1
Introduction

FINANCIAL ECONOMICS is a highly empirical discipline, perhaps the most empirical among the branches of economics and even among the social sciences in general. This should come as no surprise, for financial markets are not mere figments of theoretical abstraction; they thrive in practice and play a crucial role in the stability and growth of the global economy. Therefore, although some aspects of the academic finance literature may seem abstract at first, there is a practical relevance demanded of financial models that is often waived for the models of other comparable disciplines.[1]

Despite the empirical nature of financial economics, like the other social sciences it is almost entirely nonexperimental. Therefore, the primary method of inference for the financial economist is model-based statistical inference—financial econometrics. While econometrics is also essential in other branches of economics, what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation. The starting point for every financial model is the uncertainty facing investors, and the substance of every financial model involves the impact of uncertainty on the behavior of investors and, ultimately, on market prices. Indeed, in the absence of uncertainty, the problems of financial economics reduce to exercises in basic microeconomics. The very existence of financial economics as a discipline is predicated on uncertainty.

This has important consequences for financial econometrics. The random fluctuations that require the use of statistical theory to estimate and test financial models are intimately related to the uncertainty on which those models are based. For example, the martingale model for asset prices has very specific implications for the behavior of test statistics such as the autocorrelation coefficient of price increments (see Chapter 2). This close connection between theory and empirical analysis is unparalleled in the

[1] Bernstein (1992) provides a highly readable account of the interplay between theory and practice in the development of modern financial economics.

social sciences, although it has been the hallmark of the natural sciences for quite some time. It is one of the most rewarding aspects of financial econometrics, so much so that we felt impelled to write this graduate-level textbook as a means of introducing others to this exciting field.

Section 1.1 explains which topics we cover in this book, and how we have organized the material. We also suggest some ways in which the book might be used in a one-semester course on financial econometrics or empirical finance.

In Section 1.2, we describe the kinds of background material that are most useful for financial econometrics and suggest references for those readers who wish to review or learn such material along the way. In our experience, students are often more highly motivated to pick up the necessary background *after* they see how it is to be applied, so we encourage readers with a serious interest in financial econometrics but with somewhat less preparation to take a crack at this material anyway.

In a book of this magnitude, notation becomes a nontrivial challenge of coordination; hence Section 1.3 describes what method there is in our notational madness. We urge readers to review this carefully to minimize the confusion that can arise when $\hat{\beta}$ is mistaken for β and X is incorrectly assumed to be the same as $\mathbf{X}$.

Section 1.4 extends our discussion of notation by presenting notational conventions for and definitions of some of the fundamental objects of our study: prices, returns, methods of compounding, and probability distributions. Although much of this material is well-known to finance students and investment professionals, we think a brief review will help many readers.

In Section 1.5, we turn our attention to quite a different subject: the Efficient Markets Hypothesis. Because so much attention has been lavished on this hypothesis, often at the expense of other more substantive issues, we wish to dispense with this issue first. Much of the debate involves theological tenets that are empirically undecidable and, therefore, beyond the purview of this text. But for completeness—no self-respecting finance text could omit market efficiency altogether—Section 1.5 briefly discusses the topic.

1.1 Organization of the Book

In organizing this book, we have followed two general principles. First, the early chapters concentrate exclusively on stock markets. Although many of the methods discussed can be applied equally well to other asset markets, the empirical literature on stock markets is particularly large and by focusing on these markets we are able to keep the discussion concrete. In later chapters, we cover derivative securities (Chapters 9 and 12) and fixed-income securities (Chapters 10 and 11). The last chapter of the book presents nonlinear methods, with applications to both stocks and derivatives.

Second, we start by presenting statistical models of asset returns, and then discuss more highly structured economic models. In Chapter 2, for example, we discuss methods for predicting stock returns from their own past history, without much attention to institutional detail; in Chapter 3 we show how the microstructure of stock markets affects the short-run behavior of returns. Similarly, in Chapter 4 we discuss simple statistical models of the cross-section of individual stock returns, and the application of these models to event studies; in Chapters 5 and 6 we show how the Capital Asset Pricing Model and multifactor models such as the Arbitrage Pricing Theory restrict the parameters of the statistical models. In Chapter 7 we discuss longer-run evidence on the predictability of stock returns from variables other than past stock returns; in Chapter 8 we explore dynamic equilibrium models which can generate persistent time-variation in expected returns. We use the same principle to divide a basic treatment of fixed-income securities in Chapter 10 from a discussion of equilibrium term-structure models in Chapter 11.

We have tried to make each chapter as self-contained as possible. While some chapters naturally go together (e.g., Chapters 5 and 6, and Chapters 10 and 11), there is certainly no need to read this book straight through from beginning to end. For classroom use, most teachers will find that there is too much material here to be covered in one semester. There are several ways to use the book in a one-semester course. For example one teacher might start by discussing short-run time-series behavior of stock prices using Chapters 2 and 3, then cover cross-sectional models in Chapters 4, 5, and 6, then discuss intertemporal equilibrium models using Chapter 8, and finally cover derivative securities and nonlinear methods as advanced topics using Chapters 9 and 12. Another teacher might first present the evidence on short- and long-run predictability of stock returns using Chapters 2 and 7, then discuss static and intertemporal equilibrium theory using Chapters 5, 6, and 8, and finally cover fixed-income securities using Chapters 10 and 11.

There are some important topics that we have not been able to include in this text. Most obviously, our focus is almost exclusively on US domestic asset markets. We say very little about asset markets in other countries, and we do not try to cover international topics such as exchange-rate behavior or the home-bias puzzle (the tendency for each country's investors to hold a disproportionate share of their own country's assets in their portfolios). We also omit such important econometric subjects as Bayesian analysis and frequency-domain methods of time-series analysis. In many cases our choice of topics has been influenced by the dual objectives of the book: to explain the methods of financial econometrics, and to review the empirical literature in finance. We have tended to concentrate on topics that involve econometric issues, sometimes at the expense of other equally interesting material—including much recent work in behavioral finance—that is econometrically more straightforward.

1.2 Useful Background

The many rewards of financial econometrics come at a price. A solid background in mathematics, probability and statistics, and finance theory is necessary for the practicing financial econometrician, for precisely the reasons that make financial econometrics such an engaging endeavor. To assist readers in obtaining this background (since only the most focused and directed of students will have it already), we outline in this section the topics in mathematics, probability, statistics, and finance theory that have become indispensable to financial econometrics. We hope that this outline can serve as a self-study guide for the more enterprising readers and that it will be a partial substitute for including background material in this book.

1.2.1 Mathematics Background

The mathematics background most useful for financial econometrics is not unlike the background necessary for econometrics in general: multivariate calculus, linear algebra, and matrix analysis. References for each of these topics are Lang (1973), Strang (1976), and Magnus and Neudecker (1988), respectively. Key concepts include

- multiple integration
- multivariate constrained optimization
- matrix algebra
- basic rules of matrix differentiation.

In addition, option- and other derivative-pricing models, and continuous-time asset pricing models, require some passing familiarity with the *Itô* or *stochastic calculus*. A lucid and thorough treatment is provided by Merton (1990), who pioneered the application of stochastic calculus to financial economics. More mathematically inclined readers may also wish to consult Chung and Williams (1990).

1.2.2 Probability and Statistics Background

Basic probability theory is a prerequisite for any discipline in which uncertainty is involved. Although probability theory has varying degrees of mathematical sophistication, from coin-flipping calculations to measure-theoretic foundations, perhaps the most useful approach is one that emphasizes the intuition and subtleties of elementary probabilistic reasoning. An amazingly durable classic that takes just this approach is Feller (1968). Brieman (1992) provides similar intuition but at a measure-theoretic level. Key concepts include

- definition of a random variable
- independence
- distribution and density functions
- conditional probability
- modes of convergence
- laws of large numbers
- central limit theorems.

Statistics is, of course, the primary engine which drives the inferences that financial econometricians draw from the data. As with probability theory, statistics can be taught at various levels of mathematical sophistication. Moreover, unlike the narrower (and some would say "purer") focus of probability theory, statistics has increased its breadth as it has matured, giving birth to many well-defined subdisciplines such as multivariate analysis, nonparametrics, time-series analysis, order statistics, analysis of variance, decision theory, Bayesian statistics, etc. Each of these subdisciplines has been drawn upon by financial econometricians at one time or another, making it rather difficult to provide a single reference for all of these topics. Amazingly, such a reference does exist: Stuart and Ord's (1987) three-volume *tour de force*. A more compact reference that contains most of the relevant material for our purposes is the elegant monograph by Silvey (1975). For topics in time-series analysis, Hamilton (1994) is an excellent comprehensive text. Key concepts include

- Neyman-Pearson hypothesis testing
- linear regression
- maximum likelihood
- basic time-series analysis (stationarity, autoregressive and ARMA processes, vector autoregressions, unit roots, etc.)
- elementary Bayesian inference.

For continuous-time financial models, an additional dose of stochastic processes is a must, at least at the level of Cox and Miller (1965) and Hoel, Port, and Stone (1972).

1.2.3 Finance Theory Background

Since the *raison d'être* of financial econometrics is the empirical implementation and evaluation of financial models, a solid background in finance theory is the most important of all. Several texts provide excellent coverage of this material: Duffie (1992), Huang and Litzenberger (1988), Ingersoll (1987), and Merton (1990). Key concepts include

- risk aversion and expected-utility theory
- static mean-variance portfolio theory
- the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT)
- dynamic asset pricing models
- option pricing theory.

1.3 Notation

We have found that it is far from simple to devise a consistent notational scheme for a book of this scope. The difficulty comes from the fact that financial econometrics spans several very different strands of the finance literature, each replete with its own firmly established set of notational conventions. But the conventions in one literature often conflict with the conventions in another. Unavoidably, then, we must sacrifice either internal notational consistency across different chapters of this text or external consistency with the notation used in the professional literature. We have chosen the former as the lesser evil, but we do maintain the following conventions throughout the book:

- We use boldface for vectors and matrices, and regular face for scalars. Where possible, we use bold uppercase for matrices and bold lowercase for vectors. Thus $\mathbf{x}$ is a vector while $\mathbf{X}$ is a matrix.
- Where possible, we use uppercase letters for the levels of variables and lowercase letters for the natural logarithms (logs) of the same variables. Thus if P is an asset price, p is the log asset price.
- Our standard notation for an innovation is the Greek letter ϵ. Where we need to define several different innovations, we use the alternative Greek letters η, ξ, and ζ.
- Where possible, we use Greek letters to denote parameters or parameter vectors.
- We use the Greek letter ι to denote a vector of ones.
- We use hats to denote sample estimates, so if β is a parameter, $\hat{\beta}$ is an estimate of β.
- When we use subscripts, we always use uppercase letters for the upper limits of the subscripts. Where possible, we use the same letters for upper limits as for the subscripts themselves. Thus subscript t runs from 1 to T, subscript k runs from 1 to K, and so on. An exception is that we will let subscript i (usually denoting an asset) run from 1 to N because this notation is so common. We use t and τ for time subscripts;

i for asset subscripts; k, m, and n for lead and lag subscripts; and j as a generic subscript.
- We use the timing convention that a variable is dated t if it is known by the end of period t. Thus R_t denotes a return on an asset held from the end of period $t-1$ to the end of period t.
- In writing variance-covariance matrices, we use $\mathbf{\Omega}$ for the variance-covariance matrix of asset returns, $\mathbf{\Sigma}$ for the variance-covariance matrix of residuals from a time-series or cross-sectional model, and $\mathbf{V}$ for the variance-covariance matrix of parameter estimators.
- We use script letters sparingly. $\mathcal{N}$ denotes the normal distribution, and $\mathcal{L}$ denotes a log likelihood function.
- We use $\Pr(\cdot)$ to denote the probability of an event.

The professional literature uses many specialized terms. Inevitably we also use these frequently, and we italicize them when they first appear in the book.

1.4 Prices, Returns, and Compounding

Virtually every aspect of financial economics involves *returns*, and there are at least two reasons for focusing our attention on returns rather than on prices. First, for the average investor, financial markets may be considered close to perfectly competitive, so that the size of the investment does not affect price changes. Therefore, since the investment "technology" is constant-returns-to-scale, the return is a complete and scale-free summary of the investment opportunity.

Second, for theoretical and empirical reasons that will become apparent below, returns have more attractive statistical properties than prices, such as stationarity and ergodicity. In particular, dynamic general-equilibrium models often yield nonstationary prices, but stationary returns (see, for example, Chapter 8 and Lucas [1978]).

1.4.1 Definitions and Conventions

Denote by P_t the price of an asset at date t and assume for now that this asset pays no dividends. The *simple net return*, R_t, on the asset between dates $t-1$ and t is defined as

$$R_t = \frac{P_t}{P_{t-1}} - 1. \tag{1.4.1}$$

The *simple gross return* on the asset is just one plus the net return, $1 + R_t$.

From this definition it is apparent that the asset's gross return over the most recent k periods from date $t-k$ to date t, written $1 + R_t(k)$, is simply equal to the product of the k single-period returns from $t-k+1$ to t, i.e.,

$$\begin{aligned} 1+R_t(k) &\equiv (1+R_t)\cdot(1+R_{t-1})\cdots(1+R_{t-k+1}) \\ &= \frac{P_t}{P_{t-1}}\cdot\frac{P_{t-1}}{P_{t-2}}\cdot\frac{P_{t-2}}{P_{t-3}}\cdots\frac{P_{t-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}}, \end{aligned} \tag{1.4.2}$$

and its net return over the most recent k periods, written $R_t(k)$, is simply equal to its k-period gross return minus one. These multiperiod returns are called *compound* returns.

Although returns are scale-free, it should be emphasized that they are *not* unitless, but are always defined with respect to some time interval, e.g., one "period." In fact, R_t is more properly called a *rate* of return, which is more cumbersome terminology but more accurate in referring to R_t as a rate or, in economic jargon, a *flow* variable. Therefore, a return of 20% is not a complete description of the investment opportunity without specification of the return horizon. In the academic literature, the return horizon is generally given explicitly, often as part of the data description, e.g., "The CRSP *monthly* returns file was used."

However, among practitioners and in the financial press, a return-horizon of one year is usually assumed implicitly; hence, unless stated otherwise, a return of 20% is generally taken to mean an *annual* return of 20%. Moreover, multiyear returns are often *annualized* to make investments with different horizons comparable, thus:

$$\text{Annualized}[R_t(k)] = \left[\prod_{j=0}^{k-1}(1+R_{t-j})\right]^{1/k} - 1. \tag{1.4.3}$$

Since single-period returns are generally small in magnitude, the following approximation based on a first-order Taylor expansion is often used to annualize multiyear returns:

$$\text{Annualized}[R_t(k)] \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}. \tag{1.4.4}$$

Whether such an approximation is adequate depends on the particular application at hand; it may suffice for a quick and coarse comparison of investment performance across many assets, but for finer calculations in which the volatility of returns plays an important role, i.e., when the higher-order terms in the Taylor expansion are not negligible, the approximation (1.4.4) may break down. The only advantage of such an approximation is convenience—it is easier to calculate an arithmetic rather than a geometric average—however, this advantage has diminished considerably with the advent of cheap and convenient computing power.

Continuous Compounding

The difficulty of manipulating geometric averages such as (1.4.3) motivates another approach to compound returns, one which is not approximate and also has important implications for modeling asset returns; this is the notion of continuous compounding. The *continuously compounded return* or *log return* r_t of an asset is defined to be the natural logarithm of its gross return $(1+R_t)$:

$$r_t \equiv \log(1+R_t) = \log\frac{P_t}{P_{t-1}} = p_t - p_{t-1}, \tag{1.4.5}$$

where $p_t \equiv \log P_t$. When we wish to emphasize the distinction between R_t and r_t, we shall refer to R_t as a *simple* return. Our notation here deviates slightly from our convention that lowercase letters denote the logs of uppercase letters, since here we have $r_t = \log(1+R_t)$ rather than $\log(R_t)$; we do this to maintain consistency with standard conventions.

The advantages of continuously compounded returns become clear when we consider multiperiod returns, since

$$\begin{aligned} r_t(k) &= \log(1+R_t(k)) = \log((1+R_t)\cdot(1+R_{t-1})\cdots(1+R_{t-k+1})) \\ &= \log(1+R_t) + \log(1+R_{t-1}) + \cdots + \log(1+R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1}, \end{aligned} \tag{1.4.6}$$

and hence the continuously compounded multiperiod return is simply the sum of continuously compounded single-period returns. Compounding, a multiplicative operation, is converted to an additive operation by taking logarithms. However, the simplification is not merely in reducing multiplication to addition (since we argued above that with modern calculators and computers, this is trivial), but more in the modeling of the statistical behavior of asset returns over time—it is far easier to derive the time-series properties of additive processes than of multiplicative processes, as we shall see in Chapter 2.

Continuously compounded returns do have one disadvantage. The simple return on a portfolio of assets is a weighted average of the simple returns on the assets themselves, where the weight on each asset is the share of the portfolio's value invested in that asset. If portfolio p places weight w_{ip} in asset i, then the return on the portfolio at time t, R_{pt}, is related to the returns on individual assets, R_{it}, $i = 1\ldots N$, by $R_{pt} = \sum_{i=1}^{N} w_{ip}R_{it}$. Unfortunately continuously compounded returns do not share this convenient property. Since the log of a sum is not the same as the sum of logs, r_{pt} does not equal $\sum_{i=1}^{N} w_{ip}r_{it}$.

In empirical applications this problem is usually minor. When returns are measured over short intervals of time, and are therefore close to zero, the continuously compounded return on a portfolio is close to the weighted

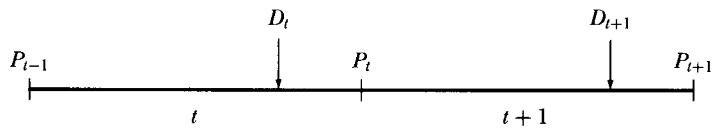

Figure 1.1. *Dividend Payment Timing Convention*

average of the continuously compounded returns on the individual assets: $r_{pt} \approx \sum_{i=1}^{N} w_{ip} r_{it}$.[2] We use this approximation in Chapter 3. Nonetheless it is common to use simple returns when a cross-section of assets is being studied, as in Chapters 4–6, and continuously compounded returns when the temporal behavior of returns is the focus of interest, as in Chapters 2 and 7.

Dividend Payments

For assets which make periodic dividend payments, we must modify our definitions of returns and compounding. Denote by D_t the asset's dividend payment at date t and assume, purely as a matter of convention, that this dividend is paid just before the date-t price P_t is recorded; hence P_t is taken to be the *ex-dividend* price at date t. Alternatively, one might describe P_t as an end-of-period asset price, as shown in Figure 1.1. Then the net simple return at date t may be defined as

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1. \tag{1.4.7}$$

Multiperiod and continuously compounded returns may be obtained in the same way as in the no-dividends case. Note that the continuously compounded return on a dividend-paying asset, $r_t = \log(P_t + D_t) - \log(P_{t-1})$, is a nonlinear function of log prices and log dividends. When the ratio of prices to dividends is not too variable, however, this function can be approximated by a linear function of log prices and dividends, as discussed in detail in Chapter 7.

Excess Returns

It is often convenient to work with an asset's excess return, defined as the difference between the asset's return and the return on some reference asset. The reference asset is often assumed to be riskless and in practice is usually a short-term Treasury bill return. Working with simple returns, the

[2] In the limit where time is continuous, Ito's Lemma, discussed in Section 9.1.2 of Chapter 9, can be used to relate simple and continuously compounded returns.

simple excess return on asset i is

$$Z_{it} = R_{it} - R_{0t}, \tag{1.4.8}$$

where R_{0t} is the reference return. Alternatively one can define a log excess return as

$$z_{it} = r_{it} - r_{0t}. \tag{1.4.9}$$

The excess return can also be thought of as the payoff on an *arbitrage portfolio* that goes long in asset i and short in the reference asset, with no net investment at the initial date. Since the initial net investment is zero, the return on the arbitrage portfolio is undefined but its dollar payoff is proportional to the excess return as defined above.

1.4.2 The Marginal, Conditional, and Joint Distribution of Returns

Having defined asset returns carefully, we can now begin to study their behavior across assets and over time. Perhaps the most important characteristic of asset returns is their randomness. The return of IBM stock over the next month is unknown today, and it is largely the explicit modeling of the sources and nature of this uncertainty that distinguishes financial economics from other social sciences. Although other branches of economics and sociology do have models of stochastic phenomena, in none of them does uncertainty play so central a role as in the pricing of financial assets—without uncertainty, much of the financial economics literature, both theoretical and empirical, would be superfluous. Therefore, we must articulate at the very start the types of uncertainty that asset returns might exhibit.

The Joint Distribution

Consider a collection of N assets at date t, each with return R_{it} at date t, where $t = 1, \ldots, T$. Perhaps the most general model of the collection of returns $\{R_{it}\}$ is its joint distribution function:

$$G(R_{11}, \ldots, R_{N1}; R_{12}, \ldots, R_{N2}; \ldots; R_{1T}, \ldots, R_{NT}; \mathbf{x} \mid \boldsymbol{\theta}), \tag{1.4.10}$$

where $\mathbf{x}$ is a vector of *state variables*, variables that summarize the economic environment in which asset returns are determined, and $\boldsymbol{\theta}$ is a vector of fixed parameters that uniquely determines G. For notational convenience, we shall suppress the dependence of G on the parameters $\boldsymbol{\theta}$ unless it is needed.

The probability law G governs the stochastic behavior of asset returns and $\mathbf{x}$, and represents the sum total of all knowable information about them. We may then view financial econometrics as the statistical inference of $\boldsymbol{\theta}$, given G and realizations of $\{R_{it}\}$. Of course, (1.4.10) is far too general to be of any use for statistical inference, and we shall have to place further restrictions on G in the coming sections and chapters. However, (1.4.10) does serve as a convenient way to organize the many models of asset returns to be developed here and in later chapters. For example, Chapters 2 through 6 deal exclusively with the joint distribution of $\{R_{it}\}$, leaving additional state variables $\mathbf{x}$ to be considered in Chapters 7 and 8. We write this joint distribution as G_R.

Many asset pricing models, such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965a, b), and Mossin (1966) considered in Chapter 5, describe the joint distribution of the cross section of returns $\{R_{1t}, \ldots, R_{Nt}\}$ at a single date t. To reduce (1.4.10) to this essentially static structure, we shall have to assert that returns are statistically independent through time and that the joint distribution of the cross-section of returns is identical across time. Although such assumptions seem extreme, they yield a rich set of implications for pricing financial assets. The CAPM, for example, delivers an explicit formula for the trade-off between risk and expected return, the celebrated security market line.

The Conditional Distribution

In Chapter 2, we place another set of restrictions on G_R which will allow us to focus on the *dynamics* of individual asset returns while abstracting from cross-sectional relations between the assets. In particular, consider the joint distribution F of $\{R_{i1}, \ldots, R_{iT}\}$ for a given asset i, and observe that we may always rewrite F as the following product:

$$\begin{aligned} F(R_{i1}, \ldots, R_{iT}) &= F_{i1}(R_{i1}) \cdot F_{i2}(R_{i2} \mid R_{i1}) \cdot F_{i3}(R_{i3} \mid R_{i2}, R_{i1}) \\ &\quad \cdots F_{iT}(R_{iT} \mid R_{iT-1}, \ldots, R_{i1}). \end{aligned} \tag{1.4.11}$$

From (1.4.11), the temporal dependencies implicit in $\{R_{it}\}$ are apparent. Issues of predictability in asset returns involve aspects of their *conditional* distributions and, in particular, how the conditional distributions evolve through time.

By placing further restrictions on the conditional distributions $F_{it}(\cdot)$, we shall be able to estimate the parameters $\boldsymbol{\theta}$ implicit in (1.4.11) and examine the predictability of asset returns explicitly. For example, one version of the random-walk hypothesis is obtained by the restriction that the conditional distribution of return R_{it} is equal to its marginal distribution, i.e., $F_{it}(R_{it} \mid \cdot) = F_{it}(R_{it})$. If this is the case, then returns are temporally independent and therefore unpredictable using past returns. Weaker versions of the random walk are obtained by imposing weaker restrictions on $F_{it}(R_{it} \mid \cdot)$.

The Unconditional Distribution

In cases where an asset return's conditional distribution differs from its marginal or unconditional distribution, it is clearly the conditional distribution that is relevant for issues involving predictability. However, the properties of the unconditional distribution of returns may still be of some interest, especially in cases where we expect predictability to be minimal.

One of the most common models for asset returns is the temporally independently and identically distributed (IID) normal model, in which returns are assumed to be independent over time (although perhaps cross-sectionally correlated), identically distributed over time, and normally distributed. The original formulation of the CAPM employed this assumption of normality, although returns were only implicitly assumed to be temporally IID (since it was a static "two-period" model). More recently, models of asymmetric information such as Grossman (1989) and Grossman and Stiglitz (1980) also use normality.

While the temporally IID normal model may be tractable, it suffers from at least two important drawbacks. First, most financial assets exhibit limited liability, so that the largest loss an investor can realize is his total investment and no more. This implies that the smallest net return achievable is -1 or -100%. But since the normal distribution's support is the entire real line, this lower bound of -1 is clearly violated by normality. Of course, it may be argued that by choosing the mean and variance appropriately, the probability of realizations below -1 can be made arbitrarily small; however it will never be zero, as limited liability requires.

Second, if single-period returns are assumed to be normal, then multiperiod returns cannot also be normal since they are the *products* of the single-period returns. Now the *sums* of normal single-period returns are indeed normal, but the sum of single-period simple returns does not have any economically meaningful interpretation. However, as we saw in Section 1.4.1, the sum of single-period continuously compounded returns does have a meaningful interpretation as a multiperiod continuously compounded return.

The Lognormal Distribution

A sensible alternative is to assume that continuously compounded single-period returns r_{it} are IID normal, which implies that single-period gross simple returns are distributed as IID *lognormal* variates, since $r_{it} \equiv \log(1 + R_{it})$. We may express the lognormal model then as

$$r_{it} \sim \mathcal{N}(\mu_i, \sigma_i^2). \qquad (1.4.12)$$

Under the lognormal model, if the mean and variance of r_{it} are μ_i and σ_i^2, respectively, then the mean and variance of simple returns are given by

$$\mathrm{E}[R_{it}] = e^{\mu_i + \frac{\sigma_i^2}{2}} - 1 \qquad (1.4.13)$$

$$\mathrm{Var}[R_{it}] = e^{2\mu_i + \sigma_i^2}[e^{\sigma_i^2} - 1]. \qquad (1.4.14)$$

Alternatively, if we assume that the mean and variance of simple returns R_{it} are m_i and s_i^2, respectively, then under the lognormal model the mean and variance of r_{it} are given by

$$\mathrm{E}[r_{it}] = \log \frac{m_i + 1}{\sqrt{1 + \left(\frac{s_i}{m_i+1}\right)^2}} \tag{1.4.15}$$

$$\mathrm{Var}[r_{it}] = \log\left[1 + \left(\frac{s_i}{m_i + 1}\right)^2\right]. \tag{1.4.16}$$

The lognormal model has the added advantage of not violating limited liability, since limited liability yields a lower bound of zero on $(1 + R_{it})$, which is satisfied by $(1 + R_{it}) = e^{r_{it}}$ when r_{it} is assumed to be normal.

The lognormal model has a long and illustrious history, beginning with the dissertation of the French mathematician Louis Bachelier (1900), which contained the mathematics of Brownian motion and heat conduction, five years prior to Einstein's (1905) famous paper. For other reasons that will become apparent in later chapters (see, especially, Chapter 9), the lognormal model has become the workhorse of the financial asset pricing literature.

But as attractive as the lognormal model is, it is not consistent with all the properties of historical stock returns. At short horizons, historical returns show weak evidence of skewness and strong evidence of excess kurtosis. The *skewness*, or normalized third moment, of a random variable ϵ with mean μ and variance σ^2 is defined by

$$S[\epsilon] \equiv \mathrm{E}\left[\frac{(\epsilon - \mu)^3}{\sigma^3}\right]. \tag{1.4.17}$$

The *kurtosis*, or normalized fourth moment, of ϵ is defined by

$$K[\epsilon] \equiv \mathrm{E}\left[\frac{(\epsilon - \mu)^4}{\sigma^4}\right]. \tag{1.4.18}$$

The normal distribution has skewness equal to zero, as do all other symmetric distributions. The normal distribution has kurtosis equal to 3, but *fat-tailed* distributions with extra probability mass in the tail areas have higher or even infinite kurtosis.

Skewness and kurtosis can be estimated in a sample of data by constructing the obvious sample averages: the sample mean

$$\hat{\mu} \equiv \frac{1}{T}\sum_{t=1}^{T} \epsilon_t, \tag{1.4.19}$$

the sample variance

$$\hat{\sigma}^2 \equiv \frac{1}{T}\sum_{t=1}^{T}(\epsilon_t - \hat{\mu})^2, \tag{1.4.20}$$

the sample skewness

$$\hat{S} \equiv \frac{1}{T\hat{\sigma}^3}\sum_{t=1}^{T}(\epsilon_t - \hat{\mu})^3, \tag{1.4.21}$$

and the sample kurtosis

$$\hat{K} \equiv \frac{1}{T\hat{\sigma}^4}\sum_{t=1}^{T}(\epsilon_t - \hat{\mu})^4. \tag{1.4.22}$$

In large samples of normally distributed data, the estimators $\hat{S}$ and $\hat{K}$ are normally distributed with means 0 and 3 and variances $6/T$ and $24/T$, respectively (see Stuart and Ord [1987, Vol. 1]). Since 3 is the kurtosis of the normal distribution, sample *excess kurtosis* is defined to be sample kurtosis less 3. Sample estimates of skewness for daily US stock returns tend to be negative for stock indexes but close to zero or positive for individual stocks. Sample estimates of excess kurtosis for daily US stock returns are large and positive for both indexes and individual stocks, indicating that returns have more mass in the tail areas than would be predicted by a normal distribution.

Stable Distributions

Early studies of stock market returns attempted to capture this excess kurtosis by modeling the distribution of continuously compounded returns as a member of the *stable* class (also called the *stable Pareto-Lévy* or *stable Paretian*), of which the normal is a special case.[3] The stable distributions are a natural generalization of the normal in that, as their name suggests, they are stable under addition, i.e., a sum of stable random variables is also a stable random variable. However, nonnormal stable distributions have more probability mass in the tail areas than the normal. In fact, the nonnormal stable distributions are so fat-tailed that their variance and all higher moments are infinite. Sample estimates of variance or kurtosis for random variables with

[3]The French probabilist Paul Lévy (1924) was perhaps the first to initiate a general investigation of stable distributions and provided a complete characterization of them through their log-characteristic functions (see below). Lévy (1925) also showed that the tail probabilities of stable distributions approximate those of the Pareto distribution, hence the term "stable Pareto-Lévy" or "stable Paretian" distribution. For applications to financial asset returns, see Blattberg and Gonedes (1974); Fama (1965); Fama and Roll (1971); Fielitz (1976); Fielitz and Rozell (1983); Granger and Morgenstern (1970); Hagerman (1978); Hsu, Miller, and Wichern (1974); Mandelbrot (1963); Mandelbrot and Taylor (1967); Officer (1972); Samuelson (1967, 1976); Simkowitz and Beedles (1980); and Tucker (1992).

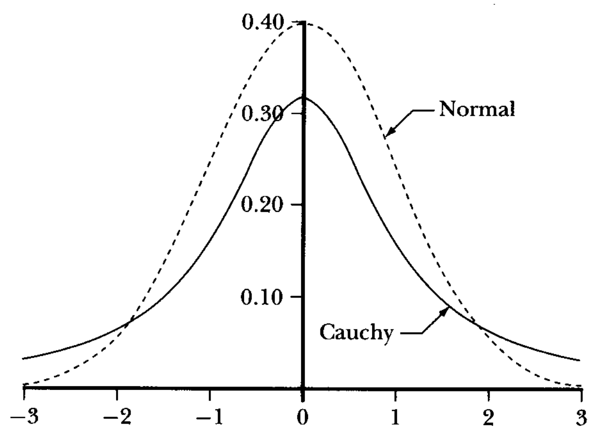

Figure 1.2. *Comparison of Stable and Normal Density Functions*

these distributions will not converge as the sample size increases, but will tend to increase indefinitely.

Closed-form expressions for the density functions of stable random variables are available for only three special cases: the normal, the Cauchy, and the Bernoulli cases.[4] Figure 1.2 illustrates the Cauchy distribution, with density function

$$f(x) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (x-\delta)^2}\,. \tag{1.4.23}$$

In Figure 1.2, (1.4.23) is graphed with parameters $\delta = 0$ and $\gamma = 1$, and it is apparent from the comparison with the normal density function (dashed lines) that the Cauchy has fatter tails than the normal.

Although stable distributions were popular in the 1960's and early 1970's, they are less commonly used today. They have fallen out of favor partly because they make theoretical modelling so difficult; standard finance theory

[4]However, Lévy (1925) derived the following explicit expression for the logarithm of the characteristic function $\varphi(t)$ of any stable random variable X: $\log\varphi(t) \equiv \log \mathrm{E}[e^{itX}] = i\delta t - \gamma|t|^{\alpha}[1 - i\beta\,\mathrm{sgn}(t)\tan(\alpha\pi/2)]$, where $(\alpha, \beta, \delta, \gamma)$ are the four parameters that characterize each stable distribution. $\delta \in (-\infty, \infty)$ is said to be the *location* parameter, $\beta \in (-\infty, \infty)$ is the *skewness index*, $\gamma \in (0, \infty)$ is the *scale* parameter, and $\alpha \in (0, 2]$ is the *exponent*. When $\alpha = 2$, the stable distribution reduces to a normal. As α decreases from 2 to 0, the tail areas of the stable distribution become increasingly "fatter" than the normal. When $\alpha \in (1, 2)$, the stable distribution has a finite mean given by δ, but when $\alpha \in (0, 1]$, even the mean is infinite. The parameter β measures the symmetry of the stable distribution; when $\beta = 0$ the distribution is symmetric, and when $\beta > 0$ (or $\beta < 0$) the distribution is skewed to the right (or left). When $\beta = 0$ and $\alpha = 1$ we have the Cauchy distribution, and when $\alpha = 1/2$, $\beta = 1$, $\delta = 0$, and $\gamma = 1$ we have the Bernoulli distribution.

almost always requires finite second moments of returns, and often finite higher moments as well. Stable distributions also have some counterfactual implications. First, they imply that sample estimates of the variance and higher moments of returns will tend to increase as the sample size increases, whereas in practice these estimates seem to converge. Second, they imply that long-horizon returns will be just as non-normal as short-horizon returns (since long-horizon returns are sums of short-horizon returns, and these distributions are stable under addition). In practice the evidence for non-normality is much weaker for long-horizon returns than for short-horizon returns.

Recent research tends instead to model returns as drawn from a fat-tailed distribution with finite higher moments, such as the t distribution, or as drawn from a mixture of distributions. For example the return might be conditionally normal, conditional on a variance parameter which is itself random; then the unconditional distribution of returns is a mixture of normal distributions, some with small conditional variances that concentrate mass around the mean and others with large conditional variances that put mass in the tails of the distribution. The result is a fat-tailed unconditional distribution with a finite variance and finite higher moments. Since all moments are finite, the Central Limit Theorem applies and long-horizon returns will tend to be closer to the normal distribution than short-horizon returns. It is natural to model the conditional variance as a time-series process, and we discuss this in detail in Chapter 12.

An Empirical Illustration

Table 1.1 contains some sample statistics for individual and aggregate stock returns from the Center for Research in Securities Prices (CRSP) for 1962 to 1994 which illustrate some of the issues discussed in the previous sections. Sample moments, calculated in the straightforward way described in (1.4.19)–(1.4.22), are reported for value- and equal-weighted indexes of stocks listed on the New York Stock Exchange (NYSE) and American Stock Exchange (AMEX), and for ten individual stocks. The individual stocks were selected from market-capitalization deciles using 1979 end-of-year market capitalizations for all stocks in the CRSP NYSE/AMEX universe, where International Business Machines is the largest decile's representative and Continental Materials Corp. is the smallest decile's representative.

Panel A reports statistics for daily returns. The daily index returns have extremely high sample excess kurtosis, 34.9 and 26.0 respectively, a clear sign of fat tails. Although the excess kurtosis estimates for daily individual stock returns are generally less than those for the indexes, they are still large, ranging from 3.35 to 59.4. Since there are 8179 observations, the standard error for the kurtosis estimate under the null hypothesis of normality is $\sqrt{24/8179} = 0.054$, so these estimates of excess kurtosis are overwhelmingly statistically significant. The skewness estimates are negative for the daily index returns, -1.33 and -0.93 respectively, but generally positive for the individual stock returns, ranging from -0.18 to 2.25. Many of the skewness estimates are also statistically significant as the standard error under the null hypothesis of normality is $\sqrt{6/8179} = 0.027$.

Panel B reports sample statistics for monthly returns. These are considerably less leptokurtic than daily returns—the value- and equal-weighted CRSP monthly index returns have excess kurtosis of only 2.42 and 4.14, respectively, an order of magnitude smaller than the excess kurtosis of daily returns. As there are only 390 observations the standard error for the kurtosis estimate is also much larger, 0.248. This is one piece of evidence that has led researchers to use fat-tailed distributions with finite higher moments, for which the Central Limit Theorem applies and drives longer-horizon returns towards normality.

1.5 Market Efficiency

The origins of the Efficient Markets Hypothesis (EMH) can be traced back at least as far as the pioneering theoretical contribution of Bachelier (1900) and the empirical research of Cowles (1933). The modern literature in economics begins with Samuelson (1965), whose contribution is neatly summarized by the title of his article: "Proof that Properly Anticipated Prices Fluctuate Randomly".[5] In an informationally efficient market—not to be confused with an allocationally or Pareto-efficient market—price changes must be unforecastable if they are properly anticipated, i.e., if they fully incorporate the expectations and information of all market participants.

Fama (1970) summarizes this idea in his classic survey by writing: "A market in which prices always 'fully reflect' available information is called 'efficient'." Fama's use of quotation marks around the words "fully reflect" indicates that these words are a form of shorthand and need to be explained more fully. More recently, Malkiel (1992) has offered the following more explicit definition:

> A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set . . . if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set

[5]Bernstein (1992) discusses the contributions of Bachelier, Cowles, Samuelson, and many other early authors. The articles reprinted in Lo (1996) include some of the most important papers in this literature.

Table 1.1. *Stock market returns, 1962 to 1994.*

Security	Mean	Standard Deviation	Skewness	Excess Kurtosis	Minimum	Maximum
		Panel A: Daily Returns				
Value-Weighted Index	0.044	0.82	−1.33	34.92	−18.10	8.87
Equal-Weighted Index	0.073	0.76	−0.93	26.03	−14.19	9.83
International Business Machines	0.039	1.42	−0.18	12.48	−22.96	11.72
General Signal Corp.	0.054	1.66	0.01	3.35	−13.46	9.43
Wrigley Co.	0.072	1.45	−0.00	11.03	−18.67	11.89
Interlake Corp.	0.043	2.16	0.72	12.35	−17.24	23.08
Raytech Corp.	0.050	3.39	2.25	59.40	−57.90	75.00
Ampco-Pittsburgh Corp.	0.053	2.41	0.66	5.02	−19.05	19.18
Energen Corp.	0.054	1.41	0.27	5.91	−12.82	11.11
General Host Corp.	0.070	2.79	0.74	6.18	−23.53	22.92
Garan Inc.	0.079	2.35	0.72	7.13	−16.67	19.07
Continental Materials Corp.	0.143	5.24	0.93	6.49	−26.92	50.00
		Panel B: Monthly Returns				
Value-Weighted Index	0.96	4.33	−0.29	2.42	−21.81	16.51
Equal-Weighted Index	1.25	5.77	0.07	4.14	−26.80	33.17
International Business Machines	0.81	6.18	−0.14	0.83	−26.19	18.95
General Signal Corp.	1.17	8.19	−0.02	1.87	−36.77	29.73
Wrigley Co.	1.51	6.68	0.30	1.31	−20.26	29.72
Interlake Corp.	0.86	9.38	0.67	4.09	−30.28	54.84
Raytech Corp.	0.83	14.88	2.73	22.70	−45.65	142.11
Ampco-Pittsburgh Corp.	1.06	10.64	0.77	2.04	−36.08	46.94
Energen Corp.	1.10	5.75	1.47	12.47	−24.61	48.36
General Host Corp.	1.33	11.67	0.35	1.11	−38.05	42.86
Garan Inc.	1.64	11.30	0.76	2.30	−35.48	51.60
Continental Materials Corp.	1.64	17.76	1.13	3.33	−58.09	84.78

Summary statistics for daily and monthly returns (in percent) of CRSP equal- and value-weighted stock indexes and ten individual securities continuously listed over the entire sample period from July 3, 1962 to December 30, 1994. Individual securities are selected to represent stocks in each size decile. Statistics are defined in (1.4.19)–(1.4.22).

> . . . implies that it is impossible to make economic profits by trading on the basis of [that information set].

Malkiel's first sentence repeats Fama's definition. His second and third sentences expand the definition in two alternative ways. The second sentence suggests that market efficiency can be tested by revealing information to market participants and measuring the reaction of security prices. If prices do not move when information is revealed, then the market is efficient with respect to that information. Although this is clear conceptually, it is hard to carry out such a test in practice (except perhaps in a laboratory).

Malkiel's third sentence suggests an alternative way to judge the efficiency of a market, by measuring the profits that can be made by trading on information. This idea is the foundation of almost all the empirical work on market efficiency. It has been used in two main ways. First, many researchers have tried to measure the profits earned by market professionals such as mutual fund managers. If these managers achieve superior returns (after adjustment for risk) then the market is not efficient with respect to the information possessed by the managers. This approach has the advantage that it concentrates on real trading by real market participants, but it has the disadvantage that one cannot directly observe the information used by the managers in their trading strategies (see Fama [1970, 1991] for a thorough review of this literature).

As an alternative, one can ask whether hypothetical trading based on an explicitly specified information set would earn superior returns. To implement this approach, one must first choose an information set. The classic taxonomy of information sets, due to Roberts (1967), distinguishes among

Weak-form Efficiency: The information set includes only the history of prices or returns themselves.

Semistrong-Form Efficiency: The information set includes all information known to all market participants (*publicly available* information).

Strong-Form Efficiency: The information set includes all information known to any market participant (*private* information).

The next step is to specify a model of "normal" returns. Here the classic assumption is that the normal returns on a security are constant over time, but in recent years there has been increased interest in equilibrium models with time-varying normal security returns.

Finally, abnormal security returns are computed as the difference between the return on a security and its normal return, and forecasts of the abnormal returns are constructed using the chosen information set. If the abnormal security return is unforecastable, and in this sense "random," then the hypothesis of market efficiency is not rejected.

1.5.1 Efficient Markets and the Law of Iterated Expectations

The idea that efficient security returns should be random has often caused confusion. Many people seem to think that an efficient security price should be smooth rather than random. Black (1971) has attacked this idea rather effectively:

> A perfect market for a stock is one in which there are no profits to be made by people who have no special information about the company, and in which it is difficult even for people who do have special information to make profits, because the price adjusts so rapidly as the information becomes available.... Thus we would like to see *randomness* in the prices of successive transactions, rather than great continuity.... Randomness means that a series of small upward movements (or small downward movements) is very unlikely. If the price is going to move up, it should move up all at once, rather than in a series of small steps.... Large price movements are desirable, so long as they are not consistently followed by price movements in the opposite direction.

Underlying this confusion may be a belief that returns cannot be random if security prices are determined by discounting future cash flows. Smith (1968), for example, writes: "I suspect that even if the random walkers announced a perfect mathematic proof of randomness, I would go on believing that in the long run future earnings influence present value."

In fact, the discounted present-value model of a security price is entirely consistent with randomness in security returns. The key to understanding this is the so-called *Law of Iterated Expectations.* To state this result we define information sets I_t and J_t, where $I_t \subset J_t$ so all the information in I_t is also in J_t but J_t is superior because it contains some extra information. We consider expectations of a random variable X conditional on these information sets, written $\mathrm{E}[X \mid I_t]$ or $\mathrm{E}[X \mid J_t]$. The Law of Iterated Expectations says that $\mathrm{E}[X \mid I_t] = \mathrm{E}[\mathrm{E}[X \mid J_t] \mid I_t]$. In words, if one has limited information I_t, the best forecast one can make of a random variable X is the forecast of the forecast one would make of X if one had superior information J_t. This can be rewritten as $\mathrm{E}[X - \mathrm{E}[X \mid J_t] \mid I_t] = 0$, which has an intuitive interpretation: One cannot use limited information I_t to predict the forecast error one would make if one had superior information J_t.

Samuelson (1965) was the first to show the relevance of the Law of Iterated Expectations for security market analysis; LeRoy (1989) gives a lucid review of the argument. We discuss the point in detail in Chapter 7, but a brief summary may be helpful here. Suppose that a security price at time t, P_t, can be written as the rational expectation of some "fundamental value" V^*, conditional on information I_t available at time t. Then we have

$$P_t = \mathrm{E}[V^* \mid I_t] = \mathrm{E}_t V^*. \tag{1.5.1}$$

The same equation holds one period ahead, so

$$P_{t+1} = \mathrm{E}[V^* \mid I_{t+1}] = \mathrm{E}_{t+1} V^*. \tag{1.5.2}$$

But then the expectation of the change in the price over the next period is

$$\mathrm{E}_t[P_{t+1} - P_t] \;=\; \mathrm{E}_t[\mathrm{E}_{t+1}[V^*] - \mathrm{E}_t[V^*]] \;=\; 0, \tag{1.5.3}$$

because $I_t \subset I_{t+1}$, so $\mathrm{E}_t[\mathrm{E}_{t+1}[V^*]] = \mathrm{E}_t[V^*]$ by the Law of Iterated Expectations. Thus realized changes in prices are unforecastable given information in the set I_t.

1.5.2 Is Market Efficiency Testable?

Although the empirical methodology summarized here is well-established, there are some serious difficulties in interpreting its results. First, any test of efficiency must assume an equilibrium model that defines normal security returns. If efficiency is rejected, this could be because the market is truly inefficient or because an incorrect equilibrium model has been assumed. This *joint hypothesis* problem means that market efficiency as such can never be rejected.

Second, perfect efficiency is an unrealistic benchmark that is unlikely to hold in practice. Even in theory, as Grossman and Stiglitz (1980) have shown, abnormal returns will exist if there are costs of gathering and processing information. These returns are necessary to compensate investors for their information-gathering and information-processing expenses, and are no longer abnormal when these expenses are properly accounted for. In a large and liquid market, information costs are likely to justify only small abnormal returns, but it is difficult to say how small, even if such costs could be measured precisely.

The notion of *relative* efficiency—the efficiency of one market measured against another, e.g., the New York Stock Exchange vs. the Paris Bourse, futures markets vs. spot markets, or auction vs. dealer markets—may be a more useful concept than the all-or-nothing view taken by much of the traditional market-efficiency literature. The advantages of relative efficiency over absolute efficiency are easy to see by way of an analogy. Physical systems are often given an efficiency rating based on the relative proportion of energy or fuel converted to useful work. Therefore, a piston engine may be rated at 60% efficiency, meaning that on average 60% of the energy contained in the engine's fuel is used to turn the crankshaft, with the remaining 40% lost to other forms of work such as heat, light, or noise.

Few engineers would ever consider performing a statistical test to determine whether or not a given engine is perfectly efficient—such an engine exists only in the idealized frictionless world of the imagination. But measuring relative efficiency—relative to the frictionless ideal—is commonplace. Indeed, we have come to expect such measurements for many household products: air conditioners, hot water heaters, refrigerators, etc. Similarly, market efficiency is an idealization that is economically unrealizable, but that serves as a useful benchmark for measuring relative efficiency.

For these reasons, in this book we do not take a stand on market efficiency itself, but focus instead on the statistical methods that can be used to test the joint hypothesis of market efficiency and market equilibrium. Although many of the techniques covered in these pages are central to the market-efficiency debate—tests of variance bounds, Euler equations, the CAPM and the APT—we feel that they can be more profitably applied to measuring efficiency rather than to testing it. And if some markets turn out to be particularly inefficient, the diligent reader of this text will be well-prepared to take advantage of the opportunity.