly diminished (as suggested by Markowitz in private communications) if the 65-day period begins on a different date for each investor.

In the first paper quoted earlier, the authors simulated a single stock and further limited the bid-offer decision to a ternary choice:

1. Bid to buy one share.
2. Offer to sell one share.
3. Do nothing.

Each agent had a collection of rules that described how he or she should behave (1, 2, or 3) in various market conditions. If the current market conditions were not covered by any of the rules, the default was to do nothing. If more than one rule applied in a certain market condition, the rule to act upon was chosen probabilistically according to the "strengths" of the applicable rules. The strength of each rule was determined according to the rule's past performance: rules that "worked" became "stronger." Thus, if a certain rule performed well, it became more likely to be used again.

The price was updated proportionally to the relative excess of offers over demands. In Arthur *et al.* (1997), the rules were used to predict future prices. The price prediction was then transformed into a buy-sell order through the use of a constant absolute risk aversion (CARA) utility function. The use of CARA utility might have been an unfortunate choice, as it leads to demands that do not depend on the investor's wealth (in the LLS model and in following studies this dependence is found to be quite crucial).

The heart of the Arthur *et al.* dynamics is in the part of each rule that describes the market conditions in which the rule is to be applied. In particular, this differentiates between fundamental rules and technical rules, and allows the study of their relative strength in various market regimes. For instance, a fundamental rule may require a market condition of the following type:

$$\text{Dividend}/\text{Current Price} > 0.04$$

in order to be applied. A technical rule may be triggered if the market fulfills a condition of the following type:

$$\text{Current price} > \text{10-Period moving average of past prices}$$

The rules undergo a genetic dynamics: the weakest rules are substituted periodically by copies of the strongest rules and all the rules undergo random mutations (or even versions of sexual crossovers: new rules are formed by combining parts from two different rules). The genetic dynamics of the trading rules represents investors' learning: new rules represent new trading strategies. Investors examine new strategies, and adopt those that tend to work best. The main results of this model are as follows:

1. *For a few agents and small dividend changes*:
   - The price converges toward an equilibrium price which is close to the fundamental value.

- Trading volumes are low.
- There are no bubbles, crashes, or anomalies.
- Agents follow a small number of homogenous simple fundamentalist rules.

2. *For a large number of agents and generic conditions*:

- There is no convergence to an equilibrium price, and the dynamics is complex.
- The price displays occasional large deviations from the fundamental value (bubbles and crashes).
- Some of these deviations are triggered by the emergence of collectively self-fulfilling agent price-prediction rules.
- The agents become heterogeneous (adopt very different rules).
- Trading volumes fluctuate (large volumes correspond to bubbles and crashes).
- The rules evolve over time to more and more complex patterns, organized in hierarchies (rules, exceptions to rules, exceptions to exceptions, etc.).
- The successful rules are time dependent: a rule that is successful at a given time may perform poorly if reintroduced after many cycles of market coevolution.

The conclusions of Arthur *et al.* are powerfully expressed in their own words:

> Experiments with a computerized version of this endogenous-expectations market explain one of the more striking puzzles in finance: Standard theory tends to see markets as efficient, with no rationale for herd effects and no possibility of systematic speculative profit, whereas traders tend to view the market as exhibiting a "psychology," bandwagon effects, and opportunities for speculative profit. Recently, the traders' view has been justified by invoking behavioral assumptions, such as the existence of noise traders. We show, without behavioral assumptions that both views can be correct. A market of inductively rational traders can exist in 2 different regimes:
>
> ... a simple regime which corresponds to the rational-expectations equilibrium of the efficient market literature ... [and] a complex regime in which rich psychological behavior emerges. Technical trading emerges as do temporary bubbles and crashes.

Similar herding emergence that is not induced by behavioral assumptions is exhibited in the simple model of Shnerb, Louzon, Bettelheim, and Solomon (1999).

8.5. LUX'S INTERMITTENT FLUCTUATIONS INDUCED BY TRADERS DYNAMICS

The Lux and Marchesi (1999) microscopic simulation, following the model introduced in Lux (1995, 1996), proposes to model in one endogenous dynamic framework the three stochastic properties of the price dynamics, which appear to be the most important generic features of the stock markets:

1. *The random walk character of the price fluctuations*. The lack of predictability of the market prices is at odds with some of the traditional models but it is intimately related to the efficient market hypothesis. A more precise mathematical formulation ("unit root property") implies the time stationarity of the stochastic distribution of the differences of logs of prices. The technical proof of this property is called the Dickey-Fuller test.

2. *The heavy tail distribution of returns*. The actual measurements of short time (less than a few weeks) returns reveal that the large returns (both positive and negative) are significantly more probable than predicted by the standard normal (Gaussian) distribution.[3] This has both theoretical and practical implications (inducing modifications in the Black-Scholes formula and modifications in the estimation of the value at risk of financial institutions). Analyzing and explaining this phenomenon in detail might give clues regarding the actual microscopic dynamics underlying stock markets.

3. *The clustering of volatility*. Volatility presents significant time autocorrelation. This effect can be formally described by ARCH/GARCH (generalized autoregressive conditional heteroscedasticity) processes.

All three of these phenomena emerge in the Lux model as soon as one assumes that in addition to the fundamentalists there are also chartists in the model. Lux and Marchesi (1999) further divide the chartists into optimists (buyers) and pessimists (sellers). The market fluctuations are driven and amplified by the fluctuations in the various populations: chartists converting into fundamentalists, pessimists into optimists, and so on.

In the Lux and Marchesi model, the stock's fundamental value is exogenously determined. The fluctuations of the fundamental value are inputted exogenously as a white noise process in the logarithm of the value (with a given time-constant variance). The market price is determined by investors' demands and by the market clearance condition.

Lux and Marchesi consider three types of traders: fundamentalists, optimists, and pessimists.

Fundamentalists look less at the current trends in the market and more at the fundamental value of the stock. The *fluctuations of this fundamental value* are inputted exogenously as a white noise process in the logarithm of the value (with a given time-constant variance). The Fundamentalists buy or sell amounts of shares proportional to the difference between the actual

[3] See Mandelbrot (1963a, 1963b), Fama (1963, 1965a), Roll (1968), Teichmoeller (1971), Officer (1972), Mantegna (1991), Zajdenweber (1994), Mantegna and Stanley (1995), and Cont, Potters, and Bouchaud (1997).

price and this fundamental price. The *actual price* in turn is the result of the buy/sell operations of the agents and is assumed to fluctuate proportionally to the excess demand.

Chartists look more at the present trends in the market price rather than at fundamental economic values; the chartists are divided into *optimists* (they buy a fixed amount of shares per unit time) and *pessimists* (they sell shares).

8.5.1 Transitions between Groups

Transitions between the three groups happen with probabilities depending on the market dynamics and on the present numbers of traders in each of the three classes.

The transition probabilities of chartists depend on the majority opinion (through an "opinion index" measuring the relative number of optimists minus the relative number of pessimists) and on the actual price trend (the current time derivative of the current market price), which determines the relative profit of the various strategies.

The fundamentalists behavior is determined by the belief that the price will eventually revert to the perceived fundamental value in the not too distant future. Excess profits, then, solely consist of the percentage deviation between the fundamental and current prices, discounted by the fact that the fundamentalists expect a certain time to pass before the reversal to the fundamental value takes place and their expected gains are realized (as opposed to the chartists who expect immediate profits).

The fundamentalists decide to turn into chartists if the profits of the later become significantly larger than their own, and vice versa (the detailed formulae used by Lux and Marchesi are inspired from the exponential transition probabilities governing the statistical mechanics of physical systems).

8.5.2 Results

1. No long lasting deviations between the current market price and the fundamental price are observed.

2. The deviations from the fundamental price, which do occur, are unsystematic.

3. In spite of the fact that the variations of the fundamental price are assumed to be normally distributed, the variations of the current price (the market returns) are not. In particular the returns exhibit

- A frequency of extreme events that is higher than expected for a normal distribution (a technical measure of it is the scaling expo-

nent, which Lux and Marchesi find to be 2.64). The authors emphasize the amplification role of the market that transforms the input normal distribution of the fundamental value variations into a leptokurtotic (heavy-tailed) distribution of price variation, which is encountered in the actual financial data.
- Clustering of volatility.[4]

The authors explain the volatility clustering (and as a consequence, the leptokurtoticity) by the following mechanism. In periods of high volatility, the fundamental information is not very useful to ensure profits, and a large fraction of the agents become chartists. The opposite is true in quiet periods when the actual price is very close to the fundamental value. The two regimes are separated by a threshold in the number of chartist agents. Once this threshold is approached (from below), large fluctuations take place that further increase the number of chartists. This destabilization is eventually dampened by the energetic intervention of the fundamentalists when the price deviates too much from the fundamental value. The authors compare this temporal instability with the on-off intermittence encountered recently in certain physical systems. According to Egenter *et al*. (1999), the fraction of chartists in the Lux and Marchesi model goes to zero as the total number of traders goes to infinity, when the rest of the parameters are kept constant.

In conclusion, the Lux model presents some of the main stochastic properties of the stock prices, in particular, the phenomena of heavy tails and clustered volatility. The model traces these phenomena to the positive feedback loop that relates the migration between the chartist and fundamentalist populations to the market price fluctuations. In the LLS model, a similar mechanism relates the wealth associated with various strategies to the price dynamics.

8.6. THE HERDING MODEL OF BAK, PACZUSKI, AND SHUBIK

The MS model of Bak, Paczuski, and Shubik (1997) attempts to explain large market fluctuations by the herding behavior of individual investors. Some of their investors imitate one another and eventually generate large herds of investors behaving in a similar way. The market fluctuations are then significantly larger that what would be expected from a population of independent heterogeneous investors. The authors assume that each trader can either own one share of the stock or none; the money not invested in

[4] The technical parameter related to it is the Hurst exponent, which Lux and Marchesi measure to be 0.85.

stocks earns a constant riskless interest rate. There are two main types of traders: rational traders, who look at the fundamental values of the company, and noise imitators, who ignore the actual performance of the company and mainly look at the market trends and at the behavior of other traders.

8.6.1 The Mechanics of the Market

- There is only one type of stock, of which each agent can own 0 or 1 shares.
- There are N agents and $N/2$ shares.
- Each shareowner i offers his share at a price $s(i)$, which is a real number from the interval [0, max]. Each potential buyer j (a trader with 0 shares) makes a bid at the price $b(j)$, which is a number from the same interval.

The microscopic simulation cycle consists of the following sequence of steps:

1. At each step one agent is chosen randomly to be the one who performs a market operation.
2. The agent observes the offers and bids presently outstanding.
3. The agent updates his bid/ask price ($s(i)$ or $b(j)$ as it may be the case) according to the market data (observed in step 2) and his strategy (discussed below).
4. If the new $s(i)$ is smaller than some outstanding $b(j)$'s (or its $b(i)$ is larger than some outstanding $s(j)$'s), a match is made. More precisely, the trader sells his share to the highest bidder (or buys from the lowest seller) and the new market price, p, is updated to that transaction value. Both i and his partner j in this transaction change status (the buyer becomes a potential seller and the seller a potential buyer).
5. Trader j, which was i's partner in the transaction, selects a value $s(j)$ (or $b(j)$ as it may be the case) as his new selling (buying) price. If the advertising of this new price by j leads to a new transaction, the process is repeated from step 4. If not, one returns to step 1, and a new a random agent is selected for trading.

The authors present a variety of models characterized by the various procedures traders employ in making their bid/ask decisions.

8.6.2 A Fundamentalist Trader Market

The benchmark model is one in which all the agents are "rational" fundamentalists. In particular, they choose their prices according to their utility function and the given dividend distribution, which is fixed. The buying $b(j)$ (and respectively selling $s(j)$) prices for all the agents are therefore fixed throughout the simulation in this version of the model. By choosing the distribution of risk aversion in the population appropriately, the authors set the distribution of agents' buying prices as uniform in an interval that partially overlaps with the (uniformly distributed, higher mean) interval of selling prices. In this setting, the market comes to rest after some time: buyers and sellers "phase separate" into nonoverlaping regions. Thus, as in many other rational trader models, there is no trade after the equilibrium is reached. However, as opposed to the traditional models, the actual equilibrium price depends on the particular history of each simulation run.

8.6.3 Independent Noise Traders Market

The noise traders model assumes that the traders' decisions ignore dividends and other fundamental data. Initially the selling prices are distributed uniformly in the interval [max/2, max] and the bids are uniformly distributed in the interval [0, max/2]. At each update step 3, the price $s(j)$ (respectively $b(j)$) changes randomly by one unit, either up or down, with equal probability. If this move leads to a matching opportunity, step 4 is initiated, a sale takes place, and the market price $p(t)$ is updated. The emerging new buyer and seller then choose new bidding prices in the interval $[0, p(t)]$ (respectively, $[p(t), \max]$) according step 7.

This model is formally equivalent with a field-theoretical model that has been studied extensively by physicists. In particular, the distribution of the price variations can be estimated analytically (by quantum field theory techniques) and it respects a scaling (power distribution) law with a Hurst exponent $H = \frac{1}{4}$, in contrast with $H = \frac{1}{2}$, which characterizes usual normal noise (Gaussian random walk).

8.6.4 Noise Traders with Imitating Behavior

The main claim in the Bak Paczuski Shubik paper refers to the mimicking noise traders. The authors relate them to the emergence of large-scale organization. In this version of the model, at step 5, after each transaction, the agents involved in it choose their new selling (or buying) prices by copying the corresponding price(s) of another (randomly chosen) trader. The actual runs show that the price variations in this "urn model" are much more dramatic than in the other versions of the model, and they look more realistic.

8.6.5 Mixing Fundamental Value Traders with Imitating Noise Traders

The authors question what happens if a market contains both fundamentalists and imitating noise traders. Would the rational players discipline the noise traders? Would they exploit them? Or maybe the noise traders will be able to "spoil" the market to a degree to which the rational fundamentalist behavior becomes suboptimal?

A simulation with 2% rational traders led to the following scenario: At the beginning, the two types of agents had similar market behavior. However, at a certain stage, the noise traders started increasing the price and produced a "bubble": a "crowd" or "herding" effect. This led the rational traders to sell their stocks, and the price eventually returned to the fundamental level. The authors find that a proportion of 20% rational traders is enough to keep the price confined most of the time in the range of the fundamental value. In this case, deviations from the fundamental value are brief and small.

8.7. THE MARKET ECOLOGY OF FARMER

Farmer's (1998) MS model aims to describe the market dynamics as an "ecology" of trading strategies. At large time scales, the changes in the market dynamics are explained endogenously by the old strategies (which are no longer successful) being replaced by new ones (just as more fit biological species replace less fit species). On short time scales, the price fluctuations are induced by buyers who drive the prices up and sellers that push them down (according to a market impact function). If too many agents try to exploit one successful niche, the advantages of that niche are diminished, and a new strategy may emerge as more successful. Thus, one has an interplay between the short time dynamics of the prices and the long time dynamics of the strategies: the implementation of the various investing strategies by the agents determines the price dynamics, which, in turn, applies selective pressure on the strategies. Similarly to the biological evolution of species, a diversity of coevolving trading styles emerges. But, in contrast to biology, this evolution takes not millions of years but at most a few decades.

The paper starts from a set of simple postulates that lead to an analytical formula for the market impact function, which relates the new price to the order size and to the old price. The next main result generalizes the model to more than one strategy and expresses the market returns in terms of the previous market history, the various investing strategies, and of the wealth associated with each strategy. The determination of these investing strategies and of their associated wealth by the price history on one hand and the generation of price by the implementation of these strategies on the other hand close a feedback loop, which endogenously drives the dynamics of the model. In particular, the system is similar to an ecology, in as far as the wealth levels associated with the various strategies are analogous to the population sizes associated with various species. The effective equations governing ecological systems and expressing the competition, procreation, death, and symbiosis relationships within and between the species were introduced long ago in the field of mathematical biology under the name of the Lotka-Volterra equations (Lotka, 1925; Volterra, 1926).

This mathematical analogy to ecology dynamics is developed in a few particular examples in which the species are various investor groups: value investors, investors using purely temporal strategies (e.g., January effect), trend followers, and so on. The interactions between these various "species" and the resulting "ecology" dynamics are studied by Farmer in a series of microscopic simulations. In particular, he finds that arbitrageurs can "feed on" (exploit) populations, which are obliged (by tax or other financial constraints) to follow predictable temporal strategies. It is also systematically found that value investing leads to negative autocorrelations in log returns while trend following leads to positive autocorrelations. A relationship between the inhomogeneity of strategies within a population and the excess volatility of prices is also observed in various contexts.

In the last chapter of this paper, the effect of the (co-)evolution of strategies is considered. It is emphasized that the capital associated with each strategy varies in time: profits are reinvested, strategies change in popularity and new strategies are discovered, leading to changes in the market dynamics. Farmer compares this with the biological evolution of species driven by natural selection mechanisms. The feedback between the longer scales at which this happens and the shorter scale daily dynamics is quite complex. According to Farmer, this market evolution is the dynamical mechanism enforcing market efficiency. Knowledge about its time scale is therefore relevant for the estimation of the time spans for which arbitrage opportunities might stay open before being exploited. Again it turns out that the relevant dynamics is described by a Lotka-Volterra (LV) set of equations.

It has been previously shown in a series of papers (Biham, Malcai, Levy, and Solomon, 1998; Solomon and Levy, 1996) that such LV systems generically lead to effects encountered in real markets: Pareto (power-law) wealth and truncated Lévy (stable-Paretian) return (Solomon, 1998). In fact, a relation between these two effects was predicted by Levy and Solomon (1997) and quite precisely validated by the data Levy (1998b). The LV interactions lead to the emergence of herding and collective adaptive behavior that, in turn, ensures the (financial) survival and global financial growth of the system. In fact, the emergent adaptive economy is so resilient that is thrives even in conditions in which an average macroscopic view of the market would predict extinction/bankruptcy (Shnerb, Louzon, Bettelheim and Solomon, 1999).

In the conclusion of the Farmer (1999) paper, the possibility is considered of introducing human behavioral characteristics in the trading strategies. Finally, the general relation between the efficient market concept and the evolutionary dynamics of ever-increasing complex patterns is discussed.

8.8. THE EFFECTS OF THE NUMBER OF INVESTORS

A common theme to several papers is the question: How do the market dynamics depend on the number of investors in the market? The number of participants in Wall Street has grown significantly during the past century, and emerging markets are currently growing rapidly. There are two different views regarding the effect of the number of investors:

1. The larger the number of investors in the market, the smaller price fluctuations should be. This view is motivated by the notion that more participants lead to more "cancellation of noise," as mathematically stated by the central limit theorem.
2. The price fluctuations do not depend significantly on the number of market participants.

Several of the models mentioned here have explicitly been investigated in this aspect. For the LLS model, Hellthaler (1995) and Kohl (1997) found a nearly periodic price evolution as N, the number of investors, becomes very large. Egenter *et al*. (1999) found similar periodic crashes in the Kim-Markowitz model, while in the Lux model they found that the proportion of noise traders (responsible for the large price fluctuations) becomes very small as the total number of traders is increased. Busshaus and Rieger (1999) mention that fluctuations in their model vanish for N going to infinity. Farmer gives an equation showing that in his model fluctuations decrease as the reciprocal square root of N, in accordance with the central limit theorem.

Thus, most of the models support view 1. It would be interesting to have empirical studies on actual markets, which have significantly grown during a short period. Of course, it would not be possible to keep all other parameters constant in such an empirical investigation.

If view 1 is correct, then the empirically observed significant price fluctuations in actual markets could mean that markets are always dominated by a rather small number of important players, while the large mass of small investors has little impact. This is, in fact, quite compatible with the large inequalities in the empirical wealth distribution among investors (Levy, 1998b; Levy and Solomon, 1997).

9

PROSPECT THEORY, ASSET PRICING, AND MARKET DYNAMICS

9.1. INTRODUCTION

Expected utility theory (EUT) is one of the pillars of modern economics and finance. While it is generally accepted as a *normative* model of rational choice, it has been challenged as an adequate *descriptive* model of human behavior. Nearly five decades ago Markowitz (1952b) postulated that investors make their decisions based on *change* in wealth rather than *terminal* wealth, which is in contradiction to the EUT. Over the years, many experiments have revealed systematic violations of the axioms of EUT.[1] The main quantitative theory suggested as an alternative for EUT in describing investors' behavior is prospect theory.

In a breakthrough article, Kahneman and Tversky (1979) cast the main experimental findings regarding choice under uncertainty into a unifying

[1] See Preston and Barrata (1948), Mosteller and Nogee (1951), Allais (1953), Edwards (1953, 1954), Swalm (1966), Kahneman and Tversky (1979), Thaler (1985), Thaler and Johnson (1990), Kahneman, Knetsch, and Thaler (1990), Tversky and Kahneman (1992), Thaler (1993), Thaler (1994), and Benartzi and Thaler (1999). For a review of the main empirical and experimental findings regarding deviations from the expected utility theory, see Chapters 2 and 4.

theory, which they coin prospect theory (PT). The three basic elements of PT are as follows:

1. Investors make decisions based on *change* in wealth, x, rather than on total *terminal* wealth, $w + x$.

2. Investors are risk averse when considering prospects with only positive outcomes but are risk seeking when considering prospects with only negative outcomes. Investors' behavior can be characterized by the maximization of the expected value of an S-shaped value function, $V(x)$, which is convex for negative x but concave for positive x.

3. Investors systematically distort probabilities and base their decisions on their subjective probabilities, rather than on the objective probabilities. The distortion is such that low probabilities are subjectively overestimated.

How do these behavioral elements influence asset pricing? What effects, if any, do these elements have on the market dynamics? This chapter is devoted to the examination of these two questions. In the next section we review the experimental findings regarding the shape of the value function and the transformation of objective probabilities. In Section 9.3 we theoretically analyze the implications of PT to asset allocation and to equilibrium pricing. In Section 9.4 we study the effects of the behavioral elements of PT in the context of a dynamical market model. We employ the LLS microscopic simulation (MS) model in order to compare the price dynamics in a market with PT investors with the dynamics of a market of expected utility maximizers who do not distort probabilities. We present our main conclusions in Section 9.5.

9.2. EXPERIMENTAL ESTIMATES OF THE VALUE FUNCTION AND THE DISTORTION OF PROBABILITIES

9.2.1 The Value Function

Tversky and Kahneman (1992) conducted a comprehensive experimental study of decision making under uncertainty. Their experimental findings lead them to suggest a mathematical form for the S-shaped value function, $V(x)$. Tversky and Kahneman find that individuals' preferences are homogeneous—that is, if the certainty equivalent of a risky prospect $\tilde{x}$ is $\text{CE}(\tilde{x})$, then the certainty equivalent of a corresponding prospect $k\tilde{x}$ (where all the outcomes of the original prospect $\tilde{x}$ are multiplied by a constant k) is just $k\text{CE}(\tilde{x})$. The unique value function that is consistent with homogeneous preferences is a power function (see Tversky and Kahneman, 1992, p. 309). This leads Tversky and Kahneman to suggest that the value function is a power function. Since Tversky and Kahneman make the distinction between positive prospects[2] and negative prospects, they suggest the following two-part power value function:

$$V(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases} \tag{9.1}$$

where x is the change in wealth, and α, β, and λ are constants satisfying $0 < \alpha < 1$, $0 < \beta < 1$, and $0 < \lambda$. This value function satisfies $V'(x) > 0$ for all $x \neq 0$, $V''(x) > 0$ for $x < 0$, and $V''(x) < 0$ for $x > 0$, and is therefore S-shaped (see Figure 9.1). The convexity of the value function for negative x, and its concavity for positive x, is in accordance with the findings of risk seeking for negative prospects ($x < 0$), and risk aversion for positive prospects ($x > 0$). From their experimental data, Tversky and Kahneman estimate the typical parameters of this value function as $\alpha = 0.88$, $\beta = 0.88$, and $\lambda = 2.25$ (see Tversky and Kahneman, 1992, pp. 311–312).

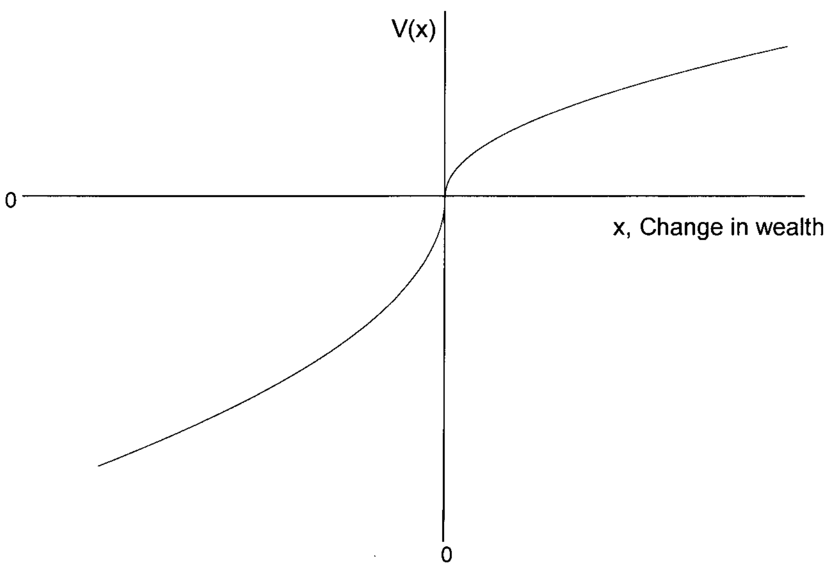


FIGURE 9.1 The S-shaped value function.

[2] A positive (negative) prospect is a prospect in which all possible outcomes are positive (negative). A mixed prospect is a prospect in which both positive and negative outcomes are possible.

9.2.2 Probability Transformation

In the original PT (Kahneman and Tversky, 1979), the outcomes of prospects are discrete, and the subjective probability weight assigned to each outcome is obtained by a transformation of the outcome's objective probability, p_i. As seen in Chapter 2, this framework leads to possible violations of first degree stochastic dominance (FSD), implying that people may prefer less over more. Quiggen (1982), Yaari (1987), Allais (1988), and Tversky and Kahneman (1992) extend PT to allow for prospects with a continuous distribution of outcomes in such a way that does not contradict FSD. Instead of the separate transformation of discrete probabilities, they suggest a transformation of the cumulative distribution function. Namely, if $F(x)$ is the objective cumulative distribution function, individuals subjectively transform the cumulative distribution to $T(F(x))$, where the transformation T satisfies $T(0) = 0$, $T(1) = 1$, and $T'(\cdot) > 0$. Tversky and Kahneman (1992) suggest the following functional form for the transformation T:

$$T(F) = \frac{F^\gamma}{\left(F^\gamma + (1 - F)^\gamma\right)^{1/\gamma}} \tag{9.2}$$

Figure 9.2a graphically describes this transformation. Notice that this transformation satisfies the conditions $T(0) = 0$, $T(1) = 1$. It can also be shown that the transformation in Eq. (9.2) is monotonic ($T'(\cdot) > 0$) for any value of γ in the range $0.279 < \gamma < 1.0$.[3] Tversky and Kahneman and Camerer and Ho (1991) estimate the parameter γ as about 0.6.[4] This transformation implies that low-probability events are overestimated, while high-probability events are underestimated. To see this, recall that the objective probability density is given by $F'(x) \equiv f(x)$ and that the subjective (distorted) probability density is given by $(TF(x))'$. The subjective probability density can also be written as

$$(TF(x))' = \frac{dT}{dF}F'(x) = \frac{dT}{dF}f(x) \tag{9.3}$$

Thus, the subjective probability density is given by the objective probability density, $f(x)$, multiplied by the factor dT/dF. For the transformation in Eq. (9.2), this factor is depicted in Figure 9.2b. Notice that the horizontal

[3]

$$T'(F) = \frac{\gamma F^{\gamma-1}\left(F^\gamma + (1-F)^\gamma\right)^{1/\gamma} - F^\gamma\left(F^\gamma + (1-F)^\gamma\right)^{(1/\gamma-1)}\left(F^{\gamma-1} - (1-F)^{\gamma-1}\right)}{\left(F^\gamma + (1-F)^\gamma\right)^{2/\gamma}}$$

Numerical analysis shows that if $\gamma > 0.279$, this expression is positive for any value of F in the relevant range $0 \leq F \leq 1$.

[4] Tversky and Kahneman estimate γ separately for gains and for losses. For gains they estimate $\gamma = 0.61$, while for losses they estimate $\gamma = 0.69$.

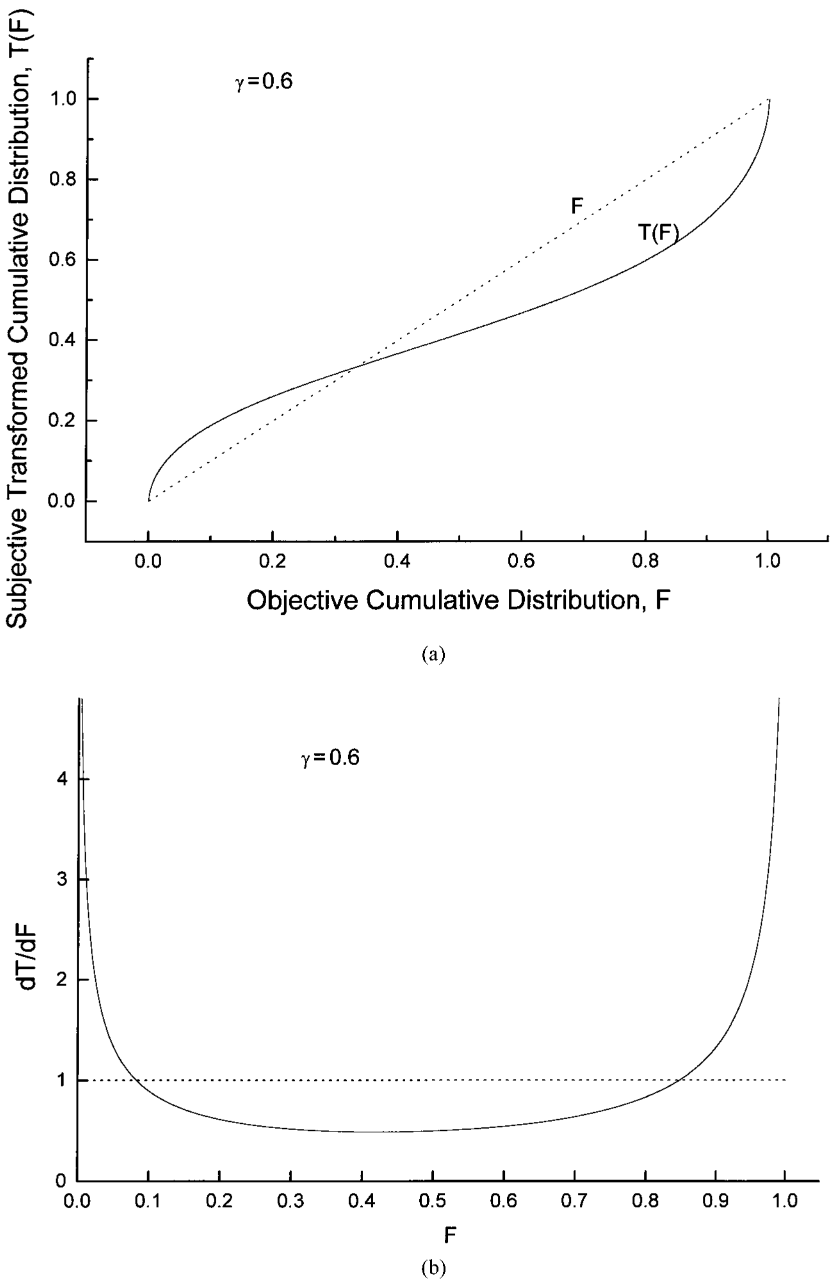


FIGURE 9.2 (a) Transformation of the cumulative probability distribution. (b) The relative weighting of probabilities.

line at 1 is the derivative of the line describing F in Figure 9.2a. The U-shaped line in Figure 9.2b is the derivative of the line describing $T(F)$ as given in Figure 9.2a, which is calculated by Eq. (9.3). It is evident from this figure that the subjective probabilities are overestimated at the extreme tails of the distribution (where the objective probability is typically small) and are underestimated at the central part of the distribution (where the objective probability is typically large).

9.3. IMPLICATIONS OF PROSPECT THEORY TO ASSET ALLOCATION AND EQUILIBRIUM PRICING

9.3.1 Rates of Return versus Dollar Returns

Prospect theory was inspired by experiments in which individuals were asked to state their certainty equivalents for risky prospects. In these experiments, the outcomes of the prospects were given in dollar terms. As a consequence, the theory was formulated in terms of change in wealth *in dollars*. While this framework is the most convenient for experiments in which certainty equivalents are determined, it is unnatural for market conditions in which the various payoffs are generally compared in terms of percent, as opposed to dollar values. To find out how much an investor with a PT value function would invest in a risky stock with a given distribution of *rates of return*, one must first translate the rates of return into changes in wealth in dollar terms and then maximize $EV(\tilde{x})$. Specifically, if an investor has an initial wealth $\$W_0$ and invests a proportion p of her money in a stock yielding a stochastic rate of return $\tilde{R}$ and a proportion $(1-p)$ in a riskless asset yielding a rate of return r, then her terminal wealth is given by

$$\tilde{W}_1 = W_0(1-p)(1+r) + W_0 p(1+\tilde{R}) \tag{9.4}$$

and the change in wealth, in dollar terms will be

$$\tilde{x} \equiv \tilde{W}_1 - W_0 = W_0(1-p)r + W_0 p\tilde{R} \tag{9.5}$$

As the change in wealth is a function of the initial wealth W_0, the initial wealth will generally affect the investment decision in the PT framework. Thus, even though the initial wealth is not explicitly relevant in the original PT terminology, in a stock market scenario, in which outcomes are given in terms of rates of return, the initial wealth is generally relevant to the investment decision. While the initial wealth does affect the optimal diversification between the stock and the bond for general S-shaped value functions, this is not so for the Tversky and Kahneman power value function, given by Eq. (9.1) for the case $\alpha = \beta$. To see this, let us write down the power value function for change in wealth induced by the stock-bond portfolio given in Eq. (9.5):

$$V = \begin{cases} \left[W_0(1-p)r + W_0 p\tilde{R}\right]^{\alpha} & \text{if } \left[W_0(1-p)r + W_0 p\tilde{R}\right] > 0 \\ -\lambda\left[-\left(W_0(1-p)r + W_0 p\tilde{R}\right)\right]^{\beta} & \text{if } \left[W_0(1-p)r + W_0 p\tilde{R}\right] < 0 \end{cases} \tag{9.6}$$

where the term in the square brackets is the change in wealth. Let us denote the rate of return on the stock that makes the change in wealth equal to 0 by $R_0(p)$. $R_0(p)$ is given by

$$W_0(1-p)r + W_0 p R_0(p) = 0$$

or

$$R_0(p) = \frac{-(1-p)r}{p} \tag{9.7}$$

If the stock's rates of return are distributed according to some probability density function $f(R)$, the expectation of the value function is then the following function of the investment proportion in the stock, p:

$$EV = \int_{R_0(p)}^{\infty} [W_0(1-p)r + W_0 pR]^{\alpha} f(R)\, dR - \lambda \int_{-\infty}^{R_0(p)} (-[W_0(1-p)r + W_0 pR])^{\beta} f(R)\, dR \tag{9.8}$$

Tversky and Kahneman (1992) empirically estimate $\alpha = \beta = 0.88$. Levy (1999b) shows that $\alpha = \beta$ is the only parameterization of the value function in Eq. (9.1), which is consistent with the key PT concept of loss aversion: the notion that individuals are more sensitive to losses than they are to gains of the same magnitude. Thus, the equality of the parameters α and β is both empirically and theoretically supported.[5] Therefore, in what follows we focus on the relevant case $\alpha = \beta$. For the $\alpha = \beta$ case, Eq. (9.8) can be rewritten as

$$EV = W_0^{\alpha}\left[\int_{R_0(p)}^{\infty} [(1-p)\mathrm{r} + pR]^{\alpha} f(R)\, dR - \lambda \int_{-\infty}^{R_0(p)} (-[(1-p)r + pR])^{\alpha} f(R)\, dR\right] \tag{9.9}$$

[5] The homogeneity of preferences with respect to mixed prospects, which is empirically observed by Tversky and Kahneman (1992), also implies the equality, $\alpha = \beta$, see Levy (1999b).

As can be seen from Eq. (9.9), in the case of the power function with $\alpha = \beta$, the initial wealth is a common factor that does not affect the optimal investment proportion in the stock. Thus, the power value function with $\alpha = \beta$ implies that given a stock with some return distribution and a riskless asset, an individual will invest the same *proportion* of her wealth in the stock, regardless of her wealth. This property constitutes further empirical support for the power value function suggested by Tversky and Kahneman, because it has been empirically and experimentally documented in various different studies that individuals' asset allocation (in terms of proportions of their wealth) is roughly independent of the level of wealth; (see Arrow (1971), Gordon, Paradis, and Rorke (1972), Friend and Blume (1975), Kydland and Prescott (1982), Kroll, Levy, and Rapoport (1988a), Levy (1994); for a review see Chapter 3.)[6]

9.3.2 Optimal Asset Allocation

In this section we investigate the diversification policy implied by the Tversky and Kahneman power value function. We restrict ourselves here to diversification between one risky stock, which serves as a proxy for the market portfolio, and one riskless bond (for the implications of PT on diversification across risky assets see H. Levy, 1999).

The diversification policy of a Tversky and Kahneman investor is rather extreme: for a given *arbitrary* rate of return distribution, the investor will typically be either fully invested in the bond or fully invested in the stock. The crossover between these two states of full investment in the bond and full investment in the stock is sharp: a small change in one of the return distribution parameters can lead to a shift from full investment in one asset to full investment in the other. Notice, however, that this does not imply that investors flip-flop between stocks and bonds. Instead, as we shall describe in the next section, in *equilibrium* the stock (which serves as a proxy for the market portfolio) will be priced such that the parameters of the rate of return distribution are at the crossover point. Let us elaborate.

To understand the sharp crossover between the two states of investment in the bond and investment in the stock, let us first examine a very simplified case in which the stock can yield only two possible rates of return. We later generalize the result to other rate of return distributions. In the simplified case with two equally probable rates of return, the rates of return can be denoted by $\mu - \sigma$ and $\mu + \sigma$, where μ is the mean rate of return and σ is the standard deviation. Let us begin by assuming that $r = 0$ and that $\mu - \sigma < 0$, $\mu + \sigma > 0$. If an investor invests a proportion

[6] In the framework of the expected utility theory, this property is known as constant relative risk aversion (CRRA), and it implies the myopic utility function $U(W) = W^{1-\alpha}/(1-\alpha)$ theoretically advocated by Merton and Samuelson (1974).

p in the stock, the value function is

$$V = \begin{cases} [p(\mu + \sigma)]^{\alpha} & \text{probability } 1/2 \\ -\lambda[-p(\mu - \sigma)]^{\alpha} & \text{probability } 1/2 \end{cases} \tag{9.10}$$

which is a special case of Eq. (9.6) with $r = 0$ and $\alpha = \beta$. Notice that as the initial wealth does not affect the optimal diversification in the $\alpha = \beta$ case, it is simply taken here as $W_0 = 1$, and hence rates of return represent change in wealth. The expectation of the value function is

$$EV = \frac{1}{2}[p(\mu + \sigma)]^{\alpha} - \frac{1}{2}\lambda[-p(\mu - \sigma)]^{\alpha} \tag{9.11}$$

or

$$EV = p^{\alpha}\left[\frac{1}{2}(\mu + \sigma)^{\alpha} - \frac{1}{2}\lambda(\sigma - \mu)^{\alpha}\right] \tag{9.12}$$

Equation (9.12) implies that if the term in the square brackets is positive, then the expected value is maximized by choosing $p \to \infty$, or borrowing and investing an infinite amount in the stock.[7] If the term in the square brackets is negative, the optimal investment proportion in the stock is 0.[8] Thus, in this example there is a discontinuous transition between zero investment in the stock and an infinite investment in the stock. This transition occurs at

$$(\mu + \sigma)^{\alpha} = \lambda(\sigma - \mu)^{\alpha} \tag{9.13}$$

or

$$\mu^{*} = \sigma\left(\frac{\lambda^{1/\alpha} - 1}{\lambda^{1/\alpha} + 1}\right) \tag{9.14}$$

where μ^{*} denotes the transition, or crossover, expected return. The optimal investment proportion in the stock as a function of the expected return μ is graphically depicted in Figure 9.3. This figure shows that the diversification policy of the Tversky and Kahneman investor is very different than the diversification policy of expected utility maximizers with

[7] As in most standard finance models, we assume that the risk-free rate is exogenous and constant. If the risk-free rate increases with the amount borrowed, the optimal investment proportion in the stock may be finite but still very large.

[8] We do not discuss short selling here.

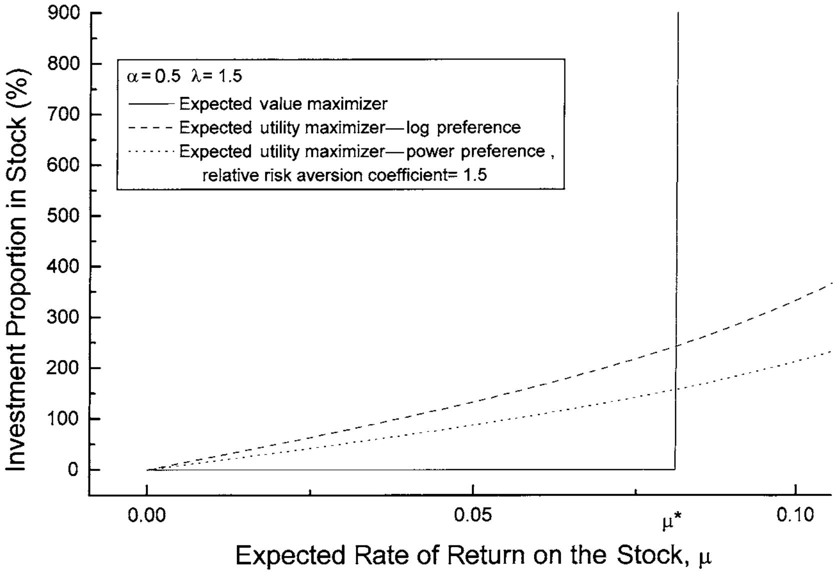


FIGURE 9.3 Optimal diversification ($\sigma = 0.2$).

typical utility functions.[9] While the investment proportion of typical expected utility maximizers varies smoothly and continuously as a function of the expected rate of return, the investment proportion of the PT expected value maximizer is characterized by a sharp kink at the crossover point μ^*.

The discontinuous crossover from the bond to the stock when the risk-free rate is 0 is not specific to the preceding return distribution with only two possible outcomes. Rather, it is a general property of the Tversky and Kahneman value function. For a general distribution of rates of return with density $f(R)$, the expectation of the value function in the $r = 0$ case is given by

$$EV = \int_0^{\infty} f(R)(pR)^{\alpha}\, dR - \lambda \int_{-\infty}^{0} f(R)(-pR)^{\alpha}\, dR \qquad (9.15)$$

[9] The expected utility is given by

$$EU(p) = \frac{1}{2}U[W_0(p(1+\mu-\sigma)+(1-p)] + \frac{1}{2}U[W_0(p(1+\mu+\sigma)+(1-p)]$$

(Recall that the risk-free rate is 0 in this example.) For the log utility function, the optimal investment proportion in the stock is given by $p^* = \mu/[(\sigma+\mu)(\sigma-\mu)]$. For the power utility function, $U(W) = W^{1-\alpha}/(1-\alpha)$, the optimal investment proportion in the stock is given by $p^* = (Z-1)/[(\sigma+\mu)+Z(\sigma-\mu)]$, where $Z = [(\sigma-\mu)/(\sigma+\mu)]^{-1/\alpha}$. (Notice that the two solutions converge for $\alpha \to 1$.)

where R is the rate of return on the stock and Eq. (9.15) is obtained as a special case of Eq. (9.9) with $r = 0$ and $W_0 = 1$. Eq. (9.15) can also be rewritten as

$$EV = p^{\alpha}\left[\int_0^{\infty} f(R)R^{\alpha}\, dR - \lambda\int_{-\infty}^{0} f(R)(-R)^{\alpha}\, dR\right] \tag{9.16}$$

If the term in the square brackets is negative, the optimal investment proportion in the stock is 0, and if this term is positive, the optimal investment proportion in the stock is infinite. Thus, if we start with some rate of return distribution $f(R)$ for which the term in the square brackets of Eq. (9.16) is negative, and we start shifting the entire distribution to the right, this term becomes less negative, until at some stage we reach a critical point where this term becomes zero. At this point investors increase their investment proportion in the stock from zero to infinity. The critical crossover point, for which the square brackets in Eq. (9.16) becomes zero, is given by

$$\int_0^{\infty} f(R)R^{\alpha}\, dR - \lambda\int_{-\infty}^{0} f(R)(-R)^{\alpha}\, dR = 0 \tag{9.17}$$

If the distribution of rates of return can be fully characterized by its mean and standard deviation (which is the case for the two-point, uniform, normal, and lognormal distributions, to name a few), Eq. (9.17) implies a relationship between the mean and the standard deviation of the distribution that must hold at the crossover point. For some distributions this relation can be derived analytically, while for others it must be solved numerically. For example, for a uniform return distribution the sharp crossover from the bond to the stock occurs at

$$\mu^* = \sqrt{3}\,\sigma\left(\frac{\lambda^{\frac{1}{\alpha+1}} - 1}{\lambda^{\frac{1}{\alpha+1}} + 1}\right) \tag{9.18}$$

(See Appendix 9.1.) For more realistic return distributions such as the normal or lognormal distributions, the crossover point can be found numerically by solving Eq. (9.17).

When the interest rate is different from 0, the crossover between investment in the bond to investment in the stock is not discontinuous, but it is still very sharp (when $r \neq 0$, the optimal investment proportion is found by maximizing the expected value given by Eq. (9.9) numerically). Figure 9.4 shows the optimal investment proportion in the stock as a function of its expected rate of return for a two-point return distribution with a standard deviation of 20% and the following value function parame-

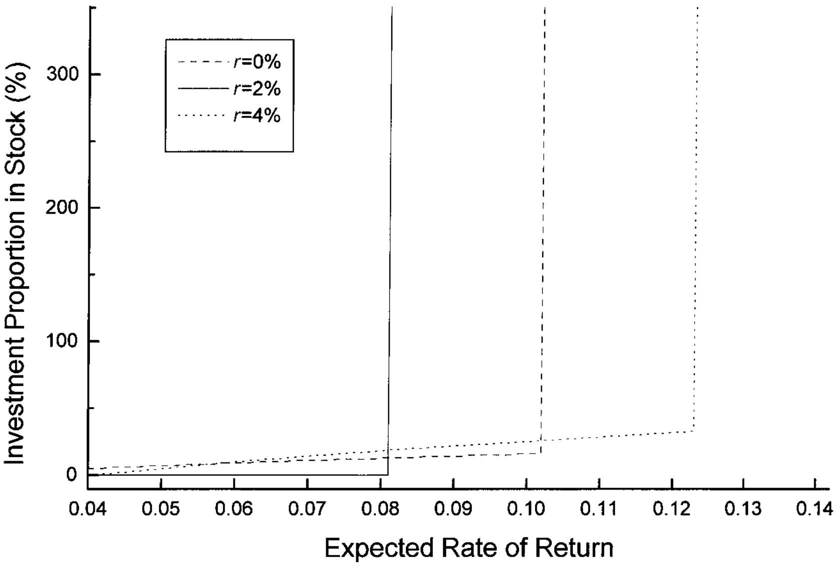


FIGURE 9.4 Optimal diversification, two-point return distribution.

ters: $\alpha = \beta = 0.8$, $\lambda = 2.0$. Three cases are shown: $r = 0\%$, $r = 2\%$, and $r = 4\%$. The crossover from the bond to the stock is very sharp, even in the last two cases in which the interest rate is different from 0. The kink in the optimal investment proportion is a robust property of the diversification policy implied by the Tversky and Kahneman value function: it holds both for various return distributions and for various parameters of the value function. Figure 9.5 shows the optimal investment proportion in the stock as a function of its expected return for normal and lognormal distributions of rates of return and for various value function parameters (the optimal investment proportion is found by maximizing Eq. (9.9) numerically).

9.3.3 Equilibrium Pricing

The equilibrium price of the stock (which is a proxy for the market portfolio) is determined by investors' portfolio optimization and by market clearance. Let us elaborate. Suppose that the market portfolio's end-of-period value is a stochastic variable $\tilde{V}_1$, and that the total aggregate wealth of investors at time 0 is W_0. (At this stage we assume that investors are homogeneous; this assumption is later relaxed.) The investors' investment proportion in the stock, p, determines the market portfolio's value at time 0. Namely, $V_0 = pW_0$ (otherwise the market is not cleared). Therefore, the

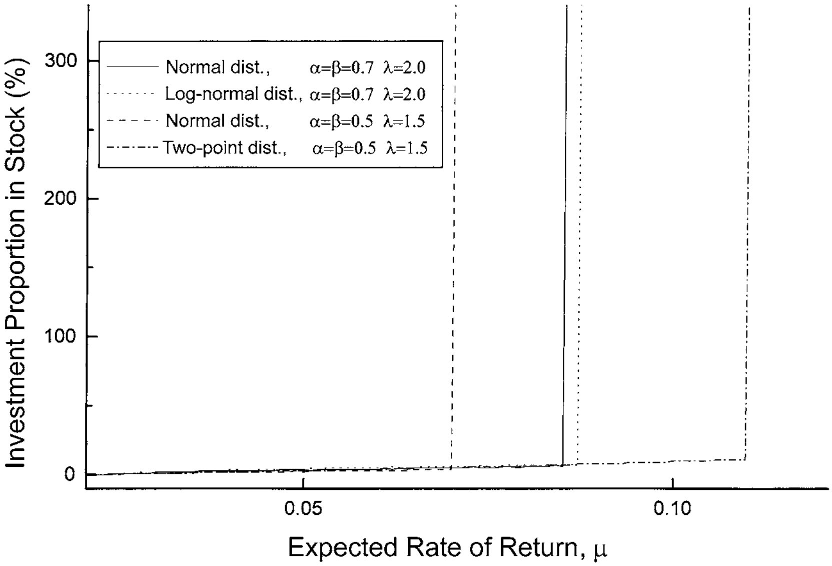


FIGURE 9.5 Optimal diversification: Different distributions and value function parameters ($r = 2\%$, $\sigma = 20\%$).

investment proportion p determines the market portfolio's rate of return distribution:

$$\tilde{R} = \frac{\tilde{V}_1}{V_0} - 1, \text{ or } \tilde{R} = \frac{\tilde{V}_1}{pW_0} - 1 \tag{9.19}$$

For a given end-of-period value distribution $\tilde{V}_1$, the investment proportion in the stock, p, determines the rate of return distribution, but on the other hand, the rate of return distribution determines the optimal investment proportion in the stock. For an expected utility maximizer, the optimal investment proportion in the stock is the proportion that maximizes the expected utility—that is, the investment proportion, p, which maximizes the expression:

$$EU = \int U[(1-p)r + pR]f(R)\,dR \tag{9.20}$$

For a PT expected value maximizer, the optimal investment proportion is the proportion that maximizes the expectation of the value function (as given in Eq. (9.9)). In both cases, for the market to be in equilibrium, the investment proportion in the stock must be self-consistent (i.e., it must generate a rate of return distribution that justifies this investment proportion as optimal). Thus, like in Lintner's version of the CAPM, the distribution of returns "adjusts and readjusts" until investors are at their optimal investment proportion, and the market is cleared simultaneously (see Lintner, 1965b, p. 598). Namely, in equilibrium there must be an equality between the market clearing investment proportion and optimal investment proportion. It is convenient to view both the market clearing investment proportion and the optimal investment proportion as a function of the expected rate of return, μ. The market clearance condition dictates the following relation between the expected rate of return and the investment proportion:

$$\mu = \frac{\bar{V}_1}{pW_0} - 1 \tag{9.21}$$

(which is obtained by taking the expectation of Eq. (9.19)). Alternatively, one can view the market clearing investment proportion as the following function of the expected rate of return:

$$p = \frac{\bar{V}_1}{(1 + \mu)W_0} \tag{9.22}$$

For given preferences (characterized either by an expected utility function or a PT value function), the expected return μ also implies an optimal investment proportion in the stock.[10] The equilibrium expected return, μ^*, is that for which the market clearing investment proportion (given by Eq. (9.22)) is equal to the optimal investment proportion (given by the optimization of Eq. (9.20) for expected utility maximizers and by the optimization of Eq. (9.9) for expected value maximizers). Graphically, μ^* is given by intersection of the line describing the market clearing investment proportion as a function of μ with the line describing the optimal investment proportion as a function of μ. Figure 9.6 shows the market clearing investment proportion given by Eq. (9.22) for the following parameters: $W_0 = 1$, $\bar{V}_1 = 1.5$.

The optimal investment proportion is calculated for two cases:

1. An expected utility (EU) maximizer with a log utility function
2. A prospect theory expected Value (EV) maximizer with $\alpha = \beta = 0.88$, $\lambda = 2.25$

(The EU optimal investment proportion is calculated by numerically maximizing the EU as given in Eq. (9.20) with $U(\cdot) = \log(\cdot)$. The PT optimal investment proportion is calculated by numerically maximizing the expected value as given in Eq. (9.9).) The risk-free real interest rate is

[10] For a given distribution of $\tilde{V}_1$, μ determines today's market portfolio value: $V_0 = \bar{V}_1/(1 + \mu)$. V_0, in turn, determines the entire rate of return distribution, as in Eq. (9.19).

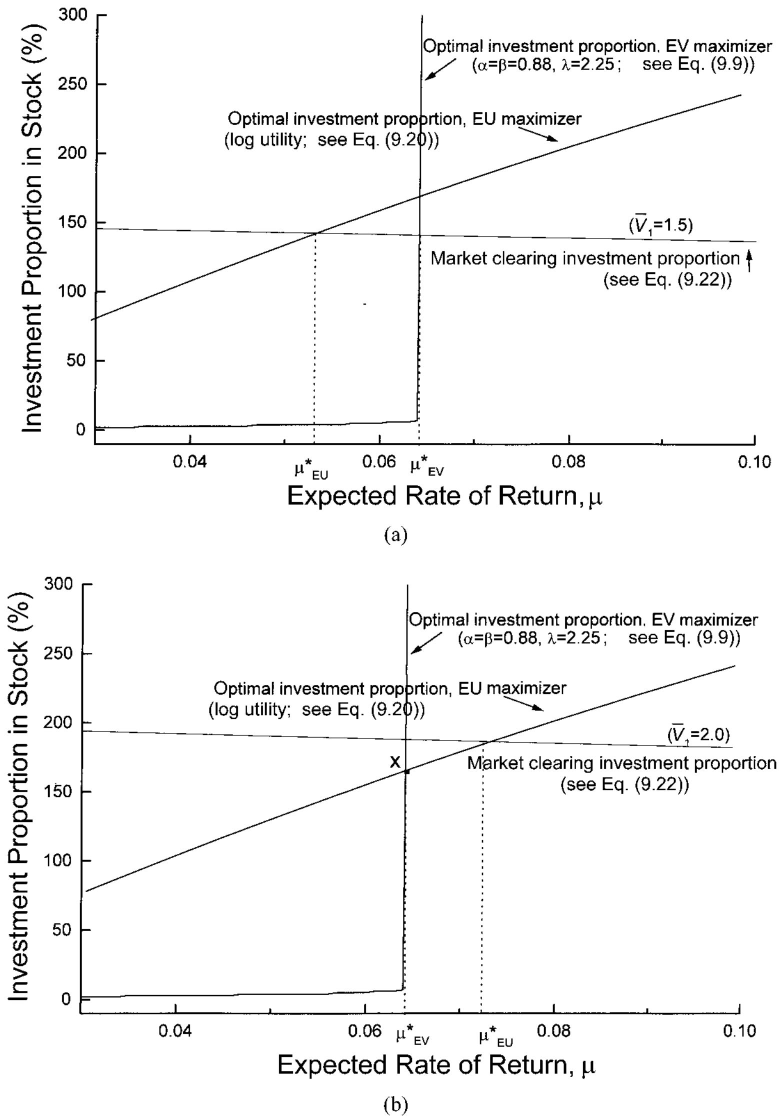


FIGURE 9.6 (a) Equilibrium expected return. (b) Equilibrium expected return.

taken as 0.8%, and the end-of-period value, $\tilde{V}_1$, is normally distributed with a mean of 1.5 and standard deviation of 0.165.[11] The equilibrium expected return is depicted in Figure 9.6a by μ^*_{EU} for the case of the expected utility maximizer and by μ^*_{EV} for the case of the expected value maximizer.

When $\bar{V}_1 = 1.5$, which is the situation depicted in Figure 9.6a, the equilibrium expected rate of return in a market of EU maximizers, μ^*_{EU}, is lower than the equilibrium expected rate of return in an identical market with EV maximizers. Different values of the end-of-period expected value, $\bar{V}_1$, correspond to different market clearing investment proportion lines (see Eq. (9.22)). The higher $\bar{V}_1$, the higher the line describing the market clearing investment proportion. Figure 9.6b describes the same market scenario as depicted in Figure 9.6a, with the exception that $\bar{V}_1$ is higher ($\bar{V}_1 = 2.0$). Since the optimal investment proportion line for EV maximizers is almost vertical in the relevant range, the different value of $\bar{V}_1$ has almost no effect on μ^*_{EV}. In contrast, the line describing the optimal investment proportion for EU maximizers is moderately upward sloping, and as a consequence when $\bar{V}_1$ is increased, μ^*_{EU} also increases (see Figure 9.6b). If the value of $\bar{V}_1$ is such that the market clearing investment proportion crosses the EU and EV optimal investment proportion lines at their intersection point (point x in Figure 9.6b), the stock is priced the same by EU maximizers and by EV maximizers. For lower values of $\bar{V}_1$, the stock is priced higher by the EU maximizers. However, the situation is reversed for higher values of $\bar{V}_1$. In this case $\mu^*_{\mathrm{EU}} > \mu^*_{\mathrm{EV}}$, which implies that the stock is priced higher by the EV maximizers (see Figure 9.6b). Thus, the higher $\bar{V}_1$, the larger the difference between the price of a stock in a market of EV maximizers and the price of an identical stock in a market of EU maximizers.

To sum up, the main findings from our theoretical analysis of asset allocation and equilibrium pricing in the PT framework are as follows:

1. Given a risky stock with a given rate of return distribution and a riskless asset, the asset allocation decision of a PT investor characterized by the Tversky and Kahneman value function (in the relevant case where $\alpha = \beta$) is independent of the investor's wealth. Namely, the optimal investment proportion in the stock is not a function of the investor's wealth.

[11] The optimal investment proportions as a function of the expected return shown in Figure 9.6 have the same features as the optimal investment proportions shown in Figure 9.3. The difference between the two figures (apart for the different parameters) is that in Figure 9.3 the standard deviation of the return distribution is constant, while in Figure 9.6 it changes as μ changes (see footnote 10). However, this does not change the general characteristics of the optimal investment proportion as a function of the expected return: a sharp kink for the EV maximizer and a smooth transition for the EU maximizer.

2. The diversification policy implied by PT is rather extreme: given one riskless asset (bond) and one risky asset (stock) with *arbitrary rate of return distribution parameters*, an investor with the Tversky and Kahneman value function will typically be in one of the following two states: (a) fully invested in the bond or (b) fully invested in the stock. The crossover between these two states is sharp: a small change in one of the rate of return distribution parameters may lead to a switch from one state to the other. However, the sharp crossover does not imply that investors flip-flop from bonds to stocks. Rather, it enables us to characterize a relationship that must hold between the parameters of the rate of return distribution *in equilibrium* when all investors are informed (i.e., the distribution of $\tilde{V}_1$ is known).

3. A risky security with some distribution of end-of-period value, $\tilde{V}_1$, will be priced higher by EV maximizers (relative to the pricing by EU maximizers) if the expectation regarding the end-of-period value, $\bar{V}_1$, is high. In contrast, if $\bar{V}_1$ is low, EV maximizers will price the security lower than EU maximizers.

In the next section we utilize these theoretical results in order to investigate the implications of PT to asset pricing and to market dynamics in a dynamic market model. As the dynamics are analytically intractable, we employ the microscopic simulation (MS) method in our analysis. We use the framework of the LLS model, described in Chapter 7. To examine the effects of PT, we compare simulations of a market that has EU maximizers with simulations of a market that is identical apart from the fact that investors' behavior is described by PT (rather than EUT).

9.4 PROSPECT THEORY AND MARKET DYNAMICS IN THE LLS MODEL

In this section we employ the LLS microscopic simulation model to investigate the effects of the behavioral elements of PT on asset pricing and on the market dynamics. The LLS model is described in detail in Chapter 7. In this model there are two main types of investors: rational informed identical (RII) investors and efficient market believers (EMB). The RII investors are essentially informed fundamentalists who estimate the stock's fundamental value and act on the assumption that the stock price will eventually converge to this value. The RII investors look for mispricings and try to exploit them. The EMB investors, in contrast, believe that the market is efficient and that the stock is correctly priced. As a consequence, their investment decisions are reduced to the optimal diversification between the stock and the riskless bond. This optimization requires the *ex ante* return distribution of the stock, but as this distribution is not available, the EMB investors employ the *ex post* return distribution in order to estimate the *ex ante* distribution.

The behavioral elements of PT have several effects on the market dynamics. Moreover, there are complex cross-influences between these various effects. To simplify the analysis, we examine each of the main effects separately. In Section 9.4.1, we examine the effects of expected value maximization (as opposed to expected utility maximization) on asset pricing when all investors are RII. In Section 9.4.2, we examine the effects of subjective probability distortions on pricing, in the same setting with RII investors. Section 9.4.3 examines the effects of value function maximization on the behavior of EMB investors and on the dynamics of a market with both RII and EMB investors.

9.4.1 Expected Value Maximization and Asset Pricing: RII Investors

One of the key elements of PT is the notion that people treat gains and losses differently and that they are significantly more sensitive to losses. Also, it has been documented that people are risk averse with respect to gains but are risk seeking with respect to losses. These properties are manifested in the S-shaped value function, which has an inflection point at the origin (see Figure 9.1). How does expected value (EV) maximization (as opposed to expected utility (EU) maximization) affect asset pricing? The answer is not clear: on the one hand, EV maximizers display increased sensitivity to losses (the value function is steeper for losses than it is for gains); on the other hand, the value function is convex in the negative region, suggesting risk-seeking behavior regarding negative prospects. Indeed, it turns out that the effect on asset pricing can go both ways. As we shall see, the effect of value function maximization on asset pricing depends not only on the value function parameters but also on the asset itself.

Our theoretical analysis in Section 9.3 has led to the prediction that EV maximizers will overprice a risky asset (with respect to the valuation of EU maximizers) if the expected end-of-period value of this asset is high, but they will underprice assets with relatively low expected end-of-period values. In this section we test this prediction in the framework of the LLS model. For the sake of simplicity, in this section we analyze the effect of EV maximization only on the informed (RII) investors. The effect of EV maximization on EMB investors will be discussed in Section 9.4.3.

Recall that at time t the RII investors estimate the stock's end-of-period ($t + 1$) payoff as

$$\overline{P}_{t+1} + \overline{D}_{t+1} = \frac{D_t(1+g)^2}{k-g} + D_t(1+g) \tag{9.23}$$

where P_{t+1} and D_{t+1} are the period $t + 1$ price and dividend, k is the discount factor, and g is the dividend expected growth rate (see Eqs. (7.3) and (7.4) in Chapter 7). Multiplying by the number of outstanding shares, N, and rearranging we obtain the following expression for the firm's expected end-of-period value:

$$\bar{V}_{t+1} = N\left(\bar{P}_{t+1} + \bar{D}_{t+1}\right) = \frac{ND_t(1+g)(1+k)}{k-g} \tag{9.24}$$

Note that the expected end-of-period value increases with the average dividend growth rate, g. Recall that the theoretical derivation of Section 9.3 predicts underpricing of the stock by EV maximizers when the dividend growth rate is low, but overpricing when the dividend growth rate is high. To check this prediction, we simulate the LLS model with RII investors. In this model the RII investors are informed about the end-of-period value distribution, $\tilde{V}_1$. They choose their investment proportions in order to maximize their expected utility (or expected value in the case of EV maximizers). Prices are then determined by market clearance. (The details of the dynamics of the LLS model are given in Chapter 7.) We compare the stock price in a market in which all investors are characterized by a power utility function with $\alpha_U = 1.5$, with the prices in a market with EV maximizers with value function parameters as estimated by Tversky and Kahneman: $\alpha = \beta = 0.88$, $\lambda = 2.25$. In both markets the exact same dividend stream is realized. At each time period the relative pricing is calculated as $(P_{EV} - P_{EU})/P_{EU}$, where P_{EV} stands for the price of the stock in the EV maximizers' market, and P_{EU} stands for the price of the stock in the EU maximizers' market. As the evaluation period of EV maximizers (the horizon over which they calculate their gains/losses) is argued to be about one year (see Benartzi and Thaler, 1995), we take the time period in the simulation as a year, and we choose the rest of the parameters accordingly: $r = 4\%$, $k = 10\%$, D_0/P_0 (the initial dividend yield) $= 4\%$. We report the average relative pricing as a function of the dividend growth rate in Figure 9.7.

For annual dividend growth rates of up to 8%, the stock is priced by the EV maximizers about 2% lower than the pricing by the EU maximizers. At a dividend growth rate of about 9%, the pattern is reversed, and the EU maximizers set a higher price. This overpricing increases dramatically as g approaches the value of the discount factor k (recall that the stock's fundamental value is derived from Gordon's dividend model, and it is defined only for $g < k$). For the average historical S & P dividend growth rate of about 6.5%,[12] the experimentally measured value function param-

[12] There are various ways to estimate the average historical dividend growth rate. If one assumes a relatively constant dividend yield, one can approximate the average dividend growth rate as being equal to the average capital appreciation, which was about 6.5% in the period 1925–1998 (see Ibbotson Associates 1999 Yearbook).

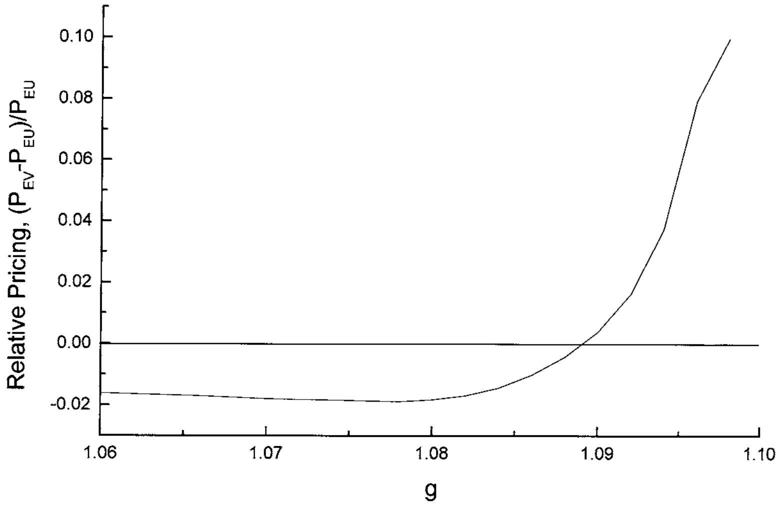


FIGURE 9.7 Under/overpricing by expected value maximizers as a function of the dividend growth rate, g.

eters yield an underpricing of about 2% with respect to the pricing by EU maximizers with the experimentally estimated relative risk aversion coefficient.

9.4.2 Probability Distortion and Asset Pricing

Many experimental studies have shown that people tend to overweigh low-probability events in their decision making (see Chapter 2). Tversky and Kahneman suggest the transformation in Eq. (9.2) as a description for the subjective distortion of probability. The extent of the probability distortion is given by the value of the parameter γ. For $\gamma = 1$, the transformation in Eq. (9.2) yields the original objective probability distribution, and there is no distortion. The lower γ, the more distortion takes place. γ is experimentally estimated as about 0.6 (see Camerer and Ho, 1991; Tversky and Kahneman, 1992).

In this section we analyze the effect of probability distortion on asset pricing. The relationship between probability distortion and asset pricing is not simple, and it depends not only on the extent of the probability distortion but also on the shape of the objective probability distribution. To illustrate the importance of the shape of the objective probability distribution, consider the following example. Table 9.1 gives the discrete rate of return distribution for assets A and B. Assume that the low probabilities of 0.1 are subjectively overweighted as 0.2 and that the

TABLE 9.1 The Effects of Probability Distortion

A			B		
Outcome	Objective probability	Subjective probability	Outcome	Objective probability	Subjective probability
− 10%	0.1	0.2	−4%	0.1	0.2
10%	0.8	0.6	−2%	0.8	0.6
30%	0.1	0.2	120%	0.1	0.2
μ	10.0%	10.0%	μ	10.0%	22%
σ	8.9%	12.6%	σ	36.7%	49%

objective probability of 0.8 is underweighted as 0.6.[13] The main difference between the two assets in Table 9.1 is that the distribution of asset A's outcomes is symmetric around the mean outcome, while asset B's outcomes are positively skewed. The table shows that the probability distortion makes investors perceive asset A as more risky than it really is (12.6% > 8.9%), while its perceived mean is not altered by the probability distortion. As a result, investors who distort the objective probabilities would price asset A lower than investors who employ the objective probabilities. In contrast, the exact same probability distortion may have an opposite effect on the pricing of asset B. For asset B, which has a return distribution with positive skewness, the distortion of probabilities increases both the perceived mean (22% > 10%) and the perceived standard deviation. In this example, the effect on the perceived mean is larger than the effect on the perceived standard deviation (49% > 36.7%). Thus, investors who distort probabilities may very well price asset B higher than investors who employ the objective probabilities in pricing this asset. The intuition for this result is as follows: the probability distortion increases the subjective weight assigned to both extreme outcomes; however, since for asset B the high-extreme outcome (120%) is much further away from the mean outcome than the low-extreme outcome (−4%), the probability distortion increases the perceived mean. Indeed, this is exactly one of the explanations suggested for the fact that people buy lotteries: although these lotteries are objectively unfair games, their outcomes are typically very highly skewed (a few very large prizes with very low probabilities), and therefore, even a slight distortion of probabilities may make people perceive them as attractive. In the present analysis we focus on the pricing of stocks, which have more-or-less symmetric return distributions (at least if relatively short-term returns are considered). We therefore concentrate on

[13] For the sake of simplicity, we use a discrete probability transformation in this example. We later consider more realistic continuous return distributions, and the cumulative probability transformation in Eq. (9.2).

assets with fairly symmetric return distributions and do not go into the effect of probability distortion on the pricing of assets with highly skewed return distributions (such as lotteries).

Even if the analysis is restricted to assets with perfectly symmetrical return distributions, the effect of probability distortion is not obvious. The probability transformation in Eq. (9.2) distorts the objective probability distribution in the following way: probability is shifted from central part of the distribution to the extremes; this is done in such a way that more probability is shifted to the upper extreme than to the lower extreme (this can be seen in Figure 9.2). Thus, the transformation typically increases both the perceived mean and the perceived standard deviation.

As the return distribution estimated by the RII investors in the LLS model is uniform (see Eq. (7.4)), we focus here on the uniform distribution. However, the general results reported here hold for other distributions such as the normal or two-point distributions as well. Figure 9.8 shows the perceived mean and standard deviation of a uniform rate of return distribution as a function of the probability distortion parameter, γ. The objective distribution is uniform in the range $[-0.1, 0.2]$. μ_0 and σ_0 denote the objective mean and standard deviation ($\mu_0 = 0.05$, $\sigma_0 = 0.087$). Both μ and σ increase as more probability distortion takes place (recall that the smaller γ, the more the probability is distorted). Notice, however, that σ increases almost linearly as γ decreases, while μ increases at an

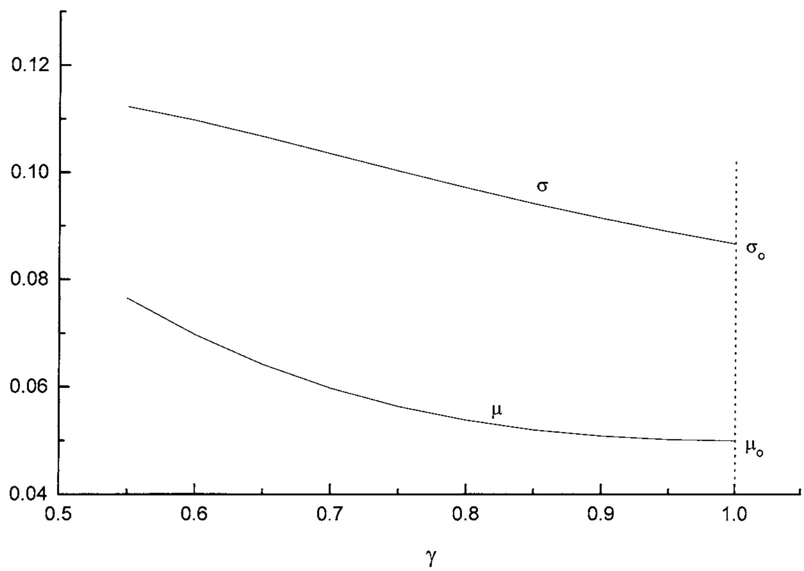


FIGURE 9.8 Subjective perception of mean μ and σ as a function of the probability distortion parameter, γ.

accelerating pace: it is first almost unaffected by reductions in γ, while for smaller γ it is much more sensitive to γ than σ is. This pattern is typical of different distributions with various parameters. This observation leads one to suspect that for relatively high values of γ (low probability distortion), the increase in σ will be more significant than the almost negligible increase in μ and, as a consequence, investors who base their decisions on the distorted probabilities will price risky assets lower than those who employ the objective probabilities. However, for low values of γ, the increase in μ becomes more significant than the increase in σ, which suggests that investors who distort probabilities will price risky assets higher than the objective investors.

To test this hypothesis, we simulate an LLS stock market with RII investors. The optimization of RII investors is given by Eq. (7.7) in Chapter 7. We record the relative pricing in the simulations as a function of γ. Namely, we simulate markets with a stock yielding the same stream of dividend realizations, but in each market we assume a different value of γ, and we record prices relative to the prices in a market of investors who employ the objective probabilities ($\gamma = 1$). The results are reported in Figure 9.9. As suspected, we observe that investors who mildly distort probabilities ($0.75 < \gamma < 1$) underprice the stock relative to unbiased investors, while more severe probability distortion ($\gamma < 0.75$) leads to

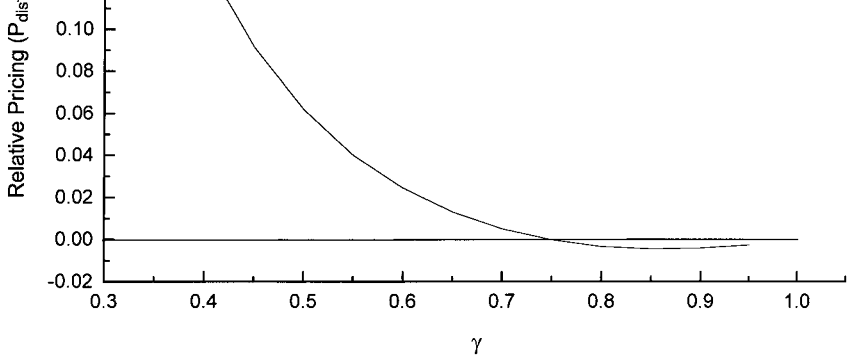


FIGURE 9.9 Relative pricing as a function of the probability distortion parameter, γ.

overpricing. The range of the experimentally estimated values of γ, $0.55 < \gamma < 0.70$, corresponds to an overpricing of about 1 to 4%.

9.4.3 Price Dynamics with EMB Expected Value Maximizers and RII Expected Utility Maximizers

In Section 9.4.1 we analyzed the effect of EV maximization on asset pricing in a market with RII investors only. Here we examine the effects of EV maximization in a market with a minority of EMB investors. Specifically, we compare the price dynamics in a market that has a minority of EV maximizing EMBs with the dynamics of a market in which the EMB investors are EU maximizers. In both cases the RII investors are EU maximizers.

One of the conclusions from our theoretical analysis in Section 9.3 is that EV maximization leads to a rather extreme asset allocation policy. Namely, for an *arbitrary* set of return distribution parameters, EV maximizers will typically be either fully invested in the stock or fully invested in the bond. The crossover between these two states is sharp (see Figures 9.3 to 9.5). Despite this sharp crossover, we argued that informed forward-looking investors (such as the RII investors) do not flip-flop between the two assets. Instead, we showed that in equilibrium, EV maximizers with information on the end-of-period value will drive the asset's expected return exactly to the crossover expected return, which we denoted by μ^*_{EV}. However, this is not the case for the EMB investors. Recall that the EMBs do not have information about the stock's end-of-period value. Instead, they estimate the stock's *ex ante* return distribution from its *ex post* return distribution (see Chapter 7). Thus, an update of the *ex post* distribution (as a consequence of the observation of a new return) leads to an update of the estimation of the *ex ante* return distribution and may certainly lead to a dramatic switch in the asset allocation of EV maximizers from full investment in one asset to full investment in the other. This is in contrast to the much smoother change in asset allocation typical of EU maximizers (see Figure 9.3).

As explained in Chapter 7, following an exceptional dividend realization, the EMBs may drive the price away from the fundamental value by initiating a positive feedback loop. Namely, a high dividend realization typically generates a high rate of return on the stock. This makes EMB investors more optimistic about the stock's *ex ante* return distribution, and they increase their investment proportion in the stock. If the EMBs have enough influence on the price, their increased demand for the stock may induce a price increase and another high rate of return. Thus, a positive feedback loop that drives the price away from the fundamentals may be initiated. For such a price deviation to occur, the EMBs must have enough impact on the price. In other words, they must either hold a large enough portion of the total wealth or, alternatively, they must shift the capital at their disposal aggressively enough to the stock. As argued earlier, when the EMB investors are EV maximizers, such an aggressive shift of funds is likely to occur (as opposed to a much more moderate shift if they were EU maximizers). Thus if EMB investors are EV maximizers, we would expect deviations of the price from the fundamental value, crashes, and booms to be more likely and to occur more frequently. Indeed, this is what we find in our simulations.

Figure 9.10 shows the price dynamics of a typical simulation. The light line shows the price in a market with 95% EU maximizing RII investors and 5% EMB investors, who are also EU maximizers. The heavy line shows the same market (with the same dividend realizations) but with EMB investors who are EV maximizers (the other 95% are again EU maximizing RII investors). In both cases the EMB population is heterogeneous. m, the number of *ex post* observations employed in the estimation of the *ex ante* distribution, is normally distributed in the EMB population with $\overline{m} = 40$, and $\sigma_m = 10$. In the case of EV maximizers, we take the value function parameters as estimated experimentally by Tversky and Kahneman: $\alpha = \beta = 0.88$, $\lambda = 2.25$. In the case of the EU maximizers, we take power preferences with a typical relative risk-aversion coefficient of 1.5. As Figure 9.10 shows, the same dividend realizations that do not generate significant price deviations in the case of EU maximizers do generate crashes, booms, and significant deviations of the price from the

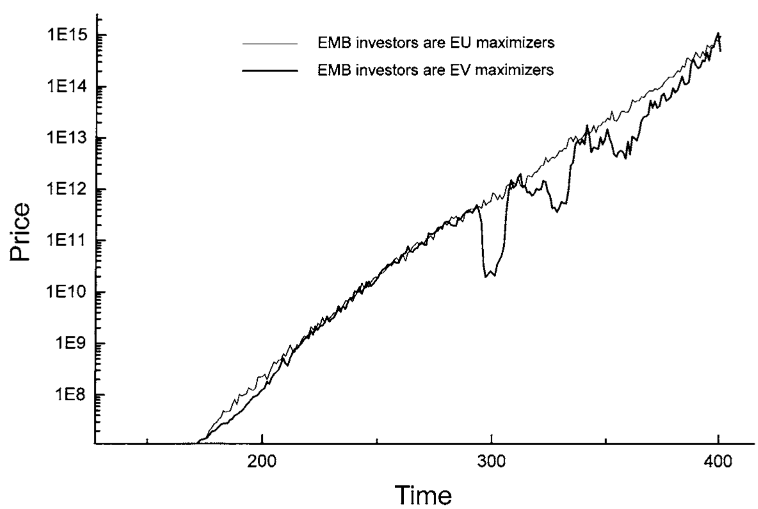


FIGURE 9.10 Price dynamics with 5% EMB investors.

fundamental value when the EMB investors are EV maximizers. This result is very typical of many other simulations with various market parameterizations.

As discussed in Chapter 7, the price deviations from the fundamental value, which are induced by the EMBs, generate high trading volume, excess volatility, short-term momentum, long-term mean reversion, and a correlation of volume with contemporaneous and lagged absolute returns. All of these phenomena are observed when the EMB investors are EU maximizers. Since the price deviations are induced more readily and more frequently when the EMB investors are EV maximizers, these phenomena are enhanced if the EMB investors (or some of them) maximize expected value rather than expected utility.

9.5. SUMMARY

Many systematic deviations from the normative predictions of the expected utility framework have been repeatedly documented in many decision-making experiments. Kahneman and Tversky have cast the main experimental findings regarding individuals' decision making into a unified framework, which is known as prospect theory. The main elements of PT are as follows:

1. Decisions are made based on *change* in wealth, x, rather than on *terminal* wealth, $w + x$.
2. Individuals' behavior can be described by the maximization of the expectation of an S-shaped value function, $EV(x)$. Tversky and Kahneman suggest the power value function in Eq. (9.1) and experimentally estimate $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 2.25$.
3. Individuals subjectively distort probabilities such that low-probability events are overestimated.

In this chapter we examined the effects of PT on asset allocation, asset pricing, and the market dynamics. Theoretical analysis yields the following results:

1. Given a risky stock with a given rate of return distribution and a riskless asset, the asset allocation decision of a PT investor characterized by the Tversky and Kahneman value function (in the relevant case where $\alpha = \beta$) is independent of the investor's wealth. Namely, the optimal investment proportion in the stock is not a function of the investor's wealth.

2. The diversification policy implied by PT is characterized by a sharp crossover between full investment in the bond to full investment in the stock.

3. A risky security with some distribution of end-of-period value, $\tilde{V}_1$, will be priced higher by EV maximizers (relative to the pricing by EU maximizers) if the expectation regarding the end-of-period value, $\bar{V}_1$, is high. In contrast, if $\bar{V}_1$ is low, EV maximizers will price the security lower than EU maximizers.

Armed with these theoretical results, we set to investigate the effects of PT on asset pricing and price dynamics in a dynamic market model. Microscopic simulation is a natural tool for this analysis, and we employ the LLS microscopic simulation model in our investigation. Our main findings in these simulations are as follows:

1. Although PT implies high sensitivity to losses, this does not necessarily mean that risky securities are priced lower by PT investors (relative to the pricing by EU maximizers). The relative pricing depends not only on the value function and utility function parameters, but also on the nature of the risky asset. Generally, PT investors price assets with low expected returns lower than EU maximizers, but they price assets with a high expected returns higher than EU maximizers. The experimentally estimated value function and utility function parameters and the historical average dividend growth rate correspond to an underpricing of about 2% by EV maximizers.

2. Probability distortion does not have a clear, one-way effect on asset pricing. Low levels of probability distortion ($0.75 < \gamma < 1$) cause the stock to be underpriced relative to the pricing based on the objective probabilities. Higher levels of probability distortion ($\gamma < 0.75$) lead to overpricing. The experimentally estimated value of $\gamma = 0.6$ leads to an overpricing of about 2%.

3. When some investors in the market employ the *ex post* return distribution in order to estimate the *ex ante* distribution (as the EMBs do), EV maximization leads to more frequent price deviations from the fundamental value. This enhances the phenomena of heavy trading volume, excess volatility, short-term momentum, long-term mean reversion, and the correlation of volume with contemporaneous and lagged absolute returns.

Clearly, this is only a preliminary analysis. There is still much to be learned about the effects of PT, and other experimentally documented behavioral elements, on asset allocation, asset pricing and market dynamics. It seems to us that microscopic simulation can be a very useful tool in this line of investigation.

APPENDIX 9.1

For a uniform rate of return distribution, the discontinuous crossover from the bond to the stock occurs at the point:

$$\mu^* = \sqrt{3}\,\sigma\left(\frac{\lambda^{1/(\alpha+1)} - 1}{\lambda^{1/(\alpha+1)} + 1}\right)$$

Proof. Consider a uniform rate of return distribution in the range $[\mu - \Delta, \mu + \Delta]$. This distribution has a mean of μ, a standard deviation of $\Delta/\sqrt{3}$, and a density of $1/(2\Delta)$. At the crossover point from the bond to the stock, the following must hold:

$$\int_0^{\infty} f(R)R^{\alpha}\,dR - \lambda\int_{-\infty}^{0} f(R)(-R)^{\alpha}\,dR = 0$$

(see Eq. (9.17)), where $f(R)$ is the rate of return density function. For the previous uniform rate of return distribution, this translates to

$$\frac{1}{2\Delta}\left[\int_0^{\mu+\Delta} R^{\alpha}\,dR - \lambda\int_{\mu-\Delta}^{0}(-R)^{\alpha}\,dR\right] = 0$$

or

$$(\mu + \Delta)^{\alpha+1} = \lambda(\Delta - \mu)^{\alpha+1}$$

Exponenting by $1/(\alpha + 1)$, we obtain

$$(\mu + \Delta) = \lambda^{1/(\alpha+1)}(\Delta - \mu)$$

Rearranging, and substituting $\Delta = \sqrt{3}\,\sigma$, we find that at the crossover point the expected return is given by

$$\mu^* = \sqrt{3}\,\sigma\left(\frac{\lambda^{1/(\alpha+1)} - 1}{\lambda^{1/(\alpha+1)} + 1}\right).$$

10

APPLICATION OF MICROSCOPIC SIMULATION TO THE CAPM: HETEROGENEOUS EXPECTATIONS AND THE NUMBER OF ASSETS IN THE PORTFOLIO[1]

10.1. INTRODUCTION

One of the cornerstones in modern finance is the capital asset pricing model (CAPM) developed by Sharpe (1964) and Lintner (1965b). The CAPM makes several assumptions regarding investors' behavior. What happens to the CAPM results when these assumptions are relaxed? In this chapter we focus on one crucial assumption—the assumption asserting that the investors have homogenous expectations. In the CAPM derivation, it is assumed that all investors have information on the mean rate of return vector $\underline{\mu}$ (where the underline denotes a vector) and on the variance—covariance matrix Σ. The homogeneous expectations assumption asserts that all investors agree about these parameters. Employing Markowitz's (1952a) mean-variance investment rule, the following main theoretical results are obtained from the CAPM:

1. All investors hold the same unlevered portfolio, called the *market portfolio*. Investors differ only with respect to the amount of borrowing and

[1] This chapter is based on a forthcoming study by Levy, H., Levy, M., and Benita, G.

lending, which reflects their preferences. Investors separate the two decisions: (a) how to construct a portfolio of risky assets and (b) how much of their wealth to allocate to the market portfolio of risky assets and how much wealth to allocate to the riskless asset. Hence the name *separation theorem*. Since all investors hold all risky assets in the same proportions, this implies that short sales cannot exist in equilibrium.

2. Beta, β (rather than σ), is the individual's asset risk measure.

3. In equilibrium, there is a linear relationship between the mean return of an individual asset and its risk. This relationship is well known as the *security market line* (SML) given by

$$\mu_i = r + (\mu_m - r)\beta_i \tag{10.1}$$

where i stands for the ith asset, μ_m is the market portfolio mean rate of return, and r is the riskless interest rate.

The CAPM is at the foundation of modern portfolio theory and provides one of the fundamental theoretical relations between risk and return in the capital market. However, empirical tests reveal only weak support or no support at all to the CAPM—see Lintner (1965b) Black, Jensen, and Scholes (1972), Miller and Scholes (1972), Levy (1978), Fama and French (1992), and Amihud, Christensen, and Mandelson (1992). In addition, surveys show that individual investors hold, on average, only 3.41 different stocks in their portfolio (see Blume, Correctt and Friend, 1974; Blume and Friend, 1975), in contrast to the CAPM which asserts that all assets should be included in all investors' portfolios.

It is quite disturbing that we find such weak empirical support for one of the most fundamental models in finance. There are two main explanations for the lack of empirical support for the CAPM. First, it is possible that the CAPM assumptions do not hold, and therefore the model's theoretical predictions are not consistent with the empirical observations. Alternatively, it is also possible that the lack of empirical support for the CAPM is due to the problematic nature of the empirical methodology employed (for criticism of the empirical tests, see Roll, 1977). Let us elaborate on the difficulties of empirically testing the CAPM.

The empirical test of the CAPM—that is, of the linear equation given in Eq. (10.1)—employs *ex post* parameters, simply because the *ex ante* parameters are not observed. Sharpe and Lintner may assert that the CAPM holds with *ex ante* parameters, hence the fact that it is rejected with *ex post* estimates does not invalidate the CAPM. It is possible, for example, that at any point in time, the sample mean, $\bar{R}_i$, does not reflect what investors think will be next year's mean return μ_i, hence, a gap between the empirical findings and Eq. (10.1) occurs. Indeed, the only pure test of the CAPM that used *ex ante* data was done in an experimental study, which showed strong support for the CAPM and the general CAPM denoted by GCAPM (see Levy, 1997; a discussion of the GCAPM is presented later).

The MS methodology provides us with a framework to test the effects of deviations from the CAPM assumptions on the model's results. In the microscopic simulation (MS) analysis one can use *ex ante* parameters, hence no statistical estimates are employed, and the CAPM is tested in Sharpe-Lintner *ex ante* framework. Thus, the argument as to whether *ex post* data reflect *ex ante* parameters is isolated. This can be done in experimental studies (see Levy, 1997) or MS studies, but not in empirical studies.

In this chapter we employ MS to test the effect of two factors related to the CAPM on equilibrium prices: one factor is related to deviations from the homogeneous expectations assumption discussed previously, and the other factor is related to the observed number of assets held in the portfolio, which is found empirically to be relatively small. We first analyze the effect of heterogeneous expectations on equilibrium prices, and then we analyze the effect of the number of assets held in the portfolio on the CAPM's results.

10.2. HOMOGENEOUS AND HETEROGENEOUS EXPECTATIONS AND EQUILIBRIUM PRICES

The classical CAPM makes the strong assumption that all investors have homogenous expectations regarding $\underline{\mu}$ and Σ. Lintner (1965a, 1969) also derives equilibrium results with heterogeneous expectations, but the results are very complex and there is no analysis regarding the possible deviation of the results of the heterogeneous case from the results corresponding to the homogeneous case. In this chapter we analyze the difference between the homogeneous and heterogeneous regimes and their effect on assets' prices.

10.2.1 The Model

Homogeneous Expectations

In the CAPM an equilibrium relationship between expected returns, variances, and covariances is derived. Prices do not play a direct significant role in this derivation.[2] However, when heterogeneous expectations are considered, prices do play a key role in the analysis. To understand the

[2] Lintner, 1965b, shows how one can start with distributions of end-of-period liquidation prices, $\tilde{P}_1$, and transform the equilibrium relationship between $\underline{\mu}$ and Σ into a statement about equilibrium prices, P_0. This transformation, however, may sometimes be problematic, as the vector of equilibrium prices, P_0, may be nonunique or even may be nonexisting (see Nielsen, 1988; M. Levy, 1997).

application of MS to price determination in the heterogeneous CAPM model, we first need to understand how equilibrium prices are calculated in the MS framework in the case of homogeneous expectations. We employ Sharpe's approach assuming that $\underline{\mu}$ and Σ are known and investors agree on these parameters. Based on $\underline{\mu}$ and Σ, we derive the efficient frontier, and for a given riskless interest rate, r, we derive the optimal unlevered portfolio (the market portfolio) and, of course, the SML, exactly as done theoretically by Sharpe (1964). We denote the following:

- N_i—The number of outstanding shares of the ith firm
- T_k—The invested wealth of the kth investor[3]
- x_i—The investment proportion in the ith stock (which is identical for all investors in the homogeneous case)
- n—The number of risky assets
- K—The number of investors

By the market clearance condition we obtain

$$P_{i0} = x_i \sum_{k=1}^{K} T_k / N_i \tag{10.2}$$

where P_{i0} is the current CAPM homogeneous expectation equilibrium price of the ith asset. The security risk, β_i is given by

$$\beta_i = \frac{x_i \sigma_i^2 + \sum\limits_{\substack{j=1 \\ j \neq i}}^{n} x_j \sigma_{ij}}{\sum\limits_{i=1}^{n} x_i^2 \sigma_i^2 + \sum\limits_{i=1}^{n} \sum\limits_{\substack{j=1 \\ j \neq i}}^{n} x_i x_j \sigma_{ij}} \tag{10.3}$$

The CAPM is a one-period model. It is best thought of in terms of firms that operate for 1 year and then liquidate their assets and distribute them to the stockholders. The asset distribution is the true future cash flows of the firm due to its operation. Given $\underline{\mu}$ and Σ, and the equilibrium prices $\underline{P}_0$, one can learn about the distribution of end-of-period liquidation values $\underline{P}_1$. Of course, the opposite is also true; given the end-of-period distribution, $\underline{P}_1$, one can assert that $\underline{\mu}$, Σ, and $\underline{P}_0$ are determined simultaneously, Lintner's (1965b) approach. We employ here the first approach—that is, based on $\underline{\mu}$ and Σ we derive $\underline{P}_0$ from which we learn the distribution of the liquidation values. Let us elaborate.

[3] We assume that the net borrowing by all investors is zero. Hence $\Sigma T_k = w_0$ is the initial wealth as well as the wealth invested in all the risky assets. This assumption can be relaxed.

Given $\underline{\mu}$, Σ, and r (defined per dollar of investment or in percentage terms), one can solve for the optimal investment proportion in the ith asset, x_i. By the market clearance condition we have $w_0 x_i = N_i P_{i0}$, where w_0 is the total wealth invested in the risky assets by all investors. By definition of the expected rate of return of the asset, μ_i, we have

$$\mu_i = \frac{\bar{P}_{i1} - P_{i0}}{P_{i0}}$$

where $\bar{P}_{i1}$ is the expected price of the ith stock at the end of the period.[4] Having μ_i and P_{i0}, one can derive also the expected implied end-of-period value $\bar{P}_{i1}$ by the CAPM equilibrium as

$$\bar{P}_{i1} = P_{i0}(1 + \mu_i)$$

Similarly, one can derive the end of period distribution of $\tilde{P}_{i1}$, the variance of this distribution, as well as the covariance of end of period values of the ith firm with any other firm.

Given the *ex ante* variance σ_i^2 (per one dollar of investment), (recall that Σ is assumed to be known) by definition,

$$\sigma_i^2 = \frac{\mathrm{var}\left(\tilde{P}_{i1}\right)}{P_{i0}^2} \quad \text{and} \quad \sigma_{ij} = \frac{\mathrm{cov}\left(\tilde{P}_{i1}, \tilde{P}_{j1}\right)}{P_{i0} P_{j0}}$$

Hence,

$$\mathrm{var}\left(\tilde{P}_{i1}\right) = P_{i0}^2 \sigma_i^2 \quad \text{and} \quad \mathrm{cov}\left(\tilde{P}_{i1}, \tilde{P}_{j1}\right) = \sigma_{ij} P_{i0} P_{j0}$$

Because by the CAPM $\underline{\mu}$ and Σ are assumed to be known, for given equilibrium price vector $\underline{P}_0$, one can derive the end-of-period expected values, variances, and covariances in terms of the stock prices. We employ these end-of-period distributions when we analyze the heterogeneous case discussed later. When we compare pricing in a heterogeneous-expectations market with the pricing in a homogeneous-expectations market, we assume identical distribution of end-of-period values, $\tilde{P}_{i1}$ (the liquidation values); in other words, we assume that the economy is the same in both cases. How do heterogeneous beliefs affect equilibrium prices?

Heterogeneous Expectations

So far, we were in the CAPM framework with homogeneous expectations. Let us now turn to the heterogeneous-expectations case, which is

[4] We assume no dividends. If dividends exist, then $\bar{P}_{i1}$ is the end of period total return, price plus expected dividends. If interim dividends are paid, they are assumed to be reinvested in the stock.

more realistic and more interesting. For simplicity we focus on one dimension of heterogeneous expectations—disagreement regarding the expected rate of return. The extension to other dimensions of heterogeneity (e.g., regarding Σ) is straightforward.

We assume that all investors agree on Σ but disagree on $\underline{\mu}$. To be more specific, the kth investor's estimate of the mean return of the ith asset is μ_{ik}, which is given by

$$\mu_{ik} = \mu_i(1 + \varepsilon_{ik}) \tag{10.4}$$

where ε_{ik} is a disagreement factor, which is assumed to be normally distributed as follows: $\varepsilon_{ik} \sim N(0, \sigma)$. Because each investor has a different realization, ε_{ik}, expectations are heterogeneous. The normality assumption is for simplicity only; the same analysis holds also with other distributions. We assume, however, that $E(\varepsilon_{ik}) = 0$, implying that on average the market is in agreement with the CAPM parameters. This assumption can easily be relaxed, but we think that to analyze the effect of heterogeneity on the CAPM results, it is best to assume that on average the deviations from the CAPM parameters are zero.

With heterogeneous expectations, the separation theorem breaks down. The investment proportions in the ith stock generally varies across investors (because of their different beliefs). We denote the investment proportion of the kth investor in the ith stock by x_{ik} (rather than x_i).

By the market clearance condition we obtain

$$P_{i0}^* = \sum_{k=1}^{K} T_k x_{ik}/N_i \tag{10.5}$$

where a superscript star denotes values with heterogeneous beliefs. Note that unlike the CAPM clearance condition, here x_{ik} rather than x_i is employed, hence P_{i0}^* may be different from P_{i0}.

Having the distributions of end-of-period liquidation values, $\tilde{P}_{i1}$ ($\tilde{P}_{i1}$ is the liquidation value of the ith firm at period 1, as determined by the CAPM with homogeneous expectations), one can derive the various heterogeneous-market parameters as follows:

$$\sigma_i^{*2} = \text{var}\left(\tilde{P}_{i1}\right)/P_{i0}^{*2}, \sigma_{ij}^* = \text{cov}\left(\tilde{P}_{i1}, \tilde{P}_{j1}\right)/P_{i0}^* P_{j0}^*$$

and

$$\beta_i^* = \frac{x_i^* \sigma_i^{*2} + \sum_{\substack{j=1 \\ j \neq i}}^{n} x_j^* \sigma_{ij}^*}{\sum_{i=1}^{n} x_i^{*2} \sigma_i^{*2} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} x_i^* x_j^* \sigma_{ij}^*} \tag{10.6}$$

where x_i^* is the fraction of the total of all investors' wealth invested in the ith stock given by

$$x_i^* = \frac{P_{i0}^* N_i}{\sum_{j=1}^{n} P_{j0}^* N_j} = \frac{P_{i0}^* N_i}{w_0} \tag{10.7}$$

where n is the number of risky assets, and $w_0 = \sum_{k=1}^{K} T_k$ is the total wealth of all investors, which is fixed by assumption (a change in this assumption will not alter the basic heterogeneous-market results).

Note that the difference between β_i^* (see Eq. (10.6)) and the CAPM beta, β_i, (see Eq. (10.3)) is that here we use the heterogeneous beliefs parameter (hence the superstars added) and the investment proportions x_i^* rather than x_i. If equilibrium prices happen to be equal in the homogenous and the heterogeneous cases, then $\sigma_{ij}^* = \sigma_{ij}$ and $\sigma_i^{*2} = \sigma_i^2$. Also, in this specific case $\beta_i^* = \beta_i$ because $x_i^* = x_i$. In the trivial case where all $\varepsilon_{ki} = 0$, we are back in the CAPM framework and obviously $\beta_i^* = \beta_i$.

At this point we would like to emphasize the differences between (a) CAPM parameters; (b) the subjective parameters, which may differ from one investor to another, and (c) market parameters in a market with heterogeneous beliefs. The CAPM parameters are $\underline{\mu}$ and Σ, from which P_{i0} and $\tilde{P}_{i1}$ can be derived for each asset. With subjective beliefs, the kth investor has an estimate μ_{ik}, hence for a given Σ and $\underline{\mu}_{ik}$, a vector of (subjectively) optimal investment proportions, $\underline{x}_k$, is derived. Investor k's estimate of β_i, β_{ik}, depends on x_{ik} and therefore differs across investors. Each investor has his own subjective μ_{ik} and β_{ik}, as well as his subjective SML. However, when we aggregate the demand for each stock across all investors, an equilibrium price P_{i0}^* is determined from which we can derive μ_i^* and β_i^*. Thus, μ_i^* and β_i^* are the equilibrium market parameters with heterogeneous beliefs, to distinguish from the subjective parameters (which differ across investors in a heterogeneous beliefs market.) For goodness of fit of the CAPM to empirical data and for the analysis of the CAPM with homogenous and heterogeneous beliefs, only the equilibrium parameters are relevant. The subjective parameters are important only as means to derive the equilibrium parameters μ_i^*, and β_i^*. Thus, in the rest of this chapter we contrast the parameters (μ_i, β_i) with (μ_i^*, β_i^*), and the SML implied by the CAPM with the heterogeneous beliefs SML denoted by SML*.

10.2.2 Data and Results

The effect of deviation from the CAPM homogeneous expectations assumption on the risk-return relationship is analyzed next in the MS framework. We focus on the effects of the following parameters:

1. The number of investors, K
2. The intensity of the heterogeneity factor, ε
3. The number of securities available in the market n

In this chapter we focus on two market sizes: a market with $n = 5$ stocks and a market with $n = 20$ stocks. Each firm has an equal number of outstanding shares (i.e., $N_i = N_j$ for all i and j). Incorporating differences in the number of outstanding shares does not change our basic results.

Table 10.1 reports the μ and σ assumed for the assets included in the study, as well as the CAPM beta as calculated from the homogeneous expectation CAPM. Note that we selected μ and σ^2 in such a way that they are positively correlated with a correlation of about 40%. We may have stocks with a relatively low μ and large σ, but in general μ and σ tend to increase together, as found in most empirical studies (see Lintner, 1965b; Douglas, 1969; Levy, 1978). For simplicity of our analysis, all correlations are assumed to be pair-wise zero, hence Σ is diagonal matrix.[5] When all correlations are zero, one is tempted to believe that σ_i is almost a perfect proxy to β_i, and the CAPM is reduced to a trivial case when σ rather than β measures risk. This is not true because β_i is affected by σ_i as well as the investment proportions. To see this, note that the CAPM beta (see Eq. (10.3)) in the specific case of zero correlations reduces to

$$\beta_i = \frac{x_i \sigma_i^2}{\sum_{i=1}^{n} x_i^2 \sigma_i^2} \tag{10.8}$$

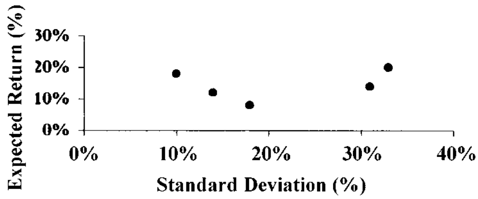
where x_i ($i = 1, 2 \ldots n$) are the homogeneous expectation CAPM investment proportions. If σ_i^2 is relatively large, x_i is generally small and β_i may also be small. Thus, even in the zero correlation case, we do not expect σ_i to be a perfect proxy to β_i. Indeed, in the $n = 5$ assets case (see Table 10.1), we find that $R^2\ (\sigma_i^2, \beta_i) = 0.40$, and in the $n = 20$ asset case, we have $R^2\ (\sigma_i^2, \beta_i) = 0.40$, which is very similar to what was found empirically (see, for example, Douglas, 1969, and Levy, 1978, who obtains $R^2 = 0.43$).

Given the expected return vector, $\underline{\mu}$, the variance-covariance matrix, Σ, and the interest rate, r, the optimal investment proportions x_i invested in the various assets are calculated. Relatively low investment proportions are allocated to assets with a low μ and high σ. To be more specific, in the zero correlation case we have the following relationship regarding the

[5] Results with a relaxation of this assumption are not much different.

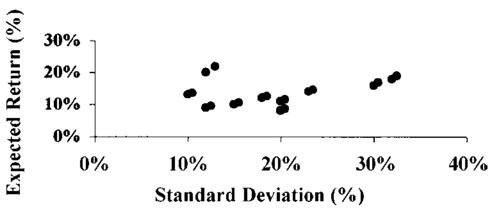
TABLE 10.1 The Parameters of Assets Corresponding to the Homogeneous Market

A market with 5 stocks



Stock no.	Expected return	Standard deviation	Beta
1	0.080	0.180	0.326
2	0.120	0.140	0.653
3	0.140	0.310	0.816
4	0.180	0.100	1.142
5	0.200	0.330	1.305

A market with 20 stocks



Stock no.	Expected return	Standard deviation	Beta
1	0.080	0.200	0.362
2	0.090	0.120	0.452
3	0.100	0.150	0.543
4	0.110	0.200	0.633
5	0.120	0.180	0.724
6	0.130	0.100	0.814
7	0.140	0.230	0.905
8	0.160	0.300	1.086
9	0.180	0.320	1.267
10	0.200	0.120	1.448
11	0.125	0.185	0.769
12	0.135	0.105	0.86
13	0.105	0.155	0.588
14	0.145	0.235	0.95
15	0.085	0.205	0.407
16	0.190	0.325	1.357
17	0.115	0.205	0.679
18	0.170	0.305	1.176
19	0.220	0.130	1.629
20	0.095	0.125	0.498

The correlation between the stock's beta and the stock's variance is 40%.
The correlation between the stock's variance and the stock's expected return is 40%.

investment proportions in any two assets i and j:

$$\frac{x_i}{x_j} = \frac{(\mu_i - r)/\sigma_i^2}{(\mu_j - r)/\sigma_j^2} \text{ (see Levy, 1973)} \tag{10.9}$$

Therefore, by the market clearance condition a relatively low equilibrium price, P_0, is determined for assets with low mean and high σ. (See the equilibrium pricing of the stocks in the homogeneous-expectations market with $n = 5$ stocks, as given in part b of Table 10.2.)

We begin to investigate the effects of heterogeneous expectations by first introducing a 10% heterogeneity factor. Namely, $\mu_{ik} = \mu_i(1 + \varepsilon_{ik})$, where ε_{ik} has a 10% standard deviation and a zero mean. To get a feeling for the meaning of 10% heterogeneity factor, recall that, if $\mu_i = 10\%$ and ε_{ik} deviates two standard deviations right to the mean (a very unlikely event), then μ_{ik} is 10% (1.2) = 12%. We see that the 10% heterogeneity factor implies a relatively low degree of heterogeneity, and investors actually do not have much different beliefs. Table 10.2 presents the MS results of 10 simulations, with five stocks available in the market, and a risk-free interest rate of 4%, for markets with various numbers of investors. Part a of the table presents the results of the regression given by Eq. (10.10):

$$\mu_i^* = \gamma_0 + \gamma_1 \beta_i^* + e_i \tag{10.10}$$

where (μ_i^*, β_i^*) are the actual parameters as determined in the heterogeneous-expectations market. A few conclusions can be drawn from the table.

1. γ_0 is generally lower than 4% and it is not always significant. This is in contrast to what has been found empirically—that is, an intercept greater than the risk free interest rate (see, for example, Miller and Scholes, 1972, and Levy, 1978).
2. γ_1 is always positive and highly significant.
3. The R^2 is very large even with a small number of investors. As the number of investors increases, the R^2 tends to increase reaching almost a perfect fit with an R^2 close to 1 for 10,000 investors.

Part b of the table reports the equilibrium prices of the 10th simulation and compares them with the homogeneous-market equilibrium prices. The results of the other simulations is very similar. We see that even with a small number of investors $k = 100$, $P_{i0} \cong P_{i0}^*$ and for $k = 10{,}000$ the CAPM with a 10% heterogeneity factor almost coincides with the Sharpe-Lintner CAPM with homogeneous expectations. Yet, γ_0 is still lower than 4% in most simulations.

TABLE 10.2 The MS Results: A Market with 5 Stocks and Heterogeneity Factor of 10%

a. The regression results

Simulation no.	100 Investors					1,000 Investors					10,000 Investors				
	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2
1	0.034	1.850	0.132	6.637	0.936	0.027	3.911	0.131	17.092	0.990	0.030	8.134	0.130	31.801	0.997
2	−0.010	−0.584	0.169	8.875	0.963	0.025	3.548	0.135	17.243	0.990	0.031	5.907	0.128	22.081	0.994
3	0.051	4.872	0.108	9.494	0.968	0.043	4.518	0.119	11.538	0.978	0.030	4.233	0.129	16.738	0.989
4	0.050	2.543	0.102	4.706	0.881	0.026	2.125	0.133	10.130	0.972	0.029	4.664	0.129	18.797	0.992
5	0.062	2.448	0.105	3.827	0.830	0.016	1.484	0.141	11.942	0.979	0.029	4.477	0.129	18.207	0.991
6	0.001	0.035	0.154	6.966	0.942	0.041	6.250	0.118	16.639	0.989	0.031	5.682	0.128	21.150	0.993
7	0.064	2.544	0.105	3.846	0.831	0.028	3.743	0.130	15.903	0.988	0.033	6.695	0.126	23.336	0.995
8	0.024	1.033	0.127	4.901	0.889	0.029	3.649	0.129	15.013	0.987	0.028	5.009	0.130	20.979	0.993
9	0.007	0.585	0.148	10.714	0.975	0.030	4.166	0.128	16.237	0.989	0.030	6.324	0.129	24.837	0.995
10	0.052	2.898	0.112	5.753	0.917	0.031	6.847	0.130	26.516	0.996	0.032	5.809	0.127	21.289	0.993
Average	0.034				0.913	0.030				0.986	0.030				0.993

b. The stocks equilibrium price, P_{i0}^{*}

Stock no.	Homogeneous market	Heterogeneous market: 100 Investors	1,000 Investors	10,000 Investors
1	2.828	2.812	2.844	2.849
2	9.350	9.148	9.388	9.375
3	2.384	2.413	2.404	2.402
4	32.072	32.278	31.994	31.988
5	3.366	3.349	3.370	3.386

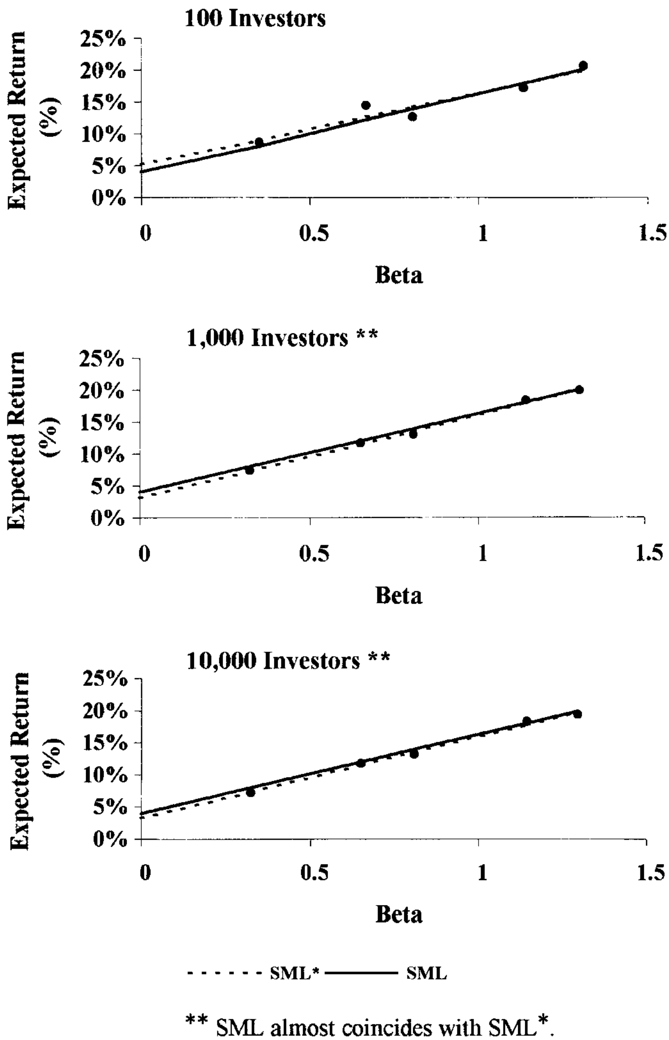


FIGURE 10.1 The SML* in heterogeneous market: A market with five stocks and heterogeneity factor of 10%.

Figure 10.1 reveals the SML and the heterogeneous expectation SML* with market data as determined in a heterogeneous market (hence the superstar is added), for the 10th simulation (the other simulations, once again, yield very similar results). Given μ, Σ, and r, we first derive the CAPM SML given by the solid line. This line crosses the vertical axis exactly at 4% and if we would have no heterogeneity, all five points, by the CAPM construction, would fall exactly on the SML. With heterogeneous expectations, P_{i0} may be different from P_i^*, hence we may find that $\mu_i^* \neq \mu_i$ and $\beta_i^* \neq \beta_i$, and because μ_i^* and β_i^* are the market parameters with heterogeneous expectation, the dots corresponding to (μ_i^*, β_i^*) do not fall necessarily on the solid straight line representing the SML. We then run regression (10.10) to draw the line SML*, which is the best fit straight line corresponding to the values (μ_i^*, β_i^*). As we can see, SML* with 10,000 investors is almost perfect. Therefore, the Sharpe-Lintner CAPM is almost intact when a 10% heterogeneity factor is allowed.

Why do we get an almost perfect fit, with SML ≅ SML*, in a large market ($K = 10{,}000$)? While it is true that generally for each investor $x_{ik} \neq x_i$, when the number of investors is very large, the deviations of x_{ik} from x_i offset each other and their net effect on prices is negligible.[6] With a very large K the total demand for the ith stock is almost the same with homogeneous and heterogeneous expectations. This implies that $P_{i0} \cong P_{i0}^*$; hence the SML* almost coincides with the SML.

Table 10.3 and Figure 10.2 show the MS results with 30% heterogeneity factor (Recall that if $\mu_i = 10\%$, this implies that there is about .95 probability that the mean will be covered by the bounds $4\% - 16\%$.) With the relatively large heterogeneity factor of 30%, we obtain many negative

[6] To see this, recall that in the zero correlation case $x_{ik} = \dfrac{\dfrac{\mu_{ik} - r}{\sigma_i^2}}{\sum_{j=1}^{n} \dfrac{\mu_{jk} - r}{\sigma_j^2}}$. Thus, x_{ik} is approximately linear in μ_{ik} (the relation is not precisely linear, because μ_{ik} also appears in the denominator; however, when the number of assets in the market, n, becomes large, the effect of μ_{ik} on the denominator becomes relatively small, and the relation becomes approximately linear). For simplicity, let us assume a linear relationship: $x_{ik} = A\mu_{ik} + B$. Similarly, the investment proportion in the ith stock in the homogeneous expectation case can be written as $x_i = A\mu_i + B$. The difference between P_{i0}^* and P_{i0} is given by

$$\Delta P_i \equiv P_{i0}^* - P_{i0} = \frac{1}{N_i} \sum_{k=1}^{K} T_k (x_{ik}^* - x_i) = \frac{A}{N_i} \sum_{k=1}^{K} T_k (\mu_{ik} - \mu_i) = \frac{A}{N_i} \sum_{k=1}^{K} T_k \varepsilon_{ik}$$

In our case, $T_k = \dfrac{w_0}{K}$, and therefore

$$\Delta P_i = \frac{A w_0}{N_i K} \sum_{k=1}^{K} \varepsilon_{ik}$$

The variance of ΔP_i is given by

$$\mathrm{Var}(\Delta P_i) = \left(\frac{A w_0}{N_i K}\right)^2 \mathrm{Var}\left(\sum_{k=1}^{K} \varepsilon_{ik}\right) = \left(\frac{A w_0}{N_i K}\right)^2 K \sigma^2 = \frac{1}{K}\left(\frac{A w_0 \sigma}{N_i}\right)^2$$

(recall that the ε_{ik}'s are uncorrelated). As the number of investors, K, increases, $\Delta P_i \to 0$, and the prices in the heterogeneous-expectations market converge to the homogeneous market prices. This result also holds for a nonuniform distribution of wealth.

TABLE 10.3 The MS Results: A Market with 5 Stocks and Heterogeneity Factor of 30%

a. The regression results

Simulation no.	100 Investors					1,000 Investors					10,000 Investors				
	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2
1	−0.094	−0.861	0.198	1.564	0.449	−0.017	−0.388	0.152	3.191	0.772	−0.047	−0.884	0.176	2.956	0.744
2	−0.056	−1.126	0.226	4.222	0.856	−0.043	−0.851	0.174	3.029	0.754	−0.039	−0.788	0.170	3.019	0.752
3	0.026	0.331	0.102	1.164	0.311	−0.032	−0.555	0.160	2.451	0.667	−0.054	−0.988	0.182	2.972	0.746
4	−0.051	−0.805	0.174	2.425	0.662	−0.066	−0.796	0.183	1.924	0.552	−0.032	−0.652	0.163	2.942	0.743
5	0.077	2.121	0.070	1.728	0.499	−0.010	−0.200	0.142	2.585	0.690	−0.042	−0.729	0.168	2.607	0.694
6	0.097	1.252	0.021	0.242	0.019	−0.047	−0.911	0.177	3.024	0.753	−0.055	−1.078	0.185	3.197	0.773
7	0.022	0.251	0.096	0.942	0.228	−0.007	−0.235	0.150	4.379	0.865	−0.053	−0.883	0.178	2.617	0.695
8	0.069	0.993	0.099	1.306	0.363	−0.069	−0.988	0.189	2.380	0.654	−0.036	−0.752	0.168	3.093	0.761
9	0.173	1.314	−0.042	−0.284	0.026	−0.046	−0.799	0.172	2.634	0.698	−0.033	−0.656	0.163	2.867	0.733
10	−0.079	−1.645	0.214	0.948	0.839	−0.039	−0.685	0.166	2.596	0.692	−0.051	−0.837	0.176	2.531	0.681
Average	0.018				0.425	−0.038				0.710	−0.044				0.732

b. The stocks equilibrium price, P_{i0}^*

Stock no.	Homogeneous market	Heterogeneous market		
		100 Investors	1,000 Investors	10,000 Investors
1	2.828	3.147	3.020	3.023
2	9.350	9.810	9.719	9.737
3	2.384	2.497	2.539	2.615
4	32.072	30.977	31.073	30.993
5	3.366	3.570	3.649	3.631

γ_0 and one negative γ_1, a relatively low R^2 (and even with 10,000 investors the R^2 remains about 70%). Similarly, part b of table 10.2 reveals relatively large price deviations of P_{i0}^* from P_{i0} even for a large number of investors. The relatively low R^2 and the deviation of the SML* from the SML are clearly observed from Figure 10.2, which corresponds to the 10th simulation.

A comparison of Tables 10.2 and 10.3 (and Figures 10.1 and 10.2) sheds light on the importance of the degree of heterogeneity. When there

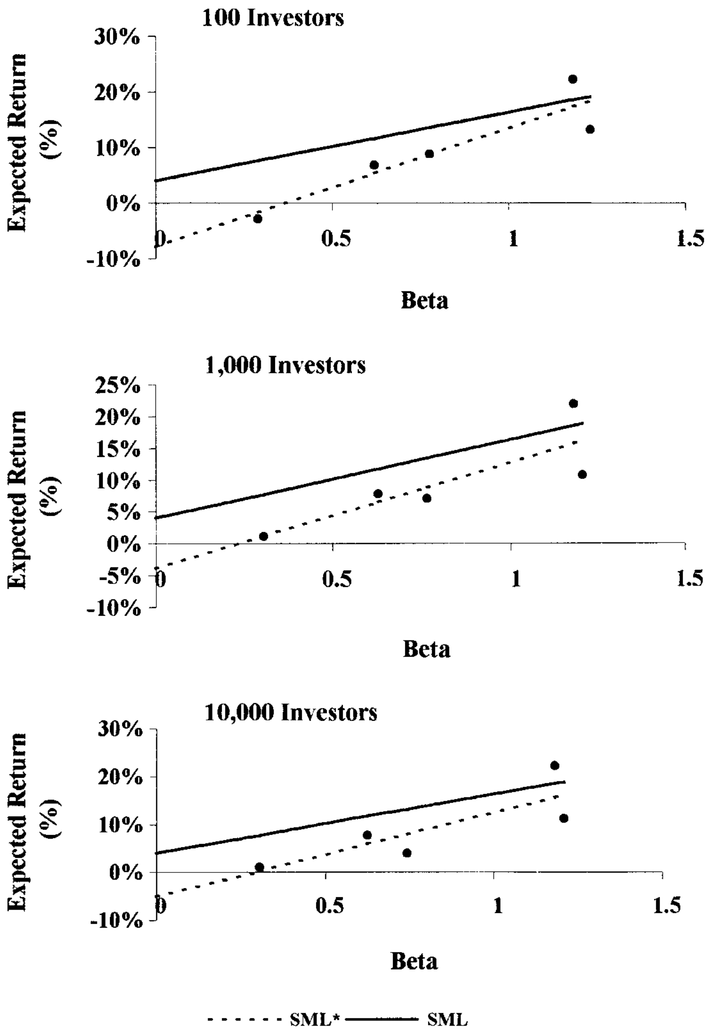


FIGURE 10.2 The SML* in heterogeneous market: A market with 5 stocks and heterogeneity factor of 30%.

is a relatively large disagreement in the market (30% heterogeneity factor), the SML has a relatively poor performance and in some cases β_i^* cannot even serve as a good risk index because γ_1 is sometimes found to be negative. Also, the gap between the SML and SML* increases as the heterogeneity increases from 10% to 30%, and the SML* even shows a negative intercept, γ_0. In all cases, $\gamma_0 < 4\%$ in contradiction to the empirical finding. Figure 10.2 reveals a systematic bias in pricing: lower expected return stocks tend to be overpriced in the heterogeneous expectations market (although the heterogeneity is perfectly symmetric). In contrast, higher expected return stocks tend to be underpriced. The explanation for this effect is due to the nonlinear effect of ε_{ik} on x_{ik}. However, as we shall see, the magnitude of this effect decreases as the number of assets in the market grows. This systematic pricing bias and the explanation for its decrease as the number of assets grows are described in Appendix 10.1.

Another result we find is related to the number of investors: with a 10% heterogeneity factor there is only a small tendency of the R^2 to increase as the number of investors increases (from about 91% to above 99%) and the fit is almost perfect even with 100 investors. This is not the case with a 30% heterogeneity factor. With a large heterogeneity factor (30%), many large deviations occur, and for $k = 100$ investors, the R^2 on average is only about 42%. As before, R^2 tends to increase with the increase in the number of investors, reaching about 73% with $k = 10{,}000$ investors. Relative to the case of a 10% heterogeneity factor, a much larger number of investors is required in order to obtain high values of R^2.

In an actual market, the number of assets is very large. One may suspect that the obtained results with a small number of assets $n = 5$ is not intact for a market with a large number of assets. This is not the case. In fact, a larger number of assets only increases the robustness of the CAPM to relaxation of the homogeneous-expectations assumption. For low degree of heterogeneity (10%), the results for $n = 5$ and for $n = 20$ are very similar, and for a 30% heterogeneity factor, a better fit between SML and SML* is obtained as n increases. Let us elaborate on the effect of the increase in the number of securities available in the market on the risk-return equilibrium relationship. Tables 10.4 and 10.5 and Figures 10.3 and 10.4 summarize the results with $n = 20$ stocks.

1. With a 10% heterogeneity factor, as with $n = 5$ stocks, we have also with $n = 20$ stocks almost a perfect fit, with SML* $\cong$ SML. Once again, as K increases, R^2 tends to increase. All γ_1 are positive and highly significant. Thus, the increase in the number of stocks, even to 20, slightly improves the fit between SML and SML*, which is almost perfect even for $n = 5$ (see Figure 10.3).

TABLE 10.4 The MS Results: A Market with 20 Stocks and Heterogeneity Factor of 10%

a. The regression results

Simulation no.	100 Investors					1,000 Investors					10,000 Investors				
	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2	γ_0	T Statistic	γ_1	T Statistic	R^2
1	0.041	5.261	0.111	13.237	0.907	0.037	13.061	0.113	37.064	0.987	0.038	43.055	0.112	117.368	0.999
2	0.047	4.031	0.108	8.547	0.802	0.038	14.737	0.113	40.859	0.989	0.036	40.357	0.114	116.895	0.999
3	0.040	5.133	0.111	13.284	0.907	0.039	15.324	0.111	40.260	0.989	0.038	60.720	0.112	165.958	0.999
4	0.016	2.195	0.135	16.891	0.941	0.040	11.213	0.110	28.811	0.979	0.036	33.863	0.114	99.727	0.998
5	0.030	4.037	0.119	14.908	0.925	0.030	9.628	0.120	36.142	0.986	0.035	44.035	0.115	132.625	0.999
6	0.034	3.757	0.119	12.430	0.896	0.034	16.455	0.116	51.951	0.993	0.035	39.871	0.115	121.574	0.999
7	0.035	3.984	0.113	11.935	0.888	0.030	8.325	0.118	30.014	0.980	0.037	31.093	0.113	88.255	0.998
8	0.028	3.447	0.123	14.056	0.917	0.037	18.360	0.114	52.095	0.993	0.036	36.560	0.114	106.706	0.998
9	0.047	3.903	0.105	8.101	0.785	0.032	11.068	0.119	38.182	0.988	0.036	40.511	0.114	119.229	0.999
10	0.030	4.934	0.119	18.200	0.948	0.039	15.138	0.110	40.269	0.989	0.038	35.521	0.112	98.464	0.998
Average	0.035				0.892	0.036				0.987	0.037				0.999

b. The stocks equilibrium price, P_{i0}^*

Stock no.	Homogeneous market	Heterogeneous market: 100 Investors	Heterogeneous market: 1,000 Investors	Heterogeneous market: 10,000 Investors	Stock no.	Homogeneous market	Heterogeneous market: 100 Investors	Heterogeneous market: 1,000 Investors	Heterogeneous market: 10,000 Investors
1	2.793	2.764	2.777	2.801	11	6.936	7.006	6.949	6.958
2	9.697	9.787	9.691	9.688	12	24.065	24.158	24.079	24.060
3	7.448	7.432	7.503	7.445	13	7.556	7.581	7.570	7.571
4	4.888	4.872	4.884	4.893	14	5.310	5.383	5.354	5.317
5	6.896	7.084	6.930	6.907	15	2.991	3.013	2.982	3.006
6	25.136	25.331	25.085	25.105	16	3.966	3.912	3.963	3.974
7	5.280	5.245	5.303	5.293	17	4.984	5.030	5.010	4.996
8	3.724	3.751	3.721	3.721	18	3.903	3.947	3.928	3.909
9	3.818	3.817	3.844	3.821	19	29.746	29.572	29.603	29.696
10	31.032	30.548	30.982	31.021	20	9.831	9.768	9.840	9.818

TABLE 10.5 The MS Results: A Market with 20 Stocks and Heterogeneity Factor of 30%

a. The regression results

Simulation no.	100 Investors γ_0	100 Investors *T* Statistic	100 Investors γ_1	100 Investors *T* Statistic	100 Investors R^2	1,000 Investors γ_0	1,000 Investors *T* Statistic	1,000 Investors γ_1	1,000 Investors *T* Statistic	1,000 Investors R^2	10,000 Investors γ_0	10,000 Investors *T* Statistic	10,000 Investors γ_1	10,000 Investors *T* Statistic	10,000 Investors R^2
1	−0.019	−0.821	0.169	7.014	0.732	0.012	1.023	0.132	10.193	0.852	0.023	5.109	0.122	24.629	0.971
2	0.003	0.090	0.152	4.434	0.522	0.005	0.638	0.141	15.848	0.933	0.022	4.458	0.123	23.296	0.968
3	−0.018	−0.751	0.151	5.924	0.661	0.018	1.938	0.131	13.450	0.910	0.024	4.744	0.121	22.517	0.966
4	0.042	2.098	0.105	4.827	0.564	0.023	2.593	0.120	12.272	0.893	0.021	3.804	0.123	20.536	0.959
5	−0.038	−2.064	0.180	9.146	0.823	0.028	3.237	0.120	12.863	0.902	0.026	6.343	0.121	27.536	0.977
6	0.011	0.453	0.138	5.414	0.620	0.007	0.587	0.132	9.707	0.840	0.022	4.529	0.123	22.930	0.967
7	−0.001	−0.065	0.152	6.688	0.713	0.016	1.862	0.133	14.390	0.920	0.023	4.918	0.123	24.653	0.971
8	0.018	0.803	0.135	5.786	0.650	0.028	2.513	0.113	9.258	0.826	0.019	3.592	0.126	22.240	0.965
9	0.034	1.627	0.105	4.659	0.547	0.018	2.579	0.128	16.638	0.939	0.027	5.527	0.119	22.282	0.965
10	0.033	1.168	0.123	4.086	0.481	0.007	0.911	0.137	16.123	0.935	0.019	4.826	0.127	30.416	0.981
Average	0.007				0.631	0.016				0.895	0.023				0.969

b. The stocks equilibrium price, P_{i0}^*

Stock no.	Homogeneous market	Heterogeneous market 100 Investors	Heterogeneous market 1,000 Investors	Heterogeneous market 10,000 Investors	Stock no.	Homogeneous market	Heterogeneous market 100 Investors	Heterogeneous market 1,000 Investors	Heterogeneous market 10,000 Investors
1	2.793	2.523	2.860	2.847	11	6.936	6.670	6.937	6.984
2	9.697	9.584	9.836	9.797	12	24.065	27.296	24.015	23.907
3	7.448	7.648	7.703	7.520	13	7.556	7.260	7.604	7.635
4	4.888	5.086	4.998	4.905	14	5.310	5.309	5.357	5.338
5	6.896	6.756	6.967	6.959	15	2.991	2.974	3.003	3.052
6	25.136	24.324	25.376	25.114	16	3.966	4.049	4.028	4.004
7	5.280	5.156	5.388	5.335	17	4.984	5.150	5.101	5.038
8	3.724	3.741	3.843	3.779	18	3.903	3.992	3.976	3.913
9	3.818	3.793	3.774	3.843	19	29.746	28.642	29.108	29.380
10	31.032	29.729	30.151	30.693	20	9.831	10.317	9.975	9.957

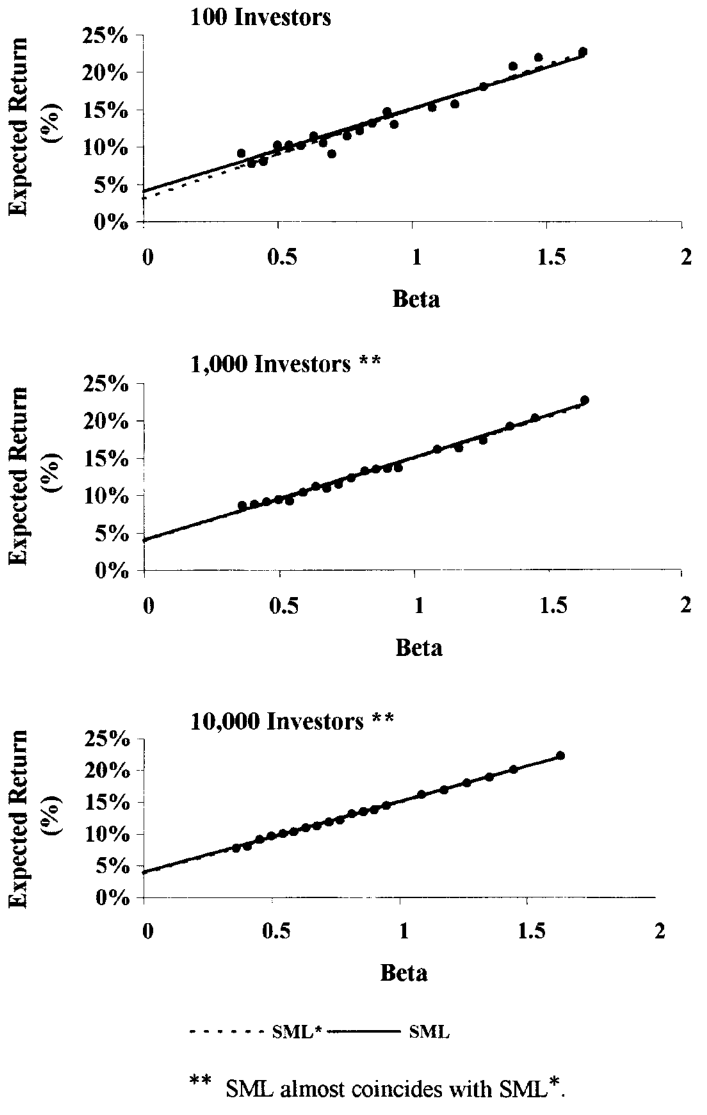


FIGURE 10.3 The SML* in heterogeneous market: A market with 20 stocks and heterogeneity factor of 10%.

2. With a 30% heterogeneity factor, which represents quite a large disagreement between investors, we obtain the following strong results. First, all γ_1 are positive and highly insignificant. For the realistic market with many investors, all γ_1 are highly significant and the R^2 is about 0.97 (i.e., an almost perfect fit is obtained). In well-developed markets with

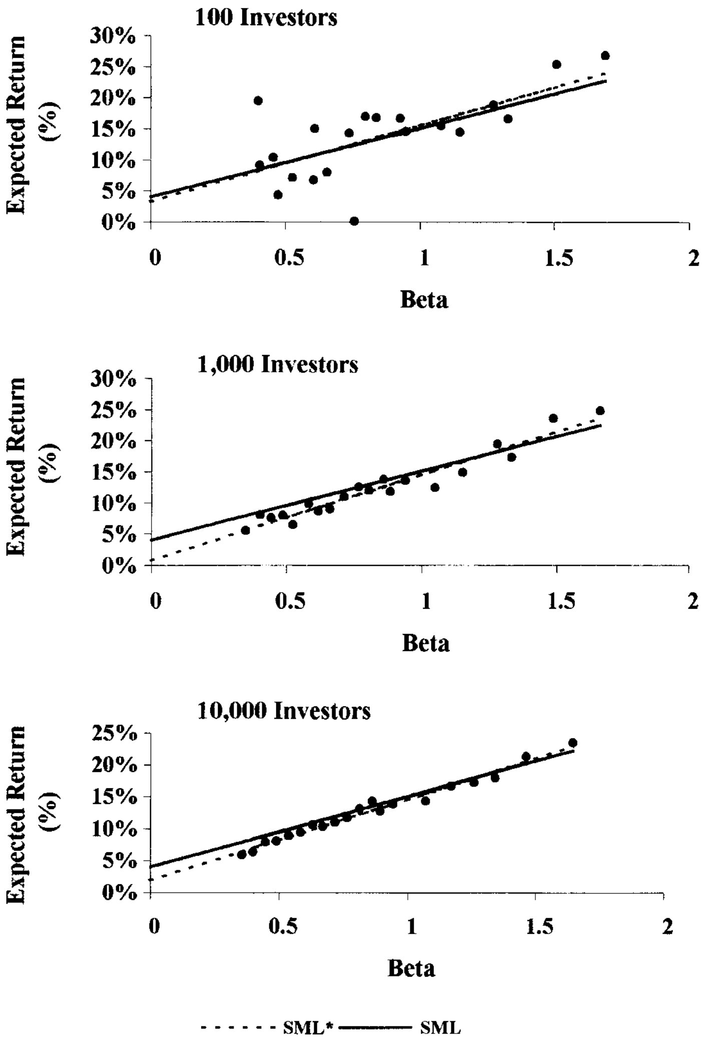


FIGURE 10.4 The SML* in heterogeneous market: A market with 20 stocks and heterogeneity factor of 30%.

many investors and many securities, the difference between SML and SML* is very small and even with a 30% heterogeneity factor there is almost 100% correlation, with SML $\cong$ SML*, quite a strong result. Most γ_0 are still smaller than 4%, but they are higher relative to the $n = 5$ case. This is in accordance with the result of Appendix 10.1, implying that $\gamma_0 \rightarrow r$ as $n \rightarrow \infty$.

A comparison between Tables 10.3 and 10.5 reveals the following results. First, as n increases, the R^2 tends to increase. Second, γ_0 tends to increase, in most cases it is positive and in better agreement with the CAPM intercept of $r = 4\%$. Finally, the SML* is much more close to the SML with $n = 20$ relative to the case $n = 5$ (compare Figures 10.3 and 10.4).

To sum up, in a large market with many investors and relatively many securities, the R^2 between μ_i^* and β_i^* is almost perfect even with 30% heterogeneity factor. What are the implications of the previous MS to the Sharpe-Lintner CAPM? We have analyzed markets with $n = 5$ assets and markets with $n = 20$ assets. However, in reality, markets are very large and the case of $n = 5$ is irrelevant. As even with only $n = 20$ we get almost a perfect fit, we can safely assume that in realistic markets with thousands of assets and millions of investors, the fit to the SML would be even better. Thus, in the most relevant case—that is, a large number of investors and relatively large number of assets—we obtain that SML $\cong$ SML* even with 30% heterogeneity factor. To understand these strong results, suppose that the mean return of the ith security is $\mu_i = 10\%$ and investors deviate up to one standard deviation to each side (actually the deviation can be larger). Then with a 30% heterogeneity factor investors' beliefs are distributed in the range (7–13%). Each investor has a different mean μ_{ik}, but still the Sharpe-Lintner CAPM is almost intact—quite a strong result!

10.3. NUMBER OF ASSETS IN THE PORTFOLIO: THE GCAPM VERSUS THE CAPM

10.3.1 The Model

In the CAPM framework with homogeneous or heterogeneous beliefs as discussed, each investor holds all available assets (i.e., each of the thousands of assets actually available should be theoretically held by each of the investors). In practice, investors commonly hold a relatively small number of assets in their portfolios. How does this affect the results of the CAPM, which implicitly assumes that all investors can hold all available risky assets? Levy (1978), Merton (1987), Markowitz (1990), and Sharpe (1991) analyze this question theoretically. The equilibrium model in which investors hold a limited number of stocks is known as the general CAPM (GCAPM) or *segmented* market model. In this chapter we employ MS to analyze equilibrium prices in a segmented market. We assume that the kth investor selects n_k stocks at random from the n available risky assets. The investor then constructs the efficient set of these n_k stocks. Given the riskless interest rate, r, the investor selects the optimum investment proportions in these stocks. By aggregating all the individual demands for each of the stocks, we derive the equilibrium prices P_{i0}. We consider two possible alternatives for selecting the n_k stocks from the n available assets:

1. Each stock has the same chance to be selected.
2. Some stocks have a lower chance to be selected hence they fall in the category of *neglected stocks*, which are also known in the financial literature as *small stocks*. We will elaborate on these two alternatives later in the chapter.

In this section we assume that $\underline{\mu}$ and Σ are given, and all investors agree on these parameters.[7] We employ MS to find out the relationship between equilibrium prices, the CAPM and the GCAPM parameters, as well as the possible justification for small firm effects in the GCAPM framework. In the GCAPM framework, there are n assets available in the market with known $\underline{\mu}$ and Σ, and the kth investor holds $n_k < n$ assets. In general, each investor may hold a different number of stocks. In this chapter, for the sake of simplicity, we assume that $n_k = n_0 < n$ for all investors. We first analyze the case in which each asset has an equal chance to be included in the individual investor's portfolio. For example, if $n = 10$ and $n_0 = 4$, each asset has an equal probability of 0.4 to be included in each portfolio. When the number of investors is very large, each asset will be included in exactly 40% of the portfolios.

When the number of investors is not very large, even with this equal selection probability model we will have a random element. To focus on the comparison of the classic CAPM and GCAPM we will assume homogeneous beliefs (i.e., $\varepsilon_{ik} = 0$ for all i, k), and a large number of investors. Thus, we have a model in which all investors agree on $\underline{\mu}$ and Σ, and it is exactly known how many portfolios include each of the assets.

The CAPM Derivation

The CAPM equilibrium derivation was discussed earlier. Let us remind the reader of the CAPM derivation. Each investor holds all n assets. We first derive the M-V efficient frontier composed of all n assets. Then, for the given riskless interest rate we derive the market portfolio with the optimal investment proportions $x_i(1 = 1, 2, \ldots n)$. As before, we assume for simplicity that all correlations are zero, hence we calculate β_i by Eq. (10.8) given earlier.

As we deal here with *ex ante* parameters and no parameter estimation is involved, we obtain precisely the CAPM. Using $\underline{\mu}$, Σ, r, and β_i, we

[7] To keep all μ and Σ unchanged, if the equilibrium prices change, the end-of-period values must also be changed.

obtain a perfect linear relation between μ_i and β_i—the well-known Sharpe-Lintner SML. The line crosses the vertical axis exactly at r.

The GCAPM Derivation

We first construct all $\binom{n}{n_0}$ portfolios, (i.e., $\frac{n!}{(n-n_0)!n_0!}$ portfolios) thus we implicitly assume a very large number of investors, because each asset has an equal chance to be included in each portfolio. For example, with $n = 3$ (assets A, B, and C) and $n_0 = 2$, we consider the three portfolios composed of assets A, B, and C as follows: (A, B), (A, C), (B, C). Generally, where n_0 assets are selected out of n assets, for each portfolio we employ the relevant elements of the CAPM parameters, $\underline{\mu}$ and Σ, to derive the "little CAPM" that is, the kth investor's efficient frontier composed of the n_0 assets in her portfolio and the kth investor investment proportions in each of these assets, denoted by x_{ik} $(i = 1, 2, \ldots n_0)$. The selected portfolio has a mean μ_k and variance σ_k^2. The aggregate demand for each asset determines the equilibrium prices, hence the market portfolio and the risk measure of each asset. As we shall see, in the segmented market the CAPM linear relationship between return and risk does not hold.

The CAPM and GCAPM Market Portfolios

One may suspect that when the number of investors is very large (hence there is no random element in the choice of the stocks), and when each asset has an equal chance to be included in each portfolio, the CAPM and GCAPM market portfolios will coincide. This is not the case, as will be shown. To see this, suppose that the kth investor holds one of the $\binom{n}{n_0}$ portfolios. Then for each portfolio we first derive x_{ik} (i.e., the kth investor's investment proportion in the ith asset). Then, for all investors we sum up the demand for the ith asset as follows:

$$\sum_{k=1}^{K} T_k x_{ik}$$

(If the ith asset is not held by the kth investor, we have $x_{ik} = 0$.) Dividing the total amount of dollars invested in the ith asset by the total wealth invested in the market, w_0, we obtain the proportion of the ith asset in the market portfolio as follows:

$$x_i^* = \sum_{k=1}^{K} T_k x_{ik} / w_0 \tag{10.11}$$

This may be much different from the CAPM proportion, x_i, which is derived from the efficient set composed of all n assets available.

Let us illustrate with a simple example of $n = 3$ assets and $n_0 = 2$. Suppose we have three assets denoted by A, B, and C, and suppose that the parameters Σ and $\underline{\mu}$ are such that the CAPM optimal investment proportions are $x_A = 0.\overline{6}$, $x_B = 0.2$ and $x_c = 0.2$. Assuming zero correlations (for simplicity only) we have

$$\frac{x_A}{x_B} = \frac{(\mu_A - r)/\sigma_A^2}{(\mu_B - r)/\sigma_B^2} = \frac{0.6}{0.2} = 3$$

Similarly,

$$\frac{x_A}{x_C} = \frac{(\mu_A - r)/\sigma_A^2}{(\mu_C - r)/\sigma_C^2} = 3$$

and

$$\frac{x_B}{x_C} = \frac{(\mu_B - r)/\sigma_B^2}{(\mu_C - r)/\sigma_C^2} = 1$$

(For the relationship between the investment proportions and the assets' parameters in the zero correlation case, see Levy, 1973.) Suppose that there are three investors in the market, each investing \$400. Thus, if \$1200 are invested in this CAPM market, a simple calculation reveals that the market portfolio will be composed of V_A = \$720, V_B = \$240, and V_C = \$240, where V stands for the value of the firm.

Let us turn to the GCAPM market values, given the same parameters $\underline{\mu}$, Σ, and r. For a very large number of investors and an equal chance of each asset to be included in each portfolio, exactly \$400 will be invested in portfolio (A, B), \$400 in portfolio (A, C), and \$400 in portfolio (B, C). Alternatively, assume that each of our three investors invests in one of the two-asset portfolios. Then, in the GCAPM framework we have

Portfolio (A, B): $\frac{X_A}{X_B} = \frac{0.6}{0.2} = 3$, hence $\$400 \cdot 3/4 = \$300$ will be invested in A and \$100 in A asset B.

Portfolio (A, C): $\frac{X_A}{A_c} = 3$, hence once again \$300 will be invested in A and \$100 in C.

Portfolio (B, C): $\frac{X_B}{X_C} = 1$, hence \$200 will be invested in B and \$200 in C.

Therefore, with the GCAPM we have the following market values:

$$V_A = \$600, V_B = \$300, V_C = \$300$$

These are quite different from the CAPM market values. Hence, the GCAPM market portfolio generally differs from the CAPM market portfolio, even if the number of investors is large, and each stock has the same probability of being selected.

To understand this result, suppose that we observe the same set of $\underline{\mu}$, Σ, and r in both the CAPM and GCAPM frameworks. If all investors hold all assets as implied by the CAPM, we obtain one set of market values and hence a market portfolio, and if the market is segmented (with $n_0 < n$ assets in each portfolio), we obtain another set of market values with another market portfolio. Obviously, the GCAPM market portfolio must be interior to the M-V efficient frontier composed of all assets available in the market. In reality, we have segmented markets. Thus, this finding may explain why actual indexes like the S & P, NYSE index, and the Dow Jones are empirically always found to be interior to the M-V efficient frontier.

As the CAPM and GCAPM market portfolios differ, beta will also not be the same in these two equilibrium frameworks. To see this, denote by x_i the CAPM market investment proportions in the ith asset, and by x_i^* the aggregate investment proportion in the ith asset as found in the GCAPM framework (see Eq. (10.11)). Having x_i^* we can calculate the assets' beta, calculated with the market portfolio as implied by the segmented GCAPM market (in contrast to the CAPM tangency market portfolio). Thus, we have

$$\beta_i = \frac{x_i \sigma_i^2}{\sum_{i=1}^{n} x_i^2 \sigma_i^2} \quad \text{and} \quad \beta_i^G = \frac{x_i^* \sigma_i^2}{\sum_{i=1}^{n} x_i^{*2} \sigma_i^2} \tag{10.12}$$

where G indicates that the market portfolio employed to calculate beta is the GCAPM market portfolio. For $x_i = x_i^*$ we have $\beta_i = \beta_i^G$.

In this framework we do not expect the pairs (μ_i, β_i^G) to be on a straight line as in the CAPM. However, we can run the regression,

$$\mu_i = \gamma_0 + \gamma_1 \beta_i^G + e_i \tag{10.13}$$

and compare it to the CAPM straight line,

$$\mu_i = r + (\mu_m - r)\beta_i$$

and analyze the effect of market segmentation on the SML. This may provide a clue regarding the results of the empirical tests of the CAPM.

Small Firm Effect

So far we have assumed that each asset has an equal chance to be included in each portfolio. We next examine the small firm or neglected stock effect. To demonstrate, suppose that there are four assets, A, B, C, and D. Suppose that asset D is a neglected stock and it has a smaller chance to be included in the portfolio. For instance, suppose that we form all possible portfolios from assets A, B, and C: (A, B), (A, C), (B, C), but asset D is included only in one portfolio, say the first portfolio, hence creating portfolio (ABD). Given the portfolios (ABD), (AC), (BC), we repeat the same analysis given earlier to calculate β_i^G, which allows us to analyze the effect of the neglected stock assumption on equilibrium prices and the risk-return relationship. By forming various portfolios, we can create situations where the neglected stock or small stock has various probabilities to be included in a portfolio.

10.3.2 The MS Segmented Market Results

Table 10.6 provides the main results regarding the effect of market segmentation on the risk-return relationship. We assume that there are $n = 10$ available assets, with μ and σ^2 given by part a of the Table. The values μ and σ^2 are selected such that the correlation, R^2, between μ_i and σ_i^2 of the individual assets is 0.46, a figure in the neighborhood of what is found in the empirical studies.

We create all $\binom{10}{4} = \frac{10!}{6!4!} = 210$ portfolios, and for each portfolio we calculate x_{ik}. For each asset we calculate the market proportion x_i^*, as explained earlier (see Eq. (10.11)). Using x_i^* we calculate the observed ith stock's beta, β_i^G. We run the regression:

$$\mu_i = \gamma_0 + \gamma_1 \beta_i^G + e_i$$

As can be seen from Table 10.6, we obtain R^2 of 0.905.[8]

To better understand the results of Table 10.6, recall that if the market would not have been segmented we would have obtained an R^2 of 1, with all the dots exactly on the same straight line, the SML. Because the market is segmented, we obtained deviations from the SML, and the R^2 is reduced to .905. It is interesting to note that the segmentation induces the intercept, γ_0, to be *larger* than $r = 4\%$ (unlike the heterogeneous-expectation case with $n_0 = n$). This is exactly as found in most empirical studies. A theoretical explanation for the fact that in the GCAPM framework the intercept is larger than the risk-free rate is given in Levy (1978).

[8] Recall that we are considering all possible portfolios (in accordance with the assumption of a very large number of investors). Thus, there is no random element in this situation, and there is no need for multiple simulations with the same parameters.

TABLE 10.6 The GCAPM MS Results

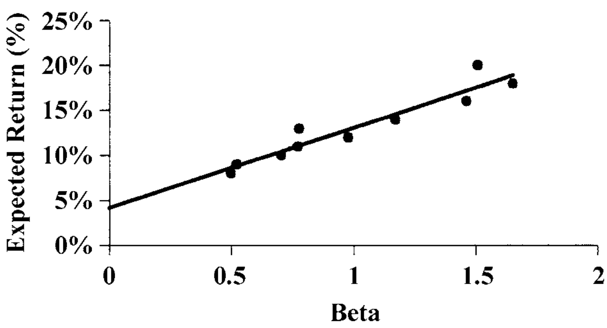
a. The stocks' parameters

Stock no.	Expected return	Variance	Beta^G
1	0.08	0.04	0.499
2	0.09	0.014	0.523
3	0.1	0.032	0.706
4	0.11	0.026	0.773
5	0.12	0.062	0.98
6	0.13	0.014	0.78
7	0.14	0.053	1.173
8	0.16	0.09	1.464
9	0.18	0.078	1.655
10	0.2	0.032	1.512

b. The regression results

$$\mu_1 = \gamma_0 + \gamma_1 \beta_i^G + e_i$$

	Intercept	Coefficient	R square
	0.042	0.089	0.905
Significance	3.8	8.724	



Next, we employ the basic data of Table 10.6 to calculate the relationship between μ_i and β_i^G, in a segmented market with neglected stocks. In this specific example, we assume that stock number 10's chance to be included in each portfolio is 50% smaller than the chance of each of the other 9 stocks. A comparison of Tables 10.6 and 10.7 reveals that due to the neglection of stock 10, the beta of stock number 10 is reduced from 1.512 to 0.981. Because μ is unchanged (20%), in the $\mu - \beta$ space the point moves to the left (see figure in Table 10.7), hence we obtain a neglected stock effect. Namely, looking at stock 10's observed μ and β^G, it would seem that this stock is undervalued according to the CAPM.

TABLE 10.7 The MS Results: Neglected Stock Effect (stock no. 10 has 50% less chance to be selected)

a. The stocks' parameters

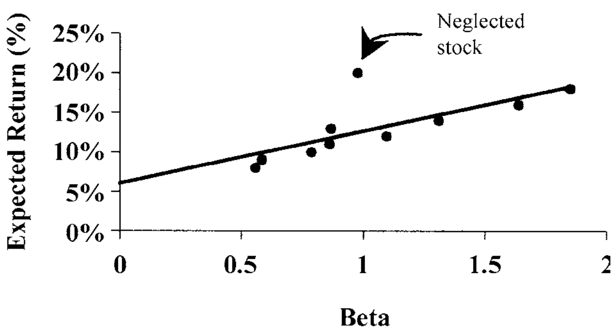
Stock no.	Expected return	Variance	Beta^G
1	0.08	0.04	0.559
2	0.09	0.014	0.585
3	0.1	0.032	0.79
4	0.11	0.026	0.864
5	0.12	0.062	1.097
6	0.13	0.014	0.869
7	0.14	0.053	1.313
8	0.16	0.09	1.639
9	0.18	0.078	1.852
10	0.2	0.032	0.981

b. The regression results

$\mu_1 = \gamma_0 + \gamma_1 \beta_i^G + e_i$

	Intercept	Coefficient	*R* square
	0.061	0.066	0.523
Significance	2.422	2.961	

Market Portfolio Expected Return is 13.07%



Because the neglected stocks are generally small stocks, we obtain the well-known small firm effect. Using the stock 10's reported beta, according to the CAPM stock 10's expected return should be

$$E(R_{10}) = 0.04 + 0.981(0.1307 - 0.04) = 12.9\%$$

while the actual return is 20%, hence there is a small firm effect of the magnitude of

$$20\% - E(R_{10}) = 20\% - 12.9\% = 7.1\%$$

To understand why beta of the neglected stock is reduced let us, for simplicity, analyze once again the zero correlation case. Recall that without the neglection, in the segmented market we have

$$\beta_i^G = \frac{x_i^* \sigma_i^2}{\sum_{j=1}^{n} x_j^{*2} \sigma_j^2}$$

Suppose now that the ith stock is neglected. Then x_i^* decreases. As the numerator decreases faster than the denominator (in percentage terms), β_i^G of a neglected stock decreases, creating the small firm or neglected stock effect.[9]

10.4 SUMMARY

The Sharpe-Lintner CAPM is a pillar of modern finance. Given μ, Σ, and r, one can derive the SML showing that μ_i is related linearly to the risk measured by β_i. All investors have homogeneous expectations and as a result all hold all available assets.

Empirical findings show a relatively poor fit to the theoretical SML, with investors holding only a relatively small number of assets in the portfolio.

In this chapter we first introduced heterogeneous expectations. We find that with *ex ante* parameters (in contrast to *ex post* empirical estimates), the CAPM is very robust even with 30% heterogeneity factor as long as the number of investors and the number of assets are large.

We next introduce a segmented market where each investor holds $n_0 < n$ assets in his portfolio. We assume homogeneous expectations and a large number of investors. When each asset has an equal chance to be included in each portfolio, we obtain an intercept that is higher than the riskless interest rate, which conforms with empirical data. When we

[9] In the independence case we have in the denominator of beta

$$\sum_{j=1}^{n} x_j^{*2} \sigma_j^2 = x_i^{*2} \sigma_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^{n} x_j^{*2} \sigma_j^2$$

when x_i^* decreases the first term decreases but all x_j^* increase. As only one term of the portfolio variance decreases, the variance of the market portfolio may even increase. In any case, the decrease in the denominator is generally smaller or equal to the decrease in the numerator in *absolute* terms. Since the denominator is greater than the numerator, it is reduced less in *percentage terms*. Hence we expect beta of neglected stocks to decrease due to the neglection (i.e., due to a decrease in x_i^*).

assume a market with neglected stocks, we obtain a strong small firm effect.

To sum up, it seems that heterogeneous beliefs are not as crucial to the CAPM results as market segmentation. Of course, one can extend this work by assuming a segmented market with heterogeneous beliefs, which is probably the most relevant case. Also, one may consider an uneven wealth distribution, with the wealthiest investors holding more assets in their portfolios than investors who are less wealthy. These issues are currently under investigation.

APPENDIX 10.1

Heterogeneous expectations induce a systematic pricing bias: low expected return stocks are overpriced relative to the homogeneous-expectations price, and high expected return stocks are underpriced. This effect explains why the heterogeneous expectations SML* is steeper than the SML and why we obtain $\gamma_0 < r$. While this effect is very substantial for a small number of stocks (see Figure 10.2), as the number of stocks grows, this effect vanishes and γ_0 converges to r. The explanation for this effect is rooted in the nonlinear dependence of the investment proportion x_{ik} on the subjective expected return μ_{ik}. Let us elaborate.

In the homogeneous expectations CAPM, the investment proportion in stock i is given by

$$x_i = \frac{\dfrac{\mu_i - r}{\sigma_i^2}}{\displaystyle\sum_{j=1}^{n} \frac{\mu_j - r}{\sigma_j^2}}$$

(Recall that we are dealing with the simplified case in which the covariance matrix is diagonal, hence Eq. (10.9) holds. A similar treatment can be done in the case where the covariances are not zero.) The denominator is a normalization term, ensuring that $\sum_{i=1}^{n} x_i = 1$. To analyze the effects of heterogeneous expectations, let us focus on a simplified scenario in which the heterogeneity of expectations is only with respect to the expected return of stock i (there is agreement regarding the expected return of all other stocks). Suppose that there are only two investors in the market, who have the following estimations regarding the expected return of stock i:

$$\text{Investor 1: } \mu_{i1} = \mu_i(1 + \varepsilon) = \mu_i + \mu_i\varepsilon$$

$$\text{Investor 2: } \mu_{i2} = \mu_i(1-\varepsilon) = \mu_i - \mu_i\varepsilon$$

Thus, the investors deviate symmetrically from the homogeneous-market expected return μ_i, and they are identical in all other respects. How will stock i be priced relative to the homogeneous CAPM price?

Investor 1's investment proportion in stock i is given by

$$x_{i1} = \frac{\dfrac{\mu_i + \mu_i\varepsilon - r}{\sigma_i^2}}{\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \dfrac{\mu_j - r}{\sigma_j^2} + \dfrac{\mu_i + \mu_i\varepsilon - r}{\sigma_i^2}} = \frac{\dfrac{\mu_i - r}{\sigma_i^2} + \dfrac{\mu_i\varepsilon}{\sigma_i^2}}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} + \dfrac{\mu_i\varepsilon}{\sigma_i^2}}$$

Investor 2's investment proportion in stock i is given by

$$x_{i2} = \frac{\dfrac{\mu_i - \mu_i\varepsilon - r}{\sigma_i^2}}{\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} \dfrac{\mu_j - r}{\sigma_j^2} + \dfrac{\mu_i - \mu_i\varepsilon - r}{\sigma_i^2}} = \frac{\dfrac{\mu_i - r}{\sigma_i^2} - \dfrac{\mu_i\varepsilon}{\sigma_i^2}}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \dfrac{\mu_i\varepsilon}{\sigma_i^2}}$$

The fraction of the total of all investors' wealth invested in stock i, x_i^*, is given by

$$x_i^* = \frac{1}{2}(x_{i1} + x_{i2})$$

(Recall that the two investors have equal wealth.) Therefore, x_i^* can be explicitly written as

$$x_i^* = \frac{1}{2}(x_{i1} + x_{i2}) = \frac{1}{2}\left(\frac{\dfrac{\mu_i - r}{\sigma_i^2} + \dfrac{\mu_i\varepsilon}{\sigma_i^2}}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} + \dfrac{\mu_i\varepsilon}{\sigma_i^2}} + \frac{\dfrac{\mu_i - r}{\sigma_i^2} - \dfrac{\mu_i\varepsilon}{\sigma_i^2}}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \dfrac{\mu_i\varepsilon}{\sigma_i^2}} \right)$$

$$= \frac{\dfrac{\mu_i - r}{\sigma_i^2} - \left(\dfrac{\left(\dfrac{\mu_i\varepsilon}{\sigma_i^2}\right)^2}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \left(\dfrac{\left(\dfrac{\mu_i\varepsilon}{\sigma_i^2}\right)^2}{\sum\limits_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)}$$

where the last equality is obtained by finding the common denominator and rearranging. Notice that $x_i^* < x_i$. This is because both the numerator and the denominator of x_i^* are reduced by the same amount with respect to x_i. Since the numerator is smaller than the denominator, $\left(\frac{\mu_i - r}{\sigma_i^2} < \sum_{j=1}^{n} \frac{\mu_j - r}{\sigma_j^2} \right)$, the reduction in the numerator is greater than the reduction in the denominator *in percentage terms*. Thus, x_i^* is smaller than x_i.

As a consequence, stock i will be priced lower in the market with (symmetric) homogeneous expectations than it would be priced in a homogeneous-expectations market. What about the pricing of the other stocks, regarding which there is agreement ($\varepsilon = 0$)? For these stocks, x_j^* must increase relative to x_j, because $\sum_{j=1}^{n} x_j^* = \sum_{j=1}^{n} x_j = 1$, and $x_i^* < x_i$ (this can also be shown from a direct calculation of x_j^*). Thus, all stocks other than stock i will be overpriced relative to the homogeneous CAPM prices.

So far, we have analyzed pricing in the case where there are heterogeneous expectations regarding only one stock. If there is heterogeneity with respect to more than one stock, there are counteracting effects on pricing. For example, if there are heterogeneous expectations regarding the returns of stocks i and j, the heterogeneity regarding stock i tends to reduce x_i^*, but the heterogeneity with respect to stock j tends to increase x_i^* (since it reduces x_j^*). Thus, the situation is quite complex. Notice, however, that while the denominator is affected in the same way for all stocks (it is $\sum_{j=1}^{n} x_j^*$ for all stocks), the effect on stock i's numerator is determined by the term $\frac{\mu_i \varepsilon}{\sigma_i^2}$. Thus, the larger $\frac{\mu_i}{\sigma_i^2}$, the larger we expect the underpricing to be (for stocks with low $\frac{\mu_i}{\sigma_i^2}$, the cross effects can be larger than the direct effects, leading to overpricing). This result is reflected in Table 10.3b. In the situation described by the table, stock 4, which has the highest $\frac{\mu_i}{\sigma_i^2}$ is underpriced, while the other stocks are overpriced. Similar results are also obtained in Table 10.4b, which describes a market with 20 stocks.

Stocks which have a high expected return in the homogeneous expectations CAPM (high μ) will *tend* to have high $\frac{\mu}{\sigma^2}$ and therefore will tend to be underpriced in a heterogeneous expectations market. As a consequence, their actual expected return, μ^*, will tend to be higher than their return in a homogeneous market, μ. In contrast, stocks with low μ will tend to be overvalued and will therefore tend to have $\mu^* < \mu$. This explains the fact that the slope of the SML* is larger than the slope of the SML (see Figures 10.2 and 10.4) and the fact that, in general, $\gamma_0 < r$.

While this systematic pricing bias is very significant in a market with $n = 5$ stocks, it is much less significant in a market with $n = 20$ stocks. In fact, as $n \to \infty$ the effect disappears altogether and the SML* coincides with the SML. To see this, suppose again for simplicity that the heterogeneous expectations are only regarding stock i and that they are symmetric ($\mu_{i1} = \mu_i(1+\varepsilon)$; $\mu_{i2} = \mu_i(1-\varepsilon)$). We have seen that in this case x_i^* is given by

$$x_i^* = \frac{\dfrac{\mu_i - r}{\sigma_i^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}}\right)}{\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}}\right)}$$

Now, assume that the number of assets in the market grows by a factor of q. To make a reasonable comparison, let us assume that

$$n' = qn, \text{ that } \sum_{j=1}^{n'} \frac{\mu_j - r}{\sigma_j^2} = q \sum_{j=1}^{n} \frac{\mu_j - r}{\sigma_j^2}, \text{ and that } w_0' = q w_0$$

where the superscript $'$ denotes the parameters in the market with the larger number of assets.[10] Following the previous derivation of x_i^* we find that $x_i'^*$ is given by

$$x_i'^* = \frac{\dfrac{\mu_i - r}{\sigma_i^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}}\right)}{q\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}}\right)}$$

[10] These assumptions imply that on average the stocks "added" have the same typical $\dfrac{\mu - r}{\sigma^2}$ ratio as the original n stocks. Thus, when increasing the number of assets in the market we do not introduce a systematic bias.

The equilibrium price of stock i is given by

$$P_i'^* = \frac{w_0' x_i'^*}{N_i} = \frac{q w_0}{N_i} \left[\frac{\dfrac{\mu_i - r}{\sigma_i^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q \sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)}{q \sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q \sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)} \right]$$

$$= \frac{w_0}{N_i} \left[\frac{\dfrac{\mu_i - r}{\sigma_i^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q \sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)}{\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2} - \left(\dfrac{\left(\dfrac{\mu_i \varepsilon}{\sigma_i^2}\right)^2}{q^2 \sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} \right)} \right]$$

As $q \to \infty$,

$$P_i'^* \to \frac{w_0}{N_i} \frac{\dfrac{\mu_i - r}{\sigma_i^2}}{\sum_{j=1}^{n} \dfrac{\mu_j - r}{\sigma_j^2}} = \frac{w_0 x_i}{N_i} = P_i$$

Thus, when the number of securities in the market grows, the heterogeneous-expectation prices converge to the homogeneous expectation price, and the classical CAPM is recovered.

11

APPLICATION OF MICROSCOPIC SIMULATION TO OPTION PRICING: UNCERTAINTY AND DISAGREEMENT ABOUT THE VOLATILITY

11.1. INTRODUCTION

The pricing of options is one of the most important and most interesting issues in finance. The cornerstone of modern option pricing theory is the Black and Scholes option pricing model. In a breakthrough article, Black and Scholes (1973) have shown that options can be dynamically replicated by a portfolio of the underlying asset and the riskless asset, and that the value of options can therefore be derived by a no-arbitrage argument. Black and Scholes have derived the following formula for the price of a European call option:

$$C_{B\&S} = S_0 N(d_1) - Ee^{-rT} N\left(d_1 - \sigma\sqrt{T}\right) \tag{11.1}$$

where S_0 is the price of the underlying asset, E is the strike price (or exercise price), σ is the standard deviation of the underlying asset, T is the time to maturity, r is the risk-free interest rate, $N(\cdot)$ is the cumulative normal distribution, and

$$d_1 = \frac{\ln(S_0/E) + (r + (\sigma^2/2))T}{\sigma\sqrt{T}}$$

Notice that all of the parameters that go into the Black and Scholes formula are known, except for the underlying asset's volatility, σ, which must be estimated. The Black and Scholes model is one of the most important models in finance. It is both a practical tool used daily by thousands of investors and a theoretical cornerstone on which further theoretical studies have been developed.

Empirical investigations have revealed some systematic deviations of option prices from the Black and Scholes model price. Most of these findings are stated in terms of the implied volatilities of different options on the same underlying asset.[1] If options are priced according to the Black and Scholes formula, options on the same underlying asset should have the same implied volatility. However, empirical investigations have shown that when the implied volatility is drawn as a function of the strike price of various options (on the same underlying asset), a "smile" rather than a horizontal line is observed—that is, deep in-the-money and deep out-of-the-money options have higher implied volatilities than near-the-money options. This finding has been termed the "volatility smile" effect.[2]

The relationship between the implied volatility and the option's maturity is known as the "implied volatility term-structure." In an empirical study, Rubinstein (1985) finds that the implied volatility decreases with the maturity of out-of-the-money options. Other researchers have found an opposite relationship (see, for example, Derman and Kani, 1994). For at-the-money options, Rubinstein does not find a significant dependence of the implied volatility on the option's maturity.

To explain these deviations of option prices from the Black and Scholes pricing formula, several extensions of the Black and Scholes model have been introduced. In the Black and Scholes model, the price of the underlying asset is assumed to follow a diffusion process. Many of the studies following Black and Scholes have been focused on extending option pricing to other underlying asset price processes such as stochastic volatility processes,[3] jump processes,[4] jump diffusion processes,[5] displaced diffusion processes,[6] and compound option models.[7] However, empirical research suggests that these alternatives to Black and Scholes do not do a much better job of explaining the observed option pricing. Hull concludes

[1] The implied volatility of an option is the volatility that is consistent with the observed option price, the observed parameters S_0, E, T, r, and the Black and Scholes formula, Eq. (11.1).

[2] For a fuller description see, for example, Derman and Kani (1994) and Hull (1997).

[3] Hull and White (1987, 1988); Heston (1993).

[4] Cox and Ross (1975); Cox, Ross, and Robinstein (1979).

[5] Merton (1976).

[6] Rubinstein (1983).

[7] Geske (1979).

a review of alternative option pricing models:

> At present, there do not seem to be any really compelling arguments for using any of the models introduced earlier in this chapter in preference to Black and Scholes for stock options. (Hull, 1997, p. 509)

A drawback of the Black and Scholes model, and most of its extensions, is that it assumes that *σ is known*, and that *all investors agree about its value*. Otherwise, the option pricing formulas break down. However, if all investors agree about the value of σ, they should also all agree about the option's value, and we would not expect any trade in the option. In fact, according to the Black and Scholes dynamic tracking logic, if σ is known the option can be perfectly replicated, and it is therefore redundant altogether. This is in sharp contrast to real markets, in which σ is not known, investors have different estimations of σ, and active option markets are observed.

While it is clear that uncertainty about the value of σ and disagreement about this value are crucial elements in actual option markets, these elements are difficult to model analytically. Specifically, if investors disagree about the value of σ, and each investor believes that he or she *knows* the true value *with certainty*, investors would take infinite positions in the option, and one cannot say anything about the equilibrium option price in this case. Thus, although uncertainty about σ is difficult to analyze analytically, it seems to be crucial in order to obtain realistic equilibrium option pricing.

In this chapter we employ microscopic simulation (MS) in order to investigate option pricing when investors have some uncertainty about the value of σ and may disagree about σ's distribution. Uncertainty and disagreement about the value of σ translate to uncertainty and disagreement about the value of the option. There are various alternatives of modeling option pricing within this framework. In the present analysis we chose a minimalistic setup in order to investigate the fundamental effects of uncertainty and disagreement about σ on pricing in a simple setting. Namely, we assume that investors translate their (subjective) probability distribution of σ to a probability distribution of the option's value, according to the Black and Scholes formula, Eq. (11.1). Thus, the option is viewed as a risky asset with an uncertain value, and investors' demands for the option are a function of their estimation regarding the distribution of σ and of the option price. The equilibrium option price is then determined by market clearance.

We find that uncertainty about σ may explain the empirically observed volatility smile and the various findings regarding the volatility term structure, while heterogeneous beliefs about σ explain the existence of active option trading. We believe that uncertainty and disagreement regarding σ will have the same general effects on option pricing in more elaborate and detailed models.

11.2. THE MODEL

We model the pricing of a European call option in a market with uncertainty and disagreement regarding the value of σ, the standard deviation of the underlying asset. We consider a two-period model. Investors do not know σ with certainty. Rather, they estimate σ, either from *ex post* data or from other sources of information. Investor k believes that σ is distributed according to some probability distribution function, $f_k(\sigma)$.[8] The investor believes that the distribution of the option's value at the future time T_1 will be determined according to the Black and Scholes formula and his subjective belief regarding the distribution of σ, $f_k(\sigma)$. Namely, according to the Black and Scholes pricing formula (Eq. (11.1)), each value of σ determines an option value $C_{B\&S}(\sigma)$. Thus, at time T_0 investor k believes that the option price at the future time T_1 is a random variable given by $\tilde{C}_1 = C_{B\&S}(\tilde{\sigma})$, where $\tilde{\sigma}$ is distributed according to $f_k(\sigma)$.[9] Investors optimize their investment in the option such as to maximize their end-of-period expected utility. If the option is traded at time T_0 at a hypothetical price C_h, and investor k buys N_k options, then

[8] If the underlying asset's *ex post* returns $R_{t-1}, R_{t-2}, \ldots, R_{t-n}$, are employed in order to estimate σ, and the rates of return are independent over time, then $\frac{\sum_{i=1}^{n}(R_{t-i}-\bar{R})^2}{\sigma^2}$ is distributed according to the χ^2 distribution with $n-1$ degrees of freedom, and σ^2 is estimated to be distributed according to $f(\sigma^2) = \frac{\sum_{i=1}^{n}(R_{t-i}-\bar{R})^2}{\sigma^4} g\left(\frac{\sum_{i=1}^{n}(R_{t-i}-\bar{R})^2}{\sigma^2}\right)$, where g is the χ^2 density function with $n-1$ degrees of freedom (see, for example, Mood and Graybill, 1963). Investors estimating σ may differ in the number of observations they employ in their estimates, the time frame over which returns are calculated, and so on. In what follows we assume more basic distributional forms of $f(\sigma)$ for the sake of simplicity; however, our results hold for general distributions $f(\sigma)$.

[9] The stock price at time T_1, $\tilde{S}_1$, is also a random variable, which determines $\tilde{C}_1$. Given the Black and Scholes assumption of a diffusion process, $\tilde{S}_1$ is distributed according to a lognormal distribution. S_0, μ, and σ determine the parameters of this distribution. Thus, μ and $f_k(\sigma)$ determine the distribution of $\tilde{C}_1$. In the present analysis we wish to isolate the direct effect of the uncertainty regarding σ on pricing, and we therefore make the simplifying assumption $S_1 = S_0$. The indirect effect of σ on pricing through the distribution of $\tilde{S}_1$ generally only enhances the effects discussed here (see M. Levy, 1999a).

his wealth at time T_1 is a stochastic variable given by[10]

$$\tilde{W}_{k,1} = W_{k,0} - N_k C_h + N_k \tilde{C}_{k,1} \tag{11.2}$$

where $W_{k,0}$ is investor k's wealth at T_0, $\tilde{C}_{k,1}$ is given by investor k's subjective estimation $f_k(\sigma)$, and the interest rate is taken as 0 for simplicity.[11] The investor calculates his expected utility as

$$EU\left(\tilde{W}_{k,1}\right) = \int U[W_{k,0} - N_k C_h + N_k C_{B\&S}(\sigma)] f_k(\sigma)\, d\sigma \tag{11.3}$$

For any hypothetical option price, C_h, investor k will choose to hold $N_k^*(C_h)$ options, where $N_k^*(C_h)$ is the number of options that maximizes the expected utility given by Eq. (11.3). ($N_k^*(C_h)$ can be either positive, which means holding a long position in the option, or negative, which means holding a short position in the option.)

Calculating $N_k^*(C_h)$ for any hypothetical price, C_h, yields investor k's demand function for the option. Different investors will generally have different demand functions, as they may disagree about $f(\sigma)$. The equilibrium option price is determined by the market clearance condition. Namely, the option is in net zero supply. Thus, the equilibrium option price at time T_0, C_0, is the price for which the total demand is zero:

$$\sum_k N_k^*(C_0) = 0 \tag{11.4}$$

where the summation is over all investors in the market.[12]

Notice that in the case where investors know σ with certainty and agree about its value (as in the Black and Scholes model assumptions), the equilibrium option price converges to the Black and Scholes price. The reason for this is that investors who are certain that the volatility is σ are confident that at T_1 the option price will be given by the Black and Scholes formula with this σ, $C_{B\&S}(\sigma)$. If we take $T_1 - T_0 \to 0$, then investors will price the option at T_0 exactly according to the Black and Scholes price,

[10] Note that in the Black and Scholes model, holding the stock and the corresponding option in the relevant proportions creates a perfect hedge. This is not the case here because the value of σ is unknown. Therefore, the option is treated as any other risky asset. For simplicity, we assume here that the investor's portfolio is composed only of the option and the riskless asset. The extension of the present analysis to a portfolio context is possible (see M. Levy, 1999a).

[11] It is straightforward to extend the model to nonzero interest rates. In this case Eq. (11.2) will read $\tilde{W}_{k,1} = (W_{k,0} - N_k C_h)(1 + r) + N_k \tilde{C}_{k,1}$, where r is the interest rate for the period $T_1 - T_0$.

[12] In general, the aggregate demand is a decreasing function of the option price, C_h. This ensures the existence of a unique equilibrium price C_0 for which Eq. (11.4) holds.

$C_{B\&S}(\sigma)$ (if the option price differs from $C_{B\&S}(\sigma)$, investors perceive an arbitrage opportunity).

Employing MS, we examine the preceding model in order to investigate the effects of uncertainty and disagreement about σ on the price of the option relative to the benchmark Black and Scholes price. We would like to stress that MS is essential to this investigation. Analyzing investors' demand functions and the equilibrium pricing is generally impossible to carry out in an analytical framework. In what follows we separately analyze the effects of *uncertainty* about σ and the effects of *heterogeneous estimation* of $f(\sigma)$.

11.3. RESULTS

11.3.1 The Effects of Uncertainty Regarding σ — Homogeneous Estimation of $f(\sigma)$

The fact that investors do not have perfect knowledge about the value of σ can explain both the empirically documented volatility smile and various findings regarding the volatility term structure. To understand these effects it is helpful to recall the dependence of the Black and Scholes price on σ. Figure 11.1 depicts the value of $C_{B\&S}$ as a function of σ for a typical call option ($S = 100$, $E = 120$, $T = 1/12$ (one month)).

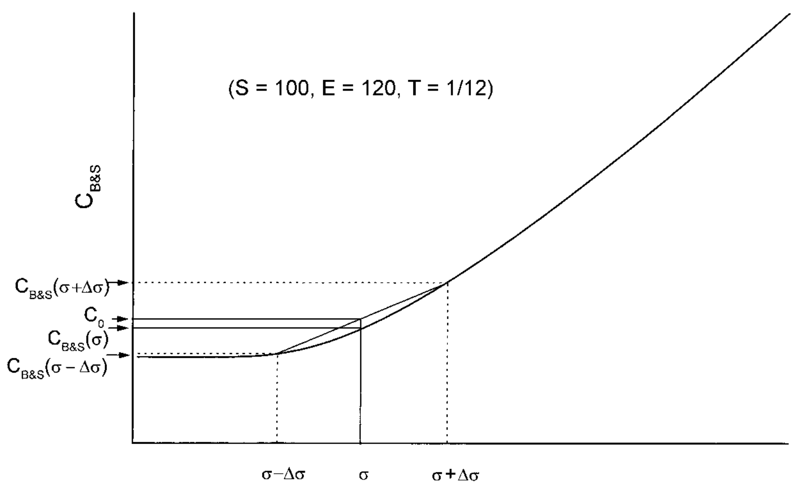


FIGURE 11.1 (a) The Black and Scholes call price as a function of σ-convex region: $C_0 > C_{B\&S}(\sigma)$. (b) The Black and Scholes call price as a function of σ-linear region: $C_0 = C_{B\&S}(\sigma)$.

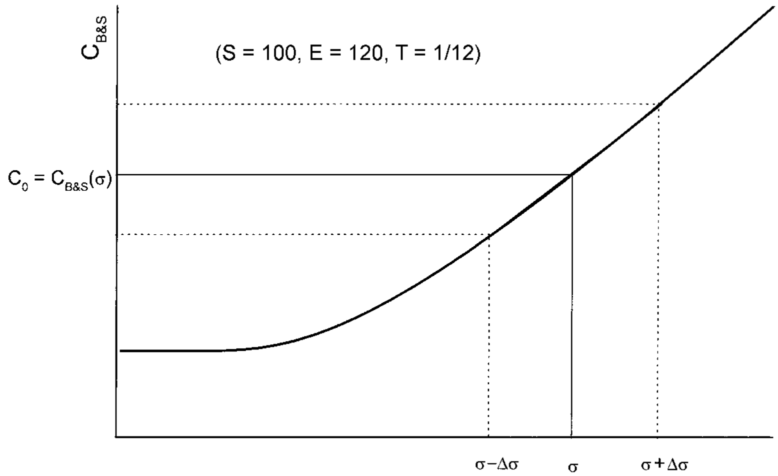


FIGURE 11.1 Continued.

If investors know with certainty that the standard deviation of the underlying asset is σ, then they know that at T_1 the option price will be $C_{B\&S}(\sigma)$, and they therefore price the option at T_0 exactly at this value.[13] However, if investors are not sure about the value of σ, the situation changes. Assume, for simplicity, that investors estimate that σ is either $\sigma - \Delta\sigma$ or $\sigma + \Delta\sigma$ with equal probability (see Figure 11.1a). In this case $\tilde{C}_1$ is given by

$$\tilde{C}_1 = \begin{cases} & \text{probability} \\ C_{B\&S}(\sigma - \Delta\sigma) & 1/2 \\ C_{B\&S}(\sigma + \Delta\sigma) & 1/2 \end{cases}$$

It is reasonable to suspect that the uncertainty about σ, and therefore about C_1, may have two effects on the option price relative to the case in which σ is known:

1. As the option value at T_1 becomes uncertain, risk averters may price the option lower.
2. The convexity of $C_{B\&S}$ as a function of σ (see Figure 11.1a) may induce higher pricing.

We will first show that in the present setting of homogeneous estimation the degree of risk aversion does not affect the option pricing (in

[13] Recall that for simplicity and in order to isolate the effects of uncertainty from time-value-of-money effects and the effect of the change in the stock price, we are assuming $S_1 = S_0$ and $r = 0$.

contrast to the idea suggested in effect 1). We also show that the equilibrium price C_0 is equal to the expected value of the option, $\bar{C}_1$. We will then analyze effect 2 with the added simplifying assumption of risk neutrality, which does not change the results.

In general, the degree of risk aversion affects asset pricing. However, risk aversion does not affect option pricing in the homogeneous estimation setting. This is counterintuitive, because we would expect risk averters to price the option lower than its expected value. The reason that risk aversion does not affect pricing in our case is related to the fact that, by its nature, the option is in zero net supply. Let us elaborate.

In general, the degree of risk aversion *does* affect the investor's demand function for the option, $N_k^*(C_h)$. However, because of the homogeneous expectation regarding $f(\sigma)$, the demand for the option is zero for each of the investors (independent of their risk aversion) exactly at the same option price C_0. In other words, $N_k^*(C_0) = 0$ for all investors k (see Figure 11.2). This is not a trivial assertion and should be elaborated. We base our analysis on Arrow's assertion that if a risky asset has a positive expected return (and $r = 0$), any risk-averse investor will demand some positive quantity of this asset (see the discussion in Arrow (1971, pp. 98–101). In the homogeneous estimation case all investors have the same estimation of $f(\sigma)$ and of $\tilde{C}_1$. For $C_0 < \bar{C}_1$ (where $\bar{C}_1$ is the expected value of $\tilde{C}_1$), all investors demand some positive quantity of the option.

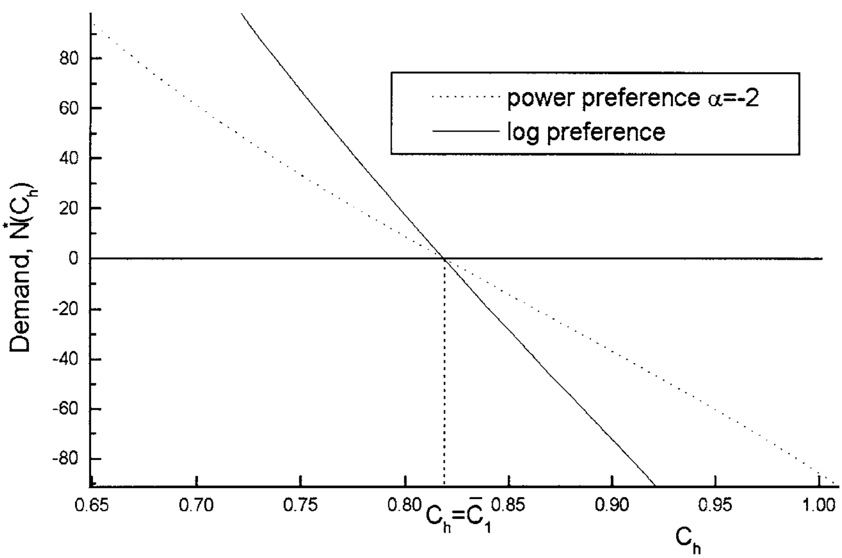


FIGURE 11.2 Demand for options—The effects of risk aversion.

For $C_0 > \overline{C}_1$, all investors demand a negative quantity of the option. This is true for any risk-averse investors.

Figure 11.2 illustrates the demand functions of two different risk averters as a function of C_h. As can be seen, the demand functions do depend on the degree of risk aversion. However, the two curves intersect at $C_h = \overline{C}_1$, at which point both investors' demands are zero. Thus, all investors' demands are zero at exactly $C_h = \overline{C}_1$, regardless of their degree of risk aversion.

As the option is in net zero supply and investors have the same estimation of $f(\sigma)$, the only market clearing price is $C_0 = \overline{C}_1$, no matter what investors preferences are. Thus, risk aversion does not affect option pricing in this setting, and the only factor that determines the effect of uncertainty on option pricing is the convexity of $C_{B\&S}(\sigma)$, (effect 2). We can therefore ignore risk aversion and assume risk neutrality without loss of generality.

If risk neutral investors believe that the standard deviation is either $\sigma - \Delta\sigma$ or $\sigma + \Delta\sigma$ with equal probability, they will price the option higher than an investor who believes with certainty that the standard deviation is σ and who prices the option at $C_{B\&S}(\sigma)$. The risk-neutral investor who is uncertain about the standard deviation prices the option at its expected value $\overline{C}_1$ (recall that $r = 0$, so there are no time-value-of-money considerations). Thus, in this market with uncertainty and homogeneous beliefs we have

$$C_0 = \overline{C}_1 = \frac{1}{2}C_{B\&S}(\sigma - \Delta\sigma) + \frac{1}{2}C_{B\&S}(\sigma + \Delta\sigma) > C_{B\&S}(\sigma) \qquad (11.5)$$

(see Figure 11.1a). The more convex $C_{B\&S}(\sigma)$ is at σ, the more significant the overpricing (relative to the Black and Scholes pricing) induced by the uncertainty. In contrast, if σ is located in a region where $C_{B\&S}(\sigma)$ is almost linear in σ, the uncertainty does not induce significant overpricing (see Figure 11.1b).

To analyze the effects of uncertainty on the pricing of options with different exercise prices, it is instructive to draw the dependence of $C_{B\&S}(\sigma)$ on σ for different exercise prices. Figure 11.3 shows $C_{B\&S}(\sigma)$ as a function of σ for typical options on the same underlying asset with ($S = 100$, $T = 1/12$) and with various exercise prices.[14] How will these options be priced? Consider a typical situation in which the standard deviation is estimated to be in the range $(\sigma - \Delta\sigma, \sigma + \Delta\sigma)$, as shown in Figure 11.3 (we consider the typical range $20\% < \sigma < 60\%$). The in-the-money options (e.g., $E = 80$, $E = 90$) will be overpriced relative to the

[14] The option value at $\sigma = 0$ is simply given by $\max\{S - E, 0\}$.

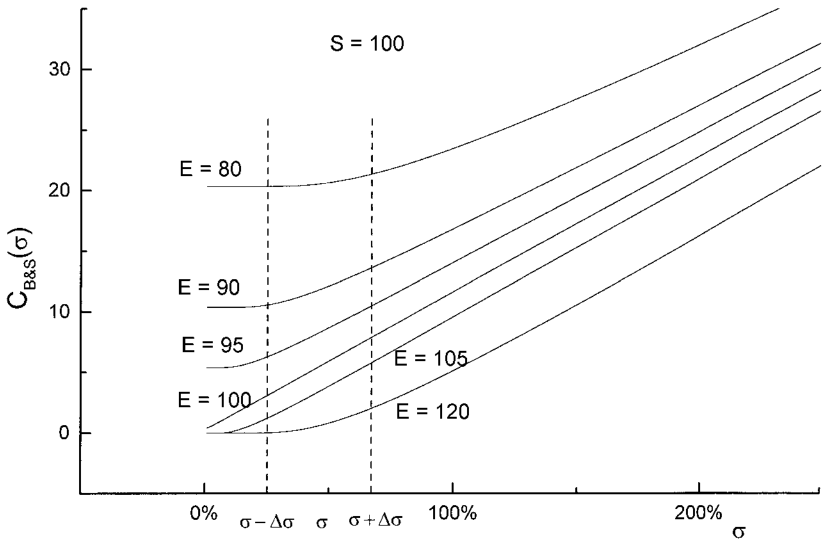


FIGURE 11.3 $C_{B\&S}$ as a function of σ, various exercise prices.

Black and Scholes price, due to the convexity of these options in this range of σ (see Figure 11.3). The same is true of out-of-the-money options (e.g., $E = 120$). However, the near-the-money options (e.g., $E = 95$, $E = 100$, $E = 105$) are almost linear in σ in this region and therefore the uncertainty will not affect their price. Translating these results into the language of implied volatilities, this implies the famous empirically documented volatility smile: in- and out-of-the-money options are overpriced relative to near-the-money options, and therefore their implied volatilities are higher. Figure 11.4 shows the implied volatilities for options on an underlying asset with ($S = 100$, $T = 1/12$) in a market of investors who believe that σ is uniformly distributed in the range [0.3–0.5].[15] Similar results are obtained with other assumptions regarding the distribution of σ.

Uncertainty about the underlying asset's volatility may also explain the various implied volatility term-structures observed. Figure 11.5 shows the Black and Scholes option price as a function of σ for out-of-the-money options with various maturities ($S = 100$, $E = 120$). If σ is estimated to be in a relatively low range, the convexity of $C_{B\&S}(\sigma)$ increases with the option's maturity (see the range between the dashed lines in Figure 11.5). This means that options with longer maturities will be overpriced relative

[15] The utility function is taken as $U(W) = \frac{W^{1-\alpha}}{1-\alpha}$ with $\alpha = 1.5$, and the initial wealth is taken as \$100. As explained above, the choice of preference and initial wealth does not affect the option pricing in the homogeneous belief case.

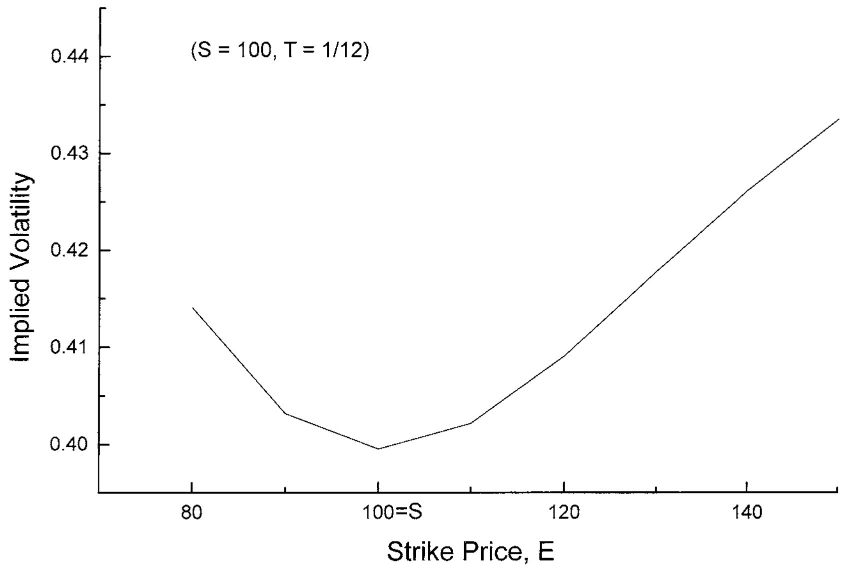


FIGURE 11.4 Implied volatility smile in a homogeneous market. Investors believe that σ is uniformly distributed in the range 0.3–0.5.

to options with shorter maturities, which leads to an upward sloping implied volatility term-structure, as found by Derman and Kani (1994).

In contrast, if σ is estimated to be in a higher range, the Black and Scholes price of the longer maturity options is almost linear in σ, while it is convex for shorter maturity options (see the range between the dotted lines in Figure 11.5). This implies a downward sloping volatility term structure as found by Rubinstein (1985). Figure 11.6 shows the implied volatility term structure obtained for options with the previous parameters in a market of investors who estimate the underlying asset's standard deviation to be uniformly distributed in the range (0.3–0.5). A similar argument holds for in-the-money options. However, for at-the-money options, the Black and Scholes option price is nearly linear in σ for all maturities and ranges of estimated σ (see Figure 11.7), implying a flat term structure, as observed by Rubinstein for at-the-money options.

While the uncertainty regarding σ may explain the volatility smile and various volatility term structures, if investors hold the same opinion regarding the distribution of σ, the uncertainty does not explain the existence of option trading volume. This is because in the homogeneous beliefs model, all investors agree on the equilibrium price C_0 and no trade takes place. As we shall see, when investors are uncertain about the value of σ and have different estimations of σ's distribution, the volatility smile

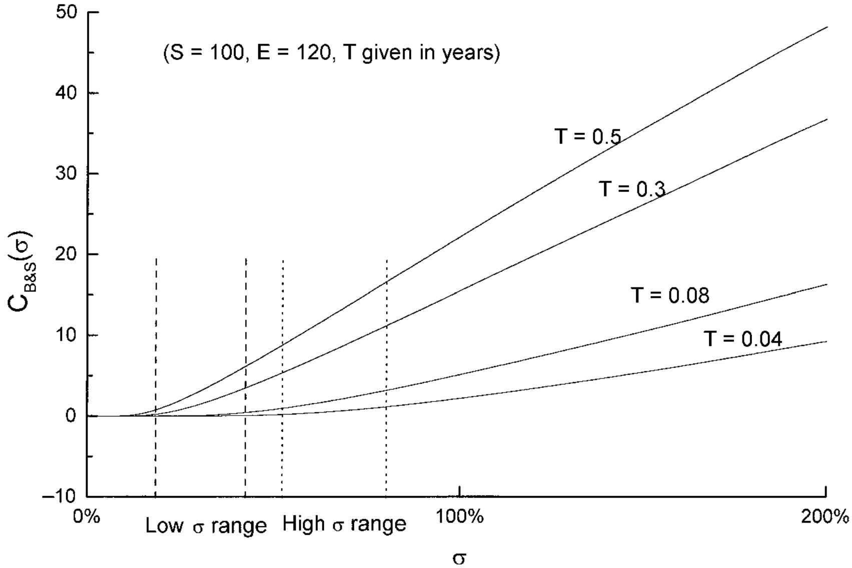


FIGURE 11.5 The Black and Scholes call price as a function of σ for out-of the-money options with various maturities.

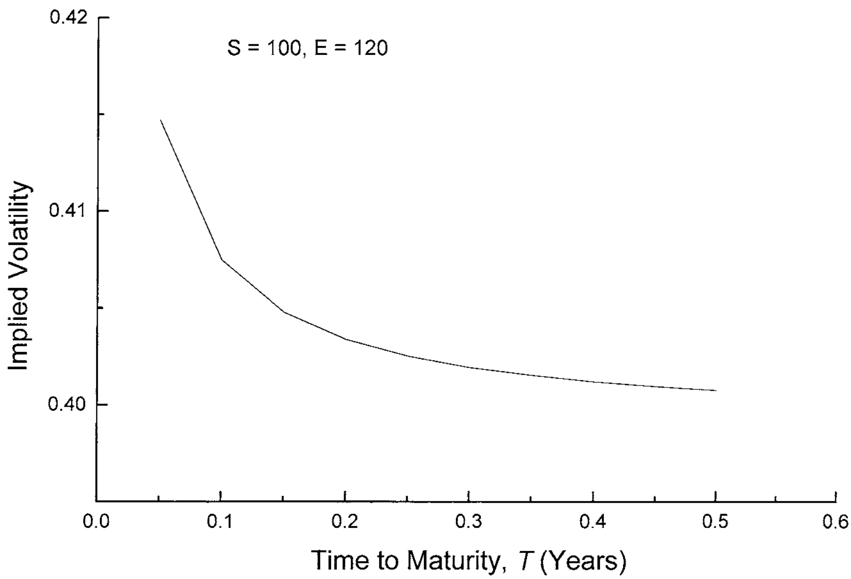


FIGURE 11.6 Implied volatility term structure for an out-of-the-money option. Investors believe that σ is uniformly distributed in the range 0.3–0.5.

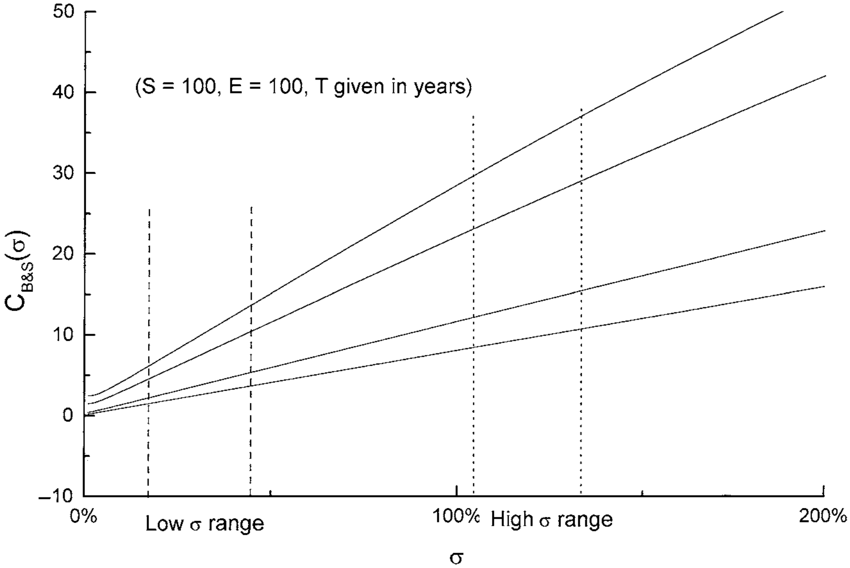


FIGURE 11.7 The Black and Scholes call price as a function of σ for at-the-money options with various maturities.

and term structure are observed as in the homogeneous-estimation case; *in addition*, trading volume is generated.

11.3.2 The Effects of Heterogeneous Estimation of $f(\sigma)$

When investors are *uncertain* about the value of σ and they are *heterogeneous* in their estimation of the distribution of σ, the same volatility smile and volatility term structure effects are induced by the uncertainty, and trading volume is generated by the heterogeneity. The uncertainty has the same general effect on pricing as in the homogeneous estimation case. For example, suppose that there are two investors with different estimations regarding the distribution of σ. Investor A estimates that σ is either $\sigma_A - \Delta\sigma$ or $\sigma_A + \Delta\sigma$ with equal probability, while investor B estimates that σ is either $\sigma_B - \Delta\sigma$ or $\sigma_B + \Delta\sigma$ with equal probability. Investor A demands positive quantities of the option if $C_0 < \overline{C}_A$ and negative quantities if $C_0 > \overline{C}_A$, where $\overline{C}_A = \frac{1}{2}[C_{B\&S}(\sigma_A - \Delta\sigma) + C_{B\&S}(\sigma_A + \Delta\sigma)]$. Investor B demands positive quantities of the option if $C_0 < \overline{C}_B$ and negative quantities if $C_0 > \overline{C}_B$, where $\overline{C}_B$ is correspondingly defined as $\overline{C}_B = \frac{1}{2}[C_{B\&S}(\sigma_B - \Delta\sigma) + C_{B\&S}(\sigma_B + \Delta\sigma)]$ (see Figure 11.8).

Since the option is in net zero supply, the equilibrium option price will be somewhere in the range $\overline{C}_A < C_0 < \overline{C}_B$ (for $C_0 < \overline{C}_A$ both investors

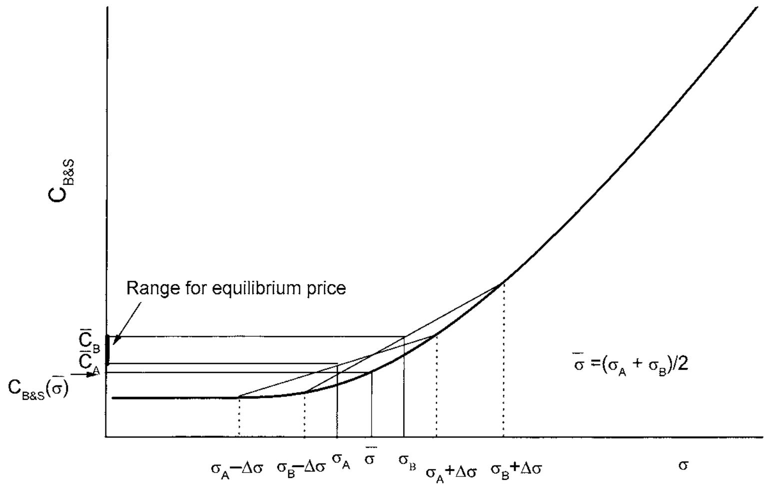


FIGURE 11.8 Pricing in a market with heterogeneous estimation of $f(\sigma)$.

will wish to buy the option, while for $C_0 > \overline{C}_B$ both investors will wish to short the option). When there is disagreement about the value of σ, one may suspect that the market price of the option will be the Black and Scholes price corresponding to the "average" belief regarding the value of σ. However, this is generally not the case. As Figure 11.8 shows, the equilibrium price will typically be higher than the Black and Scholes price corresponding to the average belief regarding σ, $C_{B\&S}\left(\frac{\sigma_A + \sigma_B}{2}\right)$.

The convexity of $C_{B\&S}(\sigma)$ with respect to σ combined with the uncertainty regarding the value of σ, have similar effects on pricing as in the homogeneous estimation case. Figure 11.9 shows the implied volatility as a function of the strike price for an option with $S = 100$, $T = 1/12$ in a market with two investor types. Type A investors believe that σ is uniformly distributed in the range (0.3−0.5), while type B investors believe that σ is uniformly distributed in the range (0.35−0.55). The investors optimize their demands for the option according to Eq (11.3), where the utility functions are taken as the log function. The equilibrium pricing is then determined by the market clearance condition according to Eq. (11.4).

In contrast to the homogeneous estimation case, when investors have different estimations of σ some investors hold long positions in the option, while others short the option, and trading volume is generated. Figure 11.10 shows the number of options traded in a market with two populations who have different estimations of σ. As before, the stock

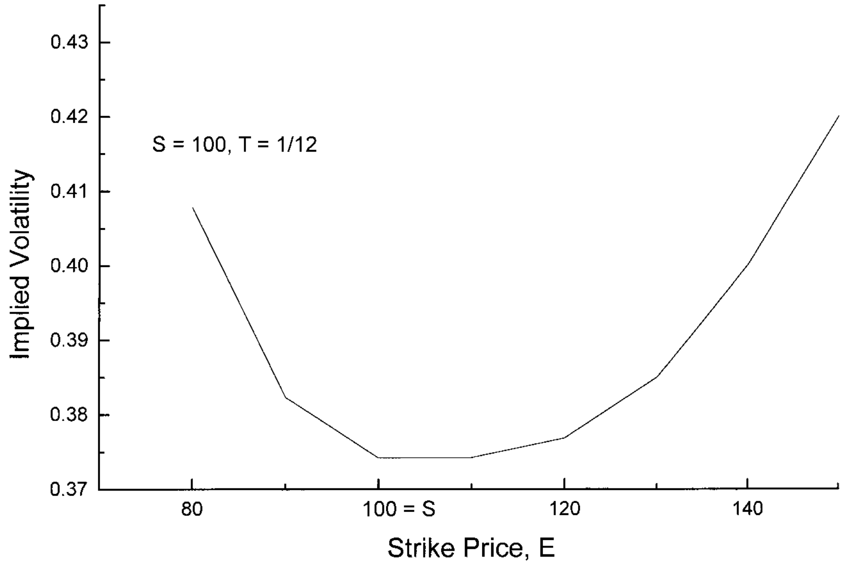


FIGURE 11.9 Implied volatility smile in a heterogeneous market. Type A investors believe that σ is uniformly distributed in the range 0.3–0.5. Type B investors believe that σ is uniformly distributed in the range 0.35–0.55.

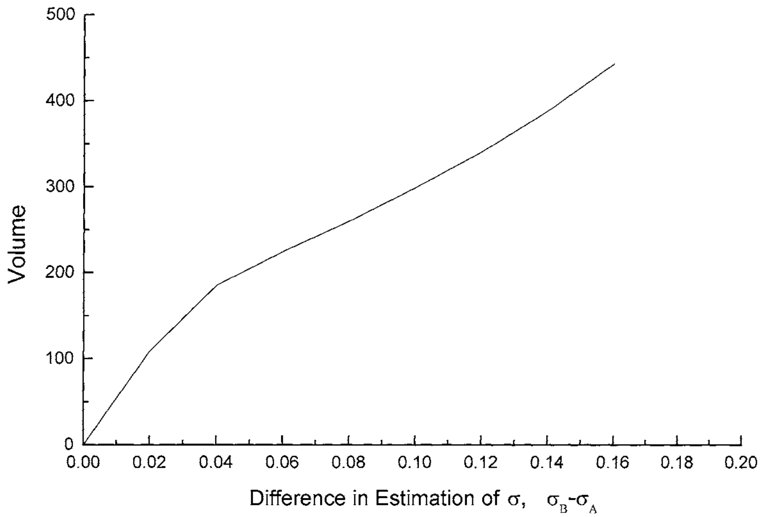


FIGURE 11.10 Trading volume in a heterogeneous market.

price is 100, the exercise price is 120, and the time to maturity is 1 month. The volume is calculated as the sum of absolute positions—that is, $\sum_k |N_k|$. Population A believes that σ is uniformly distributed in the range $[\sigma_A - 0.1, \sigma_A + 0.1]$, while population B believes that σ is uniformly distributed in the range $[\sigma_B - 0.1, \sigma_B + 0.1]$. We take $\sigma_A = 0.4$, hence investor A believes that σ is uniformly distributed in the range [0.3, 0.5], as before. We calculate the volume for different values of σ_B. Figure 11.10 shows the volume as a function of the disagreement regarding σ, $\sigma_B - \sigma_A$ $(= \sigma_B - 0.4)$. As expected, the more disagreement, the more extreme the positions taken, and the higher the trading volume. Similar results are generally obtained when one assumes heterogeneous preferences in addition to the heterogeneous estimation of $f(\sigma)$.[16]

11.4. SUMMARY

The Black and Scholes model is the cornerstone of modern option pricing theory. This model, as well as various other option pricing models based on the Black and Scholes model, rely on the assumptions that investors know the value of the underlying asset's volatility, and that they agree on this value. These assumptions are unrealistic and very problematic, as they imply that investors agree on the value of the option, and furthermore they imply that the option is redundant.

MS allows extending the Black and Scholes model to incorporate uncertainty and disagreement regarding the value of the underlying asset's volatility. We present here such an extension. While the model presented here is very simplified, we believe that it hints toward the effects of uncertainty and disagreement regarding the volatility in more complex settings. We find that uncertainty and disagreement regarding the volatility can explain the empirically observed trading volume in options, the implied volatility smile, and the various findings regarding the implied volatility term structure.

[16] When there is a correlation between investors' degree of risk aversion and their estimation of $f(\sigma)$, additional effects may be observed. See M. Levy (1999a).

BIBLIOGRAPHY

Admati, Anat, and Paul Pfleiderer. (1988). "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1.

Allais, Maurice. (1953). "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulates et Axioms de l'Ecole Americaine," *Econometrica*, 21.

Allais, Maurice. (1988). "The General Theory of Random Choices in Relation to the Invariant Cardinal Function and the Specific Probability Function. The U Model. A General Overview," in B. Munier, ed., *Risk, Decision and Rationality* (Dordrecht: Reidel.

Amihud, Y., B. J. Christensen, and H. Mandelson. (1992). "Further Evidence on the Risk-Return Relationship," Working paper, Stanford University.

Anderson, P. W., J. Arrow, and D. Pines, eds. (1988). *The Economy as an Evolving Complex System*. Redwood City, CA: Addison-Wesley.

Arkes, H. R., and C. Blummer. (1985). "The Psychology of Sunk Cost," *Organizational Behavior and Decision Processes*, 35.

Arrow, K. J. (1965). *Aspects of the Theory of Risk Bearing*. Helsinki: Helsinkiyrjö Jahnssonin Säätiö.

Arrow, Kenneth J. (1971). *Essays in the Theory of Risk Bearing*. North-Holland, Amsterdam.

Arrow, Kenneth, J. (1982, January). "Risk Perception in Psychology and Economics," *Economic Inquiry*.

Arthur, W. Brian. (1994). "Inductive Reasoning and Bounded Rationality (The El Farol Problem)," *American Economic Review* (Papers and Proceedings), 84; http://www.santafe.edu/arthur/Paper/El_Farol.html.

Arthur, W. B., J. H. Holland, B. Lebaron, R. G. Palmer, and P. Taylor. (1997). "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*. Redwood City, CA: Addison-Wesley.

Axelrod, Robert (1985). *The Evolution of Cooperation*. Basic Books, New York.

Axelrod, Robert (1988). *Complexity of Cooperation*, http://www.pscs.umich.edu/Software/ComplexCoop.html, University of Michigan Program for the Study of Complex Systems Contact pscs@umich.edu. Revised May 1998. The quote is from *Robert Axelrod, The Evolution of Cooperation*, http://socrates.berkeley.edu/~jgwacker/GA/Axelrod.html, Last modified 28 July 1997.

Bak P., M. Paczuski, and M. Shubik. (1997). "Price Variations in a Stock Market with Many Agents," *Physica A* 246.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. (1998). "A Model of Investor Sentiment," *Journal of Financial Economics*, 49.

Beebout, Harold. (1999). *What Is Microsimulation*? Washington, DC, Office Mathematica Policy Research, Inc. http://www.mathematica-mpr.com/math-2.htm, Last Mod: 19 June 1999. HBEEBOUT@MATHEMATICA-MPR.COM.

Bell, David E. (1982). "Regret in Decision Marketing under Uncertainty," *Operation Research*, 30.

Benartzi, Shlomo, and Richard Thaler. (1995). "Myopic Loss Aversion and the Equity Premium Puzzle," *The Quarterly Journal of Economics*, 110(1).

Benartzi, Shlomo, and Richard Thaler. (1999). "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments," *Management Science*, 45(3).

Bettelheim, E., and B. Lehmann. (2000). "Microscopic Simulation of Reaction-Diffusion Processes," *Annual Reviews of Computational Physics* vol. VII, D. Stauffer (ed.). Singapore: World Scientific, http://complex.fiz.huji.ac.il/~eldadb/amirim/new.html.

Biham, O., O. Malcai, M. Levy, and S. Solomon. (1998, 1999). "Generic Emergence of Power Law Distributions and Levy-Stable Intermittent Fluctuations in Discrete Logistic Systems," *Physical Review E* 58(2) Aug. 1998 and O. Malcai, O. Biham, and S. Solomon, "Power-Law Distributions and Levy-Stable Intermittent Fluctuations in Stochastic Systems of Many Autocatalytic Elements," *Physical Review E.* 60(2) Aug. 1999.

Biham, O., A. A. Middleton, and D. Levine. (1992). "Self-Organization and a Dynamical Transition in Traffic Flow Models," *Phys. Rev.*, A46.

Black, F. (1986, July). "Noise," *Journal of Finance*, 41.

Black, F., M. C. Jensen, and M. Scholes. (1972). "The Capital Asset Model: Some Empirical Tests," in Michael C. Jensen, ed., *Studies in the Theory of Capital Markets*. New York: Praeger.

Black, F., and Scholes, M. (1973, May/June). "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*.

Blume, M. E., and I. Friend. (1975). "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *Journal of Finance*, 30(2).

Blume, M., J. Crockett, and I. Friend. (1974). "Stockownership in the United States: Characteristics and Trends," *Survey of Current Business*.

Bossaerts, P., Kleiman, and C. Plott. (1999). "Price Discovery in Financial Markets: The Case of the CAPM," *California Institute of Technology*.

Bouchaud, J-P, and M. Potters. (1997). *Théorie des Risques Financiers*. Aléa Saclay, Paris: Diffusion Eyrolles.

Brennan, M. (1970). "Taxes, Market Valuation and Corporate Financial Policy," *National Tax Journal*, 23(4).

Brockner, J. (1992). "The Escalation of Commitment to a Failing Course of Action, Toward Theoretical Progress," *Academy of Management Review*, 17(1).

Bromley, D. W. (1982). "Entitlements, Missing Markets, and Environmental Uncertainty," *Journal of Environmental Economics and Management*, 18.

Budescu, D. V., and W. Weiss. (1987). "Reflection of Transitive and Intransitive Preferences, a Test of Prospect Theory," *Organizational Behavior and Human Decision Processes*, 39.

Bunde, A., and S. Havlin. (1996). *Fractals and Disordered Systems*. Berlin-Heidelberg: Springer.

Busshaus, C., and H. Rieger. (1999). "A Prognosis Oriented Microscopic Stock Market Model," *Physica A*, 267.

Cadsby, C. B., and E. Maynes. (1998). "Laboratory Experiments in Corporate and Investment Finance: A Survey," *Managerial and Decision Economics*.

Caldarelli, G., M. Marsili, and Y.-C. Zhang. (1998). "A Prototype Model of Stock Exchange," *Europhysics Letters*, 40. See also M. Marsili, S. Maslov, and Y-C Zhang, "Dynamical Optimization Theory of a Diversified Portfolio," cond-mat/9801239.

Caldwell, S. B. (1993). "What Is Dynamic Microsimulation and the Corsim 3.0 Model?" sbc1@cornell.edu. *Institute for Public Affairs, Department of Sociology, Cornell University*, http://www.strategicforecasting.com/pubs/99-misc/theory93.html.

Camerer, Colin, F., and Teck-Hua Ho, "Nonlinear Weighting of Probabilities and Violations of the Betweenness Axiom," Unpublished manuscript, The Wharton School, University of Pennsylvania.

Campbell, J. Y., A. W. Lo, and A. C. Mackinlay. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Casey, J. T. (1994). "Buyers' Pricing Behavior for Risky Alternatives, Encoding Processes and Preference Reversals. *Management Science*, 40(6).

Castiglione, R., G. Mannella, S. Motta, and G. Nicosia. (1999). *International Journal of Modern Physics C*, 10.

Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza. (1996). *The History and Geography of Human Genes*, Princeton, NJ: Princeton University Press. See, however, Renfrew Colin, *Archeology and Language: The Puzzle of Indo-European Origins*, Cambridge University Press, 1990; S. E. van der Leeuv, ed., *Archeological Approaches to the Study of Complexity*, University of Amsterdam, 1981; N. Winder, "Jonah and the Flatworm," *Human Ecology Review*, 1999.

Chang, I., and D. Stauffer. (1999). "Fundamental Judgement in Cont-Bouchaud Herding Model of Market Fluctuations," *Physica A* 264.

Chang, O. H., D. R. Nichols, and J. J. Schultz. (1987). "Taxpayer Attitudes toward Tax Audit Risk, *Journal of Economic Psychology*, 8.

Chatagny, R., and B. Chopard. (1997). *International Conference on High Performance Computing and Networks*, Vienna.

Chew, Soo Hong, and Kenneth MacCrimmon. (1979). "Alpha-nu Choice Theory: An Axiomation of Expected Utility." University of British Columbia, Faculty of Commerce and Business Administration Working Paper 669.

Chowdhury, D., L. Santen and A. Shadshaneider. (2000). Physics Reports preprint.

Chowdhury, D., and D. Stauffer. (1999). "A Generalized Spin Model of Financial Markets," *Eur. Phys. J.* B8.

Clyde J., J. Schleier-Smith, G. Tseng, R. Latham, and D. Higley. (1996). http://tqd.advanced.org/3471/nuclear_weapons_fission_diag.html, tq-nuke@tjhsst.edu *Nuclear Physics Past Present and Future: Nuclear Weapons*; Site last updated: October 28, 1996.

Coady, D. P. (1995, May). "An Empirical Analysis of Fertilizer Use in Pakistan," *Economica*. 62.

Cont, R., M. Potters, J-P. Bouchaud. (1997). "Scaling in Stock Market Prices: Stable Laws and Beyond." *Science and Finance Group*, Working Paper 97-02.

Cox, J. C., and S. A. Ross. (1975). "The Pricing of Options for Jump Processes," Working Paper 2-75, Rodney L. White Center for Financial Research, University of Pennsylvania.

Cox, J. C., S. A. Ross, and M. Rubinstein. (1979). "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, 7.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyan. (1999). "Investor Psychology and Market Under- and Over-Reactions," *Journal of Finance*, 53.

D'Aveni, R. A., "Dependability and Organizational Bankruptcy: An Application of Agency and Prospect Theory," *Management Science*, 35(9).

Davis, D. D., and C. A. Holt. (1993). *Experimental Economics*, Princeton, NJ: Princeton University Press.

Dawkins, R. (1987). *Climbing Mount Improbable*. New York: W. W. Norton.

De Bondt, Werner, and Richard Thaler. (1985). "Does the Stock Market Overreact?" *Journal of Finance*, 40.

DeLong, J. Bradford, Andrei Shleifer, H. Lawrence Summers, and Robert J. Waldmann. (1990). "Noise Traders Risk in Financial Markets," *The Journal of Financial Economics*, 98.

Demographic-Sociological Microsimulation Program. (1976). Hammel, E. A., D. Hutchinson, K. W. Wachter, R. Lundy, and R. Deuel. *The SOCSIM Demographic-Sociological Microsimulation Program: Operating Manual*, Institute for International Studies, University of California, Berkeley, 1976, 1. http://arrow.qal.berkeley.edu/programs/socsim/, Socsim@demog.berkeley.edu, Last modified 10-Apr-96.

Derman, E., and I. Kani. (1994, Jan.). "The Volatility Smile and Its Implied Tree," *Quantitative Strategies Publications*. Goldman Sachs.

Diamond, W. D. (1988). "The Effect of Probability and Consequence Levels on the Focus of Consumer Judgements in Risky Situations," *Journal of Consumer Research*, 5.

Douglas, A. A., and R. J. Lewis. (1970, 1971, 1998). "Trip Generation Techniques," *Traffic Engineering and Control*, Vols. 12.7, 1970, and 12.10, 1971; The quotation is from *Microsimulation*, http://www.glue.umd.edu/~jrect/Microsimulation.htm; Last Mod: 08 Dec. 1998.

Douglas, G. W. (1969, Spring). "Risk in the Equity Markets: An Empirical Appraisal of Market Efficiency," *Yale Economics Essays*, 9.

Drake, P. R., and J. L. Freedman. (1993). "Deciding Whether to Seek a Bargain, Effects of Both Amount and Percentage Off," *Journal of Applied Psychology*, 78(6).

Edgar, P. E. (1991). *Chaos and Order in Capital Markets: A New View of Cycles, Prices and Market Volatility*. New York: John Wiley and Sons.

Edwards, Ward. (1953). "Probability Preference in Gambling," *American Journal of Psychology*, 66.

Edwards, Ward. (1954). "Probability-Preferences among Bets with Differing Expected Values," *American Journal of Psychology*, 67.

Egenter, E., T. Lux, and D. Stauffer. (1999). "Finite Size Effects in Monte Carlo Simulations of Two Stock Market Models," *Physica A* 268.

Epstein, Joshua M., and Robert L. Axtell. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge, MA: Complex Adaptive Systems, MIT Press; see also Giles Wright, "Rise and Fall," *New Scientist*, 4 October 1997. http://www.newscientist.com/ns/971004/features.html.

Fama, E. F. (1963). "The Distribution of the Daily First Differences of Stock Prices: A Test of Mandelbrot's Stable Paretian Hypothesis." Unpublished doctoral dissertation, University of Chicago.

Fama, E. F. (1965a). "The Behavior of Stock Prices," *Journal of Business*, 38(1).

Fama, E. F. (1965b). "Mandelbrot and the Stable Paretian Hypothesis," *Journal of Business*, 4(36).

Fama, E., and K. French. (1988). "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96.

Fama, E., and K. French. (1992). “The Cross-Section of Expected Stock Returns,” *Journal of Finance*.

Farmer, J. D. (1998). “Market Force, Ecology and Evolution,” e-print adap-org/9812005. See also J. D. Farmer and S. Joshi, “Market Evolution Toward Marginal Efficiency” *SFI Report* 1999.

Fiegenbaum, A. (1990). “Prospect Theory and the Risk-Return Association: An Empirical Examination in 85 Industries,” *Journal of Economic Behavior and Organizations*, 14.

Fiegenbaum, A., and H. Thomas. (1988). “Attitudes Toward Risk and the Risk-Return Paradox, Prospect Theory Explanations,” *Academy of Management Journal*, 31(1).

Fiorina, M. P., and C. R. Plott. (1978). “Committee Decisions under Majority Rule: An Experimental Study,” *American Political Science Review*, 72(2).

Fishburn, Peter, C. (1982). “Nontransitive Measurable Utility,” *Journal of Mathematical Psychology*, 26.

Friedman, Daniel. (1989). “The S-Shaped Value Function as a Constrained Optimum,” *American Economic Review*, 79(5).

Friedman, M. (1953a). “The Methodology of Positive Economics,” in *Essays in Positive Economics*. Chicago: University of Chicago Press.

Friedman, Milton. (1953b). “The Case for Flexible Exchange Rates,” in *Essays in Positive Economics*. Chicago: University of Chicago Press.

Friedman, M., and L. J. Savage. (1948, August). “The Utility Analysis of Choices Involving Risk,” *Journal of Political Economy*.

Friend, I., and M. E. Blume. (1975, December). “The Demand for Risky Assets,” *The American Economic Review*.

Galam, S. (1990). “Social Paradoxes of Majority Rule Voting and Renormalization Group,” *Journal of Statistical Physics*, 61.

Galam, S. (1996). “Fragmentation versus Stability in Bimodal Coalitions,” *Physica A* 230.

Galam, S. (1997, May 30). “Le Dangereux Seuil Critique du FN (Front National),” *Le Monde*.

Galam, S. (1998). “Comment on a Landscape Theory of Aggregation,” *British Journal of Political Science*, 28.

Garland, H., and S. Newport. (1991). “Effects of Absolute and Relative Sunk Costs on the Decision to Persist with a Course of Action,” *Organizational Behavior and Human Decision Processess*, 48.

Geske, R. (1979). “The Valuation of Compound Options,” *Journal of Financial Economics*, 7.

Giffith, R. M. (1949). “Odds Adjustments by American Horse Race Bettors,” *American Journal of Psychology*, 62.

Gilovich, T., R. Vallone, and A. Tversky. (1985). “The Hot Hand in Basketball: On the Misperception of Random Sequences,” *Cognitive Psychology*, 17.

Glance, N. S., and B. A. Huberman. (1994, March). “Dynamics of Social Dilemmas,” *Scientific American*.

Glance Natalie Sandrine. (1983, June). *Dynamics with Expectations*, PhD thesis, Physics Department, Stanford University. http:// messenger.netscape.com/ bookmark/ 4_5/ messengerstart.html http:// www.parc.xerox.com/ spl/ groups/ dynamics/ www/ thesis.html glance@xerox.fr.

Glosten, L. R., and P. R. Milgrom. (1985). “Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders,” *Journal of Financial Economics*, 14.

Goldenberg J., D. Mazursky, and S. Solomon. (1999, May). “Towards Identifying the Inventive Templates of New Products: Channeled Ideation Approach,” *Journal of Marketing Research*, 36.

Gordon, M. J., G. E. Paradis, and C. H. Rorke. (1972). “Experimental Evidence on Alternative Portfolio Decision Rules,” *American Economic Review*, 62(1).

Gregory, R. (1986). "Interpreting Measures of Economic Loss, Evidence from Contingent Valuation and Experimental Studies," *Journal of Environmental Economics and Management*, 13.

Grossman, S., and J. Stiglitz. (1980). "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70.

Gudowski, W., and J. Kierkegaard. (1999). http://www.neutron.kth.se/research/mcsim.html; Last update: 1999-06-14 by Janne Wallenius; Monte-Carlo simulations of the nuclear reactor core. Nuclear Technology Center at KTH.

Hadar, J., and W. Russell. (1969). "Rules for Ordering Uncertain Prospects," *American Economic Review*, 59.

Haka, S., L. Friedman, and V. Jones. (1986). "Functional Fixation and Interference Theory: A Theoretical and Empirical Investigation," *The Accounting Review*, 61(3).

Hakansson, N. H. (1971, January). "Capital Growth and the Mean-Variance Approach to Portfolio Selection," *Journal of Financial and Quantitative Analysis*.

Hanoch, G., and H. Levy. (1969). "The Efficiency Analysis of Choices Involving Risk," *Review of Economic Studies*, 36.

Hardie, B. G. S., E. J. Johnson, and P. S. Fader, (1993). "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12(4).

Hardin, G. (1968). "The Tragedy of the Commons," *Science*, 162.

Harrison, G. W. (1986). "An Experiment for Risk Aversion," *Economic Letters*.

Hellthaler, T. (1995). "The Influence of Investor Number on a Microscopic Market," *International Journal of Modern Physics*, C6.

Hellwig, M. (1980). "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 22.

Hershey, J. C., and Schoemaker, P. H. (1980). "Prospect Theory's Reflection Hypothesis: A Critical Examination," *Organizational Behavior and Human Performance*, 25.

Heston, S. L. (1993). "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6(2).

Hirst, D. E., E. J. Joyce, and M. S. Schadewald. (1994). "Mental Accounting and Outcome Contiguity in Consumer-Borrowing Decisions," *Organizational Behavior and Human Decision Processes*, 58.

Hogarth, R. M., and H. J. Einhorn. (1990). "Venture Theory, a Model of Decision Weights," *Management Science*, 36(7).

Hong, H., and J. C. Stein. (1999). "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets," *Journal of Finance*. forthcoming.

Huberman Bernardo. (1997). *Xerox Palo Alto Research Center Internet Page* http://www.parc.xerox.com/spl/groups/dynamics/dynamics.shtml This page has been last changed on August 18, 1997 hogg@parc.xerox.com.

Hull, John, C. (1997). *Options, Futures, and other Derivatives*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall.

Hull, J. C., and A. White. (1987). "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42.

Hull, J. C., and A. White. (1988). "An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility," *Advances in Futures and Options Research*, 3.

Ibbotson Associates. (1999). *Stocks, Bonds, Bills and Inflation*. Chicago: Ibbotson Associates.

Jagannathan, R., and Z. Wang. (1993, November). "The CAPM Is Alive and Well," Federal Reserve Bank of Minneapolis, Staff Report 165.

Jegadeesh, N., and S. Titman. (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48.

Jegers, M. (1991). "Prospect Theory and the Risk-Return Relation: Some Belgian Evidence, *Academy of Management Journal*, 34(1).

Kahneman, Daniel, J. Knetsch, and R. Thaler. (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98.

Kahneman, Daniel, and A. Tversky. (1979). "Prospect Theory of Decisions under Risk," *Econometrica*, 47(2).

Kameda, T., and J. H. Davis. (1990). "The Function of the Reference Point in Individual and Group Risk Decision Making," *Organizational Behavior and Human Decision Process*, 46.

Kanto, A., G. Rosenqvist, and A. Suvas. (1992). "On Utility Function Estimation of Race Track Bettors," *Journal of Economic Psychology*, 13.

Karpoff, Jonathan. (1987). "The Relationship between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22.

Kaufman, M., J. Urbain, and R. Thomas. (1985). *Journal of Theoretical Biology*, 114.

Kim, G. W., and H. M. Markowitz. (1989). "Investment Rules, Margin, and Market Volatility," *Journal of Portfolio Management*, 16.

Kiviat, Philip, J., Richard Villanueva, and Harry Markowitz. (1968). *The Simscript II Programming Language*. Englewood Cliffs, NJ: Prentice Hall.

Kohl, R. (1997). "The Influence of the Number of Different Stocks on the Levy, Levy Solomon Model," *International Journal of Modern Physics C*, 8.

Kohring, G. A. (1996). "Ising models of Social Impact: The Role of Cumulative Advantage," *Journal Physique*, 6.

Kroll, Y., and H. Levy. (1992). "Further Tests of the Separation Theorem and the Capital Asset Pricing Model," *American Economic Review*.

Kroll, Y., H. Levy, and H. M. Markowitz. (1984). "Mean-Variance versus Direct Utility Maximization," *Journal of Finance*.

Kroll, Y., H. Levy, and A. Rapoport. (1988a). "Experimental Test of the Mean-Variance Model for Portfolio Selection," *Organizational Behavior and Human Decision Processes*, 42.

Kroll, Y., H. Levy, and A. Rapoport. (1988b). "Experimental Tests of the Separation Theorem and the Capital Asset Pricing Model," *American Economic Review*.

Kydland, F. E., and E. C. Prescott. (1982). "Time to Build and Aggregate Fluctuations," *Econometrica*, 50.

Kyle, Albert S. (1984). "Market Structure, Information, Futures Markets, and Price Formation." In Gary G. Storey, Andrew Schmitz, and Alexander H. Sarris, eds., *International Agricultural Trade*. Boulder and London: Westview Press.

Kyle, Albert S. (1985b). "Informed Speculation with Imperfect Competition," Unpublished manuscript.

Kyle, Albert S. (1985a). "Continuous Auctions and Insider Trading," *Econometrica*, 53.

La Recherche. (1999). 322 July/August 1999 contains various articles on aging. See also La Science, *Les Mathematiques Sociales*, Dossier Hors-Serie. Edition Francaise de Scientific American, Juillet 1999.

Latané, H. A. (1959, April). "Criteria for Choices Among Risky Ventures," *Journal of Political Economics*.

Leclerc, F., B. H. Schmidt, and L. Dube. (1995). "Waiting Time and Decision Making, is Time Like Money?" *Journal of Consumer Research*, 22.

Lee, C. K., R. G. Klopp, R. Weindruch, and T. A. Prolla. (1999). "Gene Expression Profile of Aging and Its Retardation by Caloric Restriction, *Science*, 285.

Levy, H. (1973, March). "The Demand for Asset under Conditions of Risk," *Journal of Finance*.

Levy, H. (1978). "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio," *American Economic Review*, 68.

Levy, H. (1992, April). "Stochastic Dominance and Expected Utility: Survey and Analysis," *Management Science*.

Levy, H. (1994). "Absolute and Relative Risk Aversion: An Experimental Study," *Journal of Risk and Uncertainty*.

Levy, H. (1997, February). "Risk and Return: An Experimental Analysis," *International Economic Review*.

Levy, H. (1998). *Stochastic Dominance: Investment Decision Making under Uncertainty*. Kluwer: Academic Publishers.

Levy, H. (1999). "Cumulative Prospect Theory and the CAPM," Hebrew University working paper.

Levy, H., and K. C. Lim. (1998). "The Economic Significance of the Cross-Sectional Autoregressive Model: Further Analysis," *Review of Quantitative Finance and Accounting*, 11.

Levy, H., and Z. Wiener. (1999). "Prospect Theory, Utility Theory, and Market Inefficiency," Hebrew University, working paper.

Levy, Haim, and Z. Wiener. (1998). "Stochastic Dominance and Prospect Dominance with Subjective Weighting Functions," *Journal of Risk and Uncertainty*, 16.

Levy, M., and H. Levy. (1996, May). "The Danger of Assuming Homogeneous Expectations," *The Financial Analyst Journal*.

Levy, M., Levy H., and S. Solomon. (1994). "A Microscopic Model of the Stock Market: Cycles, Booms, and Crashes," *Economics Letters*, 45.

Levy, M., Levy H., and S. Solomon. (1995). "Simulation of the Stock Market: The Effects of Microscopic Diversity," *Journal de Physique I*, 5.

Levy, M., N. Persky, and S. Solomon. (1996). "The Complex Dynamics of a Simple Stock Market Model." *International Journal of High Speed Computing*, 8.

Levy, M., and S. Solomon. (1996). "Power Laws are Logarithmic Boltzmann Laws," *International Journal of Modern Physics C*, 7(4).

Levy, M., and S. Solomon. (1997). "New Evidence for the Power Law Distribution of Wealth," *Physica*, A, 242.

Levy, Moshe. (1997). "Nonuniqueness and Nonexistence of the CAPM Equilibrium," Hebrew University working paper.

Levy, Moshe. (1998a). "Are Rich People Smarter?" UCLA Anderson School of Management working paper.

Levy, Moshe. (1998b). "Wealth Inequality and the Distribution of Stock Returns," UCLA Anderson School of Management working paper.

Levy, Moshe. (1999a). "Option Pricing With Uncertainty and Disagreement Regarding the Underlying Asset's Volatility," Hebrew University working paper.

Levy, Moshe. (1999b). "Prospect Theory, Asset Allocation, and the Equity Premium," Hebrew University working paper.

Lintner, J. (1969, December). "The Aggregation of Investors' Diverse Judgements and Preferences in Purely Competitive Security Markets," *Journal of Finance Quantitative Analysis*, 4.

Lintner, John. (1965a). "Security Prices and Risk: The Theory of Comparative Analysis of AT & T and Leading Industrials," Paper presented at the conference on the economics of regulated public utilities, Chicago.

Lintner, John. (1965b, December). "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20.

Loewenstein, G. G. (1988). "Frames of Mind in Intertemporal Choice," *Management Science*, 34(2).

Loomes, G., and R. Sugden, R. (1982). "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty." *The Economic Journal*, 92.

Lotka, A. J. ed. (1925). *Elements of Physical Biology*. Baltimore: Williams and Wilkins.

Louzon, Y., S. Solomon, H. Atlan, and A. Cohen. (2000). "Discretion in Modeling the Immune System." In preparation.

Lux, T. (1995). "Herd Behaviour, Bubbles and Crashes," *Economic Journal*, 105, 881.

Lux, T. (1996). "The Stable Paretian Hypothesis and the Frequency of Large Returns: An Examination of Major German Stocks," *Applied Financial Economics*, 6, 463.

Lux, T., and M. Marchesi. (1999). "Volatility Clustering in Financial Markets: A Micro-Simulation of Interacting Agents," *Nature*, 397, 498; and "Scaling and Criticality in a Stochastic Multi-Agent Model of a Financial Market" to appear in *IJTAF*.

Machina, Mark J. (1982). "Expected Utility Analysis without the Independence Axioms," *Econometrica*, 50.

Mandelbrot, B. (1963a). "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 4.

Mandelbrot, B. (1963b). "New Methods in Statistical Economics," *Journal of Political Economy*, 61.

Mantegna, R. N. (1991). "Lévy Walks and Enhanced Diffusion in the Milan Stock Exchange," *Physica A*, 179.

Mantegna, R. N., and H. E. Stanley (1995). "Scaling Behavior in the Dynamics of an Economic Index." *Nature*, 376.

Mantegna, R. N. and H. E. Stanley. (1999). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge, England: Cambridge University Press.

Marchak, J. (1964, April). "Actual vs. Consistent Decision Behavioral," *Behavioral Science*.

Markowitz, H. M. (1952a, March). "Portfolio Selection," *Journal of Finance*.

Markowitz, H. M. (1952b). "The Utility of Wealth," *Journal of Political Economy*, 60.

Markowitz, Harry M. (1963). *Simscript: A Simulation Programming Language*, Englewood Cliffs, NJ: Prentice Hall.

Markowitz, H. M. (1976, December). "Investment for the Long Run: New Evidence for an Old Rule," *Journal of Finance*.

Markowitz, H. M. (1990). "Risk Adjustment," *Journal of Finance, Accounting and Auditing*.

Martinez-Vazquez, J., G. B. Harwood, and E. R. Larkins. (1992). "Withholding Position and Income Tax Compliance, Some Experimental Evidence," *Public Finance Quarterly*, 20(2).

Mayer, P. C. (1995). "Electricity Conservation, Consumer Rationality versus Prospect Theory," *Contemporary Economic Policy*, 13.

Mayshar, Y. (1979). "Transaction Costs in a Model of Capital Market Equilibrium," *Journal of Political Economy*, 87.

McKee, M. (1989). "Intra-Experimental Income Effects and Risk Aversion," *Economic Letters*.

Merton, R. C. (1971). "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3.

Merton, R. C. (1976). "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3.

Merton, R. C. (1987). "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42.

Merton, Robert C., and Paul A. Samuelson. (1974). "Fallacy of the Log-Normal Approximation to Optimal Decision Making over Many Periods," *Journal of Financial Economics*, 1(1).

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller. (1953). "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21.

Meyer, R. J., and J. Assuncao. (1990). "The Optimality of Consumer Stockpiling Strategies," *Marketing Science*, 9(2).

Miele, Frank. (1995). "Darwin's Dangerous Disciple, An Interview with Richard Dawkins," *Skeptic*, 3(4). http://www.skeptic.com/03.4.miele-dawkins-iv.

Miller, M., and M. Scholes. (1972). "Rate of Return in Relation to Risk: A Reexamination of Some Recent Findings," in Michael C. Jensen, ed., *Studies in the Theory of Capital Markets*. New York: Praeger.

Modigiliani, F., and M. Miller. (1958, June). "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*.

Modigiliani, F., and M. Miller. (1969, September). "Replying to Heins and Sprenkle," *American Economic Review*.

Mood, A. M., and F. A. Graybill. (1963). *Introduction to the Theory of Statistics*. New York: McGraw-Hill.

Mori, T., *et al*. (1997). "Persistent Quest–Research Activities 1996," The Japan Atomic Energy Institute.

Moss de Oliveira S., H. de Oliveira, and D. Stauffer. (1999). *Evolution, Money, War and Computers*. B. G. Teubner Stuttgart-Leipzig.

Mosteller, F., and P. Nogee. (1951). "An Experimental Measurement of Utility," *Journal of Political Economy*, 59.

Nagel, K., J. Esser, and M. Rickert. (2000). *Annual Reviews of Computational Physics, vol. VII*, D. Stauffer, ed. Singapore: World Scientific,

New Scientist. (1997, July 5). "Ancestral Echoes," http://marijuana.newscientist.com/ns/970705/features.html, ©Copyright New Scientist, IPC Magazines Limited, 1997.

Nielsen, L., T. (1988). "Uniqueness of Equilibrium in the Classical CAPM," *Journal of Financial and Quantitative Analysis*, 23, 3.

Officer, R. R. (1972). "The Distribution of Stock Returns," *Journal of the American Statistical Association*, 67.

Ohbuchi, Yutaka. (1999). "Tamagotchi Eggs Go Bad," *videogames.com Game Spot News*:©Zdnet, 02/25/99, http://headline.gamespot.com/news/99_02/25_vg_tama/index.html.

Orcutt, G. H., S. B. Caldwell, and R. Wertheimer. (1976). *Policy Exploration through Microanalytic Simulation*. Washington, DC: The Urban Institute.

Palmer, R. G., W. B. Arthur, J. H. Holland, B. LeBaron, and P. Tayler. (1994). "Artificial Economic Life: A Simple Model of a Stock Market," *Physica D*, 75.

Payne, J. W., D. J. Laughhunn, and R. Crum. (1984). "Multiattribute Risky Choice Behavior: The Editing of Complex Prospects," *Management Science*, 30(11).

Penna, T. J. P. (1995). "A Bit-String Model for Biological Aging." *Journal of Statistical Physics*, 78.

Plott, C. R. (1979). "The Application of Laboratory Experimental Methods in Public Choice," in C. S. Russell, ed., *Collective Decision Making*, Washington, DC: Resources of the Future.

Plott, C. R. (1986). "Rational Choice in Experimental Markets," *Journal of Business*.

Plott, C. R., and V. L. Smith. (1978). "An Experimental Examination of Two Exchange Institutions," *Review of Economic Studies*.

Plott, C. R., and S. Sunder. (1982). "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational-Explanation Models," *Journal of Political Economy*, 90.

Poterba, J. M., and Summers, L. H. (1988). "Mean Reversion in Stock Returns: Evidence and Implications," *Journal of Financial Economics*, 22.

Pratt, J. W. (1964). "Risk Aversion in the Small and in the Large," *Econometrica*, 32.

Preston, G., and Philip Baratta. (1948). "An Experimental Study of the Action-Value of an Uncertain Outcome," *American Journal of Psychology*, 61.

Pryor, Rich. (1999). *ASPEN, A smart agent-based economics simulation model*, Sandia National Laboratories, http://www-aspen.cs.sandia.gov/e-mail: rjpryor@sandia.gov.

Qualls, W. J., and Pluto, C. P. (1989). "Organizational Climate and Decision Framing: An Integrated Approach to Analyzing Industrial Buying Decisions," *Journal of Marketing Research*, 26.

Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank Dependent Model*. Boston, Kluwer Academic Publishers.

Quiggin, John. (1982). "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3.

Renfrew, Colin. (1990). *Archeology and Language: The Puzzle of Indo-European Origins*. Cambridge: Cambridge University Press. See also *Archeological Approaches to the Study of Complexity*, S. E. van der Leeuv ed., University of Amsterdam, 1981, and N. Winder, "Jonah and the Flatworm," *Human Ecology Review*, 1999.

Roll, R. (1968). "The Efficient Market Model Applied to U.S. Treasury Bill Rates," Unpublished Doctoral dissertation, University of Chicago.

Roll, R. (1977). "A Critique of the Asset Pricing Theory's Test, Part I: On Past and Potential Testibility of Theory," *Journal of Financial Economics*.

Roll, Richard. (1986). "The Hubris Hypothesis of Corporate Takeovers," *Journal of Business*, 59(2), Part 1.

Ross, S. A. (1976). "The Arbitrage Theory of Capital Asset Pricing," *Econometrica*.

Roth, Alvin, E., ed. (1987). *Laboratory Experimentation in Economics: Six Points of View*. Cambridge: Cambridge University Press.

Rubinstein, M. (1983). "Displaced Diffusion Option Pricing," *Journal of Finance*, 38.

Rubinstein, M. (1985). "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978," *Journal of Finance*, 40.

Russell, Thomas, and Richard H. Thaler. (1985). "The Relevance of Quasi Rationality in Competitive Markets," *American Economic Review*, 75.

Sahimi, M. (1994). *Application of Percolation Theory*. London: Taylor and Francis.

Salminen, P. (1994). "Solving the Discrete Multiple Criteria Problem Using Linear Prospect Theory," *European Journal of Operational Research*, 72.

Samuelson, P. A. (1989). "The Judgement of Economic Science on Rational Portfolio Management: Timing and Long Horizon Effects," *The Journal of Portfolio Management*.

Samuelson, P. A. (1994). "The Long Term Case for Equities and How It Can Be Oversold," *The Journal of Portfolio Management*.

Samuelson, Paul A. (1990). "Asset Allocation Could Be Dangerous to Your Health," *Journal of Portfolio Management*.

Samuelson, Paul A. (1997). "Estimating Probabilities Relevant to Calculating Relative Risk-Corrected Returns of Alternative Portfolios," *Journal of Risk and Uncertainty*, 15(3).

Schaubroeck, J., and Davis, E. (1994). "Prospect Theory Predictions When Escalation Is Not the Only Chance to Recover Sunk Costs," *Organizational Behavior and Human Decision Processes*, 57.

Schelling, Thomas C. (1978). *Micro Motives and Macro Behavior*. W. W. Norton, New York. See also Internet Education Project page of the Carl Vinson Institute of Government. *Artificial Societies Simulations* http://iep.cviog.uga.edu:2000/SIM/intro.htm: Schelling Model.

Schoemaker, P. J., and H. C. Hunreuther. (1979). "An Experimental Study of Insurance Decisions," *The Journal of Risk and Insurance*, 46(4).

Sebora, T. C., and J. R. Cornwall. (1995). "Expected Utility Theory vs. Prospect Theory," *Journal of Mathematical Economics*, 7(1).

Sharpe, W. (1991, June). "Capital Asset Pricing with and without Negative Holdings," *Journal of Finance*.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19.

Shefrin, H., and M. Statmen. (1993). "Behavioral Aspect of the Design and Marketing of Financial Products," *Financial Management*, 22(2).

Shefrin, Hersh M., and Meir Statman. (1985). "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance* 40.

Shiller, Robert J. (1981). "Do Stock Returns Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*.

Shnerb, N., Y. Louzon, E. Bettelheim, and S. Solomon. (1999). "The Importance of Being Discrete," submitted to Proceedings of the National Academy of Sciences and http://xxx.lanl.gov/ps/adap-org/9912005.

Smith, V. L. (1962). "An Experimental Study of Comparative Market Behavior," *Journal of Political Economics*.

Smith, V. L. (1976). "Experimental Economics: Induced Value Theory," *American Economic Review*.

Smith, V. L. (1982). "Microeconomic Systems as an Empirical Science," *American Economic Review*.

Smith, V. L. (1991). "Rational Choice: The Contrast between Economics and Psychology," *Journal of Political Economy*.

Solomon, S. (1995). "The Microscopic Representation of Complex Macroscopic Phenomena," in D. Stauffer, ed. *Annual Reviews of Computational Physics II*. Singapore: World Scientific.

Solomon, S. (1998). "Stochastic Lotka-Volterra systems of competing auto-catalytic agents lead generically to truncated Pareto power wealth distribution, truncated Levy distribution of market returns, clustered volatility, booms and craches," in *Computational Finance 97*, Eds. A. P. N. Refenes, A. N. Burgess, J. E. Moody. Kluwer Academic Publishers.

Solomon, S. (1999). "Behaviorly realistic simulations of stock market; traders with a soul," *Computer Physics Communications*, 121 and 122.

Solomon, S., and M. Levy. (1996). "Spontaneous Scaling Emergence in Generic Stochastic Systems," *International Journal of Modern Physics C*, 7(5).

Solomon S., Gerard Weisbuch, Lucilla de Arcangelis, Naeem Jan, and Dietrich Stauffer. (2000). "Social Percolation Models," *Physica A*.

Sornette, D., and A. Johansen. (1997/1998). "Large Financial Crashes," *Physica A* 245, 1997; and "A Hierarchical Model of Financial Crashes," *Physica A* 261, 1998.

Stauffer, D. (1998). "Can Percolation Theory Be Applied to the Stock Market?" *Annalen der Physik*, and http://xxx.lanl.gov/cond-mat/9810162.

Stauffer, D. (1999). "Finite Size Effects in the Lux-Marchesi and Other Microscopic Market Models," Talk at the *Workshop on Economics with Heterogeneous Interacting Agents*, Genoa.

Stauffer, D., and A. Aharony. (1994). *Introduction to Percolation Theory*. London: Taylor and Francis.

Stauffer, D., and N. Jan. (Submitted). "Sharp Peaks in the Percolation Model for Stock Markets," submitted *Physica A*.

Stauffer, D., P. M. C. de Oliveira, and A. T. Bernardes. (1999). "Monte Carlo Simulation of Volatility Correlation in Microscopic Market Model," *International Journal of Theoretical and Applied Finance*, 2.

Stauffer, D., and D. Sornette. (1999). "Self-Organized Percolation Model for Stock Market Fluctuations," *Physica A* 271.

Steiglitz, K., M. L. Honig, L. M. Cohen. (1996). Chapter 1 in *Market-Based Control: A Paradigm for Distributed Resource Allocation*, S. Clearwater, ed. Hong Kong: World Scientific.

Stigler, G. J. (1964). "Public Regulation of the Securities Markets," *Journal of Business*, 37.

Stigler, George. (1966). *The Theory of Price*. New York: Macmillen.

Sullivan, Kevin. (1997; January 25). "A Chicken in Every Pocket," *Washington Post Foreign Service*, p. A01, http://www.washingtonpost.com/wp-srv/ inatl/ asia/ jan/ 29/ chicken.htm.

Summers, Lawrence, H. (1986, July). "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance*.

Swalm, R. O. (1966, November/December). "Utility Theory—Insights into Risk Taking," *Harvard Business Review*.

Takayasu, H. Miura, T. Hirabashi, and K. Hamada. (1992). "Statistical Properties of Deterministic Threshold Elements—The Case of Market Price," *Physica A* 184.

Teichmoeller, J. (1971). "A Note on the Distribution of Stock Price Changes," *Journal of the American Statistical Association*.

Thaler, R. (1985). "Mental Accounting and Consumer Choice," *Marketing Science*, 4(3).

Thaler, R. H., and E. Johnson. (1990). "Gambling with the House Money and Trying to Break Even: The Effect of the Prior Outcomes on Risky Choices," *Management Science*, 36.

Thaler, Richard, ed. (1993). *Advances in Behavioral Finance*. New York: Russel Sage Foundation.

Thaler, Richard. (1994). *Quasi Rational Economics*. New York: Russel Sage Foundation.

Thaler, Richard, and Eric J. Johnson. (1990). "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science*, 36(3).

Troitzsch K. G., U. Mueller, G. N. Gilbert, J. E. Doran, eds. (1996). *Social Science Microsimulation, Mathematical Methods and Models (Appl. Math.)*, Simulation and Modeling. New York: Springer-Verlag, http://www.springer-ny.com/catalog/np/nov96np/DATA/3-540-61572-5.html.

Tversky, A. (1989). "Quantifying the Market Risk Premium Phenomenon for Investment Decision Making," in W. Sharp and Katrina F. Sheferred, eds. The Institute of Chartered Financial Analysts (CFA).

Tversky, A., and Kahneman, D. (1981). "The Framing of Decisions and the Psychology of Choice," *Science*, 211.

Tversky, A., and Kahneman, D. (1986). "Rational Choice and the Framing of Decision," *Journal of Business*, 59(4).

Tversky, Amos, and Daniel Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5.

Uecker, W., A. Schepanski, and J. Shin. (1985). "Toward a Positive Theory of Information Evaluation, Relevant Tests of Competing Models in a Principal-Agency Setting," *The Accounting Review*, 60(3).

Volterra, V. (1926). "Fluctuations in the Abundance of a Species Considered Mathematically," *Nature*, 118, 1926, and "Variazioni e fluttuazioni del numero d'individui in specie animali conviventi," *Mem. R. Accad. Naz. dei Lincei.* Ser. VI, 2, 1926.

von Neuman, J., and O. Morgenstern. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

von Neuman, J., and O., Morgenstern. (1947). *Theory of Games and Economic Behavior*, 2nd ed. Princeton, NJ: Princeton University Press.

Weisbuch, G., A. Kirman, and D. Herreiner. (1999). "Market Organisation and Trading Relationships" *Economic Journal*.

Whyte, G. (1993). "Escalating Commitment in Individual and Group Decision Making: A Prospect Theory Approach," *Organizational Behavior and Human Decision Processes*, 54.

Wilde, L. (1980). "On the Use of Laboratory Experiments in Economics," in J. Pitt, ed., *The Philosophy of Economics*. Dordrecht, Holland: Reidel.

Wright, Giles. (1997, October 4). "Decline and Fall," *New Scientist*, http://www.newscientist.com/ ns/ 971004/ features.html.

Yaari, Menachem. (1987). "The Dual Theory of Choice Under Risk," *Econometrica*, 55(1).

Zajdenweber, D. (1994). "Propriétés Autosimilaires du CAC40," *Review d'Economie Politique*, 104.

Zorzenon dos Santos, R. M. (1999). *Annual Reviews of Computational Physics VI*, D. Stauffer, ed. Singapore: World Scientific.

INDEX