CHAPTER 2

Market Risk Basics

In this chapter, we begin studying how to quantify *market risk*, the risk of loss from changes in market prices, via statistical models of the behavior of market prices. We introduce a set of tools and concepts to help readers better understand how returns, risk, and volatility are defined, and we review some basic concepts of portfolio allocation theory.

The statistical behavior of asset returns varies widely, across securities and over time:

- There is enormous variety among assets, as we saw in Chapter 1, ranging from simple cash securities, to fixed-income securities, which have a time dimension, to derivative securities, whose values are functions of other asset prices, to, finally, a bewildering array of indexes, basket products, tradeable fund shares, and structured credit products.
- The market risks of many, if not most, securities must be decomposed into underlying *risk factors*, which may or may not be directly observable. In addition to statistical models, therefore, we also need models and algorithms that relate security values to risk factors. We also need to accurately identify the important risk factors.

  We discuss the relationship between asset prices and risk factors further in Chapter 4. For now, we will be a bit loose and use the terms "asset prices" and "risk factors" synonymously.
- Some risk factors are far from intuitive. It is relatively straightforward to understand fluctuations in, say, equity prices. But options and option-like security returns are driven not only by price fluctuations, but also by the actual and anticipated behavior of volatility. Markets can even, as we see in Chapter 14, be affected by the volatility of volatility.

It is impossible to capture the variety of real-life security price fluctuations in a single, simple model. But the standard model we develop in this and the next chapter is useful as a starting point for quantitative modeling. It will also help us understand the issues and pitfalls involved in security price modeling. *Nobody* believes in the standard model literally, but it provides a framework for thinking about problems. If we study the standard model, and come away more conscious of its flaws than of its virtues, we will have accomplished a great deal. Finally, as we see in Chapter 15, the standard model informs to a large extent the regulatory approach to major financial institutions.

We want to understand the relationship between risk and security price behavior. Finance theory has developed a framework for this called *modern finance theory*. It has plagued researchers with far more questions than answers, but it helps organize thinking about the subject. Modern finance theory has its roots in economic theory, where much of the theoretical apparatus it employs, such as constrained expected utility maximization and the conditions for a set of prices and quantities to be mutually consistent and individually and socially optimal, were developed.

Choices are made through time, and the environment in which choices are made is subject to uncertainty. To characterize uncertainty, modern finance theory has made extensive use of tools, particularly certain classes of *stochastic processes*, that describe uncertain environments over time and permit tractable models to be developed.

2.1 ARITHMETIC, GEOMETRIC, AND LOGARITHMIC SECURITY RETURNS

We begin by laying out the algebra of returns. Returns on an asset can be defined two different ways, as a proportional change called the *arithmetic* or *simple rate of return*, or as a logarithmic change, called the *logarithmic rate of return*. Let's assume for now that the asset pays no cash flows between t and $t + \tau$. The two definitions are expressed in the following pair of identities:

$$S_{t+\tau} \equiv (1 + \tau r^{\text{arith}}_{t,\tau}) S_t$$

$$S_{t+\tau} \equiv e^{\tau r_{t,\tau}} S_t$$

where $r^{\text{arith}}_{t,\tau}$ is the time-t, τ-period arithmetic rate of return, and $r_{t,\tau}$ the time-t τ-period rate of logarithmic rate of return on the asset. Time is usually measured in years. If returns did not vary over time, we could simplify the notation by dropping the first subscript, t. If the periodic rate of return did not vary with the horizon τ, we could drop the second subscript, t, and write r^{arith}_{t} and r_t. In this book, we'll generally adhere to a convention that Roman symbols such as t, T, and t_1, represent dates, while Greek symbols such as τ and τ_1 represent an elapsed time.

The rates of return are the increments to the asset's value. The quantities $1 + r_{t,\tau}^{\text{arith}}$ and $e^{r_{t,\tau}}$ are called the *gross rates of return*, since they include the value of the asset, not just the increment. The gross return is just the proportional change in the asset price:

$$\frac{S_{t+\tau}}{S_t} = 1 + \tau r_{t,\tau}^{\text{arith}} = e^{\tau r_{t,\tau}}$$

The change in the asset price itself is

$$S_{t+\tau} - S_t = \tau r_{t,\tau}^{\text{arith}} S_t = (e^{\tau r_{t,\tau}} - 1)S_t$$

The logarithmic return is the logarithmic change in asset price or the increment to the logarithm of the price:

$$r_{t,\tau} = \frac{1}{\tau}\log\left(\frac{S_{t+\tau}}{S_t}\right) = \frac{1}{\tau}[\log(S_{t+\tau}) - \log(S_t)]$$

Putting these definitions together gives us two equivalent ways of expressing the relationship between arithmetic and log returns:

$$\begin{aligned} r_{t,\tau}^{\text{arith}} &= \frac{1}{\tau}(e^{\tau r_{t,\tau}} - 1) \\ r_{t,\tau} &= \frac{1}{\tau}\log(1 + \tau r_{t,\tau}^{\text{arith}}) \end{aligned}$$

The choice of units is a common source of confusion. In the definitions above, and in most computations, we treat $r_{t,\tau}$ and $r_{t,\tau}^{\text{arith}}$ as decimals. We have to multiply by 100 to get a return in percent, the format in which returns are most often reported. Yet another unit for expressing returns is *basis points*, 0.0001 or hundredths of a percent, typically used to report changes, spreads, or differences in returns. A full percentage point is often called a "point" in market jargon. The gross return and the change in asset price are measured in currency units. For example, if the price was 100 a year ago, and 103 today, we have $103 = (1.03)100$, so the arithmetic rate of return is 0.03 expressed as a decimal, or 3 percent, or 300 bp. The logarithmic rate of return is $\log\left(\frac{103}{100}\right) = 0.02956$, as a decimal.

In the end, we care primarily about profit and loss (P&L) in dollars, so it doesn't really matter whether we describe returns as arithmetic or logarithmic. But there can be practical advantages to using one rather than the other, and logarithmic returns also have a conceptual advantage over arithmetic. Logarithmic returns represent the constant proportional rate at which an asset price must change to grow or decline from its initial to its terminal level, taking into account the growth in the "base" on which returns are measured. We'll discuss this in the context of stochastic processes shortly.

It is therefore easier to aggregate returns over time when they are expressed logarithmically, using the rule $e^{r_1\tau_1}e^{r_2\tau_2} = e^{r_1\tau_1+r_2\tau_2}$. Suppose we are interested in monthly and in annual returns, and are measuring time in months. We have a series of month-end asset prices $S_0, S_1, S_2, \ldots, S_{12}$, with S_0 the asset price at the end of the last month of the prior year. The one-month logarithmic returns are

$$r_{t,t+1} = \log(S_t) - \log(S_{t-1}) \quad t = 1, 2, \ldots, 12$$

The one-year logarithmic return is the sum of the one-month logarithmic returns:

$$\log\left(\frac{S_{12}}{S_0}\right) = \log(S_{12}) - \log(S_0) = \sum_{t=1}^{12} r_{t,t+1}$$

The arithmetic rates of return are not conveniently additive in this way. The one-year arithmetic rate of return is the geometric average of one-month rates:

$$\frac{S_{12}}{S_0} - 1 = \frac{S_1}{S_0}\frac{S_2}{S_1}\cdots\frac{S_{12}}{S_{11}} - 1 = \prod_{t=1}^{12}(1 + r^{\text{arith}}_{t,t+1}) - 1$$

This is a more cumbersome calculation than for logarithmic returns. For example, if we know that an investment had quarterly logarithmic returns of 1, 1.5, −0.5, and 2.0 percent in successive quarters, we can see immediately that its logarithmic return for the year was 4.0 percent. If an investment had quarterly arithmetic returns of 1, 1.5, −0.5, and 2.0 percent, the annual arithmetic return is a harder-to-compute 4.0425 percent.

An even more striking example is this: If you have a positive logarithmic return of 10 percent one day, and negative 10 percent the next, you get back to the original asset price; that is, the gross two-day return is zero. If you have a positive arithmetic return of 10 percent one day, and negative 10 percent the next, the gross two-day return is *negative*. One context in which the distinction between arithmetic and logarithmic returns is quite important when returns are compounded is *leveraged ETFs*, ETFs that are constructed using derivatives contracts and pay a multiple of the daily arithmetic return on an index. Because their returns are defined daily, in a volatile market there can be a large difference between the average daily return on a leveraged ETF and the return over longer periods of time.

However, arithmetic returns are easier to aggregate over positions. Let's introduce a bit of notation we'll use frequently. Let $S_{t,n}, n = 1, 2, \ldots, N$ represent the prices of N securities and $x_n, n = 1, 2, \ldots, N$ the amounts (e.g., the number of shares or ounces of gold) of those N securities in a portfolio. To keep things simple, assume all the positions are long. The time-t value of the portfolio is

$$V_t = \sum_{n=1}^{N} x_n S_{t,n}$$

and the arithmetic return between time t and $t + \tau$ is

$$\begin{aligned} \frac{V_{t+\tau}}{V_t} - 1 &= \frac{1}{V_t} \sum_{n=1}^{N} x_n (S_{t,n} - S_{t+\tau,n}) \\ &= \frac{1}{V_t} \tau \sum_{n=1}^{N} x_n r_{t,\tau,n}^{\text{arith}} S_{t,n} \\ &= \frac{1}{V_t} \tau \sum_{n=1}^{N} \omega_n r_{t,\tau,n}^{\text{arith}} \end{aligned}$$

where

$$\omega_n = \frac{x_n S_{t,n}}{V_t} \quad n = 1, 2, \ldots, N$$

are the weights of the positions in the portfolio. (Yes, we needed three subscripts on $r_{t,\tau,n}^{\text{arith}}$ to distinguish the date, the time horizon, and the security.) The portfolio return can be expressed as a weighted average of the arithmetic returns of the assets. There is no corresponding simple relationship between the logarithmic returns of the portfolio and those of the positions.

Another advantage of logarithmic returns becomes clearer later in this chapter when we discuss the standard asset pricing model. No matter how large in magnitude a logarithmic return is, an asset price will remain

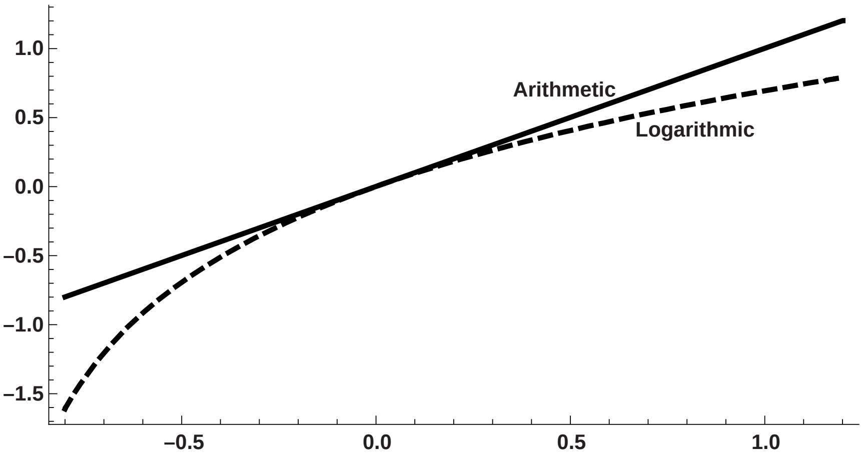

FIGURE 2.1 Approximating Logarithmic by Arithmetic Returns
Arithmetic (x) and logarithmic ($\log(1+x)$) returns for x between 0 and 80 percent.

positive; as the return approaches negative infinity, the asset price will approach, but never reach, zero. An arithmetic return less than -100 percent, in contrast, leads to a negative asset price, which generally makes no sense. This convenient property of logarithmic returns is on motivation for the typical assumption in financial modeling that asset returns are *lognormally distributed*.

For small values of $r_{t,\tau}$ or $r_{t,\tau}^{\text{arith}}$ (that is, small increments $S_{t+\tau} - S_t$), the approximate relationship

$$r_{t,\tau} \approx r_{t,\tau}^{\text{arith}}$$

holds. This approximation of logarithmic by arithmetic returns is a useful and usually innocuous simplification that is widely used in risk measurement practice. As seen in Figure 2.1, the approximation breaks down for large positive or negative values of r_t, which arise when the time interval τ is large, or the return volatility is high. We discuss the approximation further in the context of risk measurement in Chapters 3 and 5.

To see the accuracy of the approximation more exactly, express the log return by expanding $\log(1 + r_{t,\tau}^{\text{arith}})$ around the point $r_{t,\tau}^{\text{arith}} = 0$ to get

$$r_{t,\tau} = r_{t,\tau}^{\text{arith}} - \frac{1}{2}(r_{t,\tau}^{\text{arith}})^2 + \frac{1}{3}(r_{t,\tau}^{\text{arith}})^3 + \cdots$$

The Taylor expansion is equal to the arithmetic return, plus a series in which negative terms are succeeded by smaller-magnitude positive terms.[1]

The log return is smaller than the log of the arithmetic gross return. Thus treating log returns as though they were arithmetic results in a smaller-magnitude estimate of the change in asset price. The opposite is true for negative returns. The table in Example 2.1 compares return computations.

Example 2.1 (Accuracy of Arithmetic Return Approximation) The return in the first column is an arithmetic return. The corresponding log return, that is, the log return that would produce the same change in asset price $S_{t+\tau} - S_t$, is displayed in the second column, and is always a smaller number.

$r_{t,\tau}^{\text{arith}}$	$\log(1 + r_{t,\tau}^{\text{arith}})$
−0.50	−0.693
−0.10	−0.105
−0.01	−0.010
0.01	0.010
0.10	0.095
0.50	0.405

2.2 RISK AND SECURITIES PRICES: THE STANDARD ASSET PRICING MODEL

The standard model in finance theory begins by relating security prices to people's preferences and to security payoffs via an *equilibrium model* of asset prices and returns. Such a model explains observed asset price behavior as the outcome of a market-clearing process rooted in fundamentals: investor preferences and the fundamental productivity and risks of the economy. The mechanisms behind this relationship are individual portfolio optimization and equilibrium relationships. A common approach is the *representative agent model*, which abstracts from differences among individuals or households by defining equilibrium as a state of affairs in which prices have adjusted so that the typical household is just content to hold its asset portfolio and has no incentive to trade assets.

[1]If $r_{t,\tau} < 0$, all the terms are negative.

2.2.1 Defining Risk: States, Security Payoffs, and Preferences

Investing can be thought of as choices among distributions. To reduce uncertainty to probability distributions is a gross oversimplification, since many of the most important uncertainties are "Knightian," and not susceptible to quantification. But in modeling with a distributional hypothesis, even if the distribution is not normal, or not even explicitly stated, this is an oversimplification we have to make.

To set up the standard model, we define the "primitives," the facts about the world. Preferences, in the model world of standard theory, are about two things:

Intertemporal consumption decisions People can be motivated to defer consumption now only in exchange for higher consumption in the future.

Risk aversion People prefer a sure thing to any risky prospect with the same expected value.

In this model world, choices are predictable and take place in a well-specified environment. Outside this world, preferences may include other motivations, such as avoiding losses, and behavior can be surprisingly susceptible to influences that have little bearing on getting "more," and which the standard theory can only consider distractions. These phenomena are studied by *behavioral finance*, and there is a lively debate on whether they are incidental to the more important regularities identified by the standard theory, or are in fact central to explaining financial phenomena in the real world.

We need a probability model, that is, a *state space* and probabilities assigned to each state and all their possible combinations. This tells us what states of the world are possible. State spaces can be discrete and finite, discrete and countable, or continuous. If the state space is finite, with I possible states, we can denote the probability of each by $\pi_i, i = 1, \ldots, I$. Security returns or payoffs are different in each state, that is, they have a probability distribution, with the probabilities of particular values equal to the probabilities of the states in which the payoffs have those values. Risk is then defined as the positive probability of ending up in a state of the world in which returns or consumption are low.

As clear as this definition of risk seems, one immediately runs into knotty problems, particularly that of ranking securities that have different return probability distributions. Suppose we're comparing two securities. If one has a higher payoff than the other in every state of the world, it is clearly preferable. But what if it has a higher payoff in some, but a lower payoff in other states? There are a few common approaches to ranking securities:

Mean-variance dominance provides a partial ordering of investments, ranking those with the same variance but a higher mean, or those with the same mean but a lower variance, as preferable. But this approach doesn't rank securities with both a different mean and variance. Also, the next section describes why the correlation of a security's returns with the general state of the economy in the future is important. A security that pays off when times are bad can be preferred to another that has a higher expected return and lower volatility but pays off most in good times when you least need the extra income.

Stochastic dominance focuses on the entire distribution of payoffs, and stipulates says that payoff distributions that are skewed more to high payoffs are ranked higher. Comparing two payoff distributions, if one generally has a lower probability of very low returns and a higher probability of very high payoffs than the other, it may be ranked higher, even if it has a lower mean and higher variance. We examine "fat-tailed" and skewed distributions in Chapter 10.

Expected utility is an approach in which investments are ranked by the expected value of the investor's *utility function*, a measure of the satisfaction or welfare he derives from different amounts consumption.

Preferences can be summarized in an *expected utility function*, the arguments of which are consumption or wealth, now and at future dates. For simplicity, we will deal with utility as a function of consumption at a set of discrete dates $t = 0, 1, \ldots$ Time 0 represents the present, and $u(c_t)$ represents utility at time t as a function of consumption c_t. This leads to a *time-separable* utility function, in which expected utility is a sum of current and expected utilities. There is no money in this asset pricing model: the units of c_t, the consumption good, are called the *numéraire* for the model.

To be consistent with the basic premisses of the model, the utility function has to *discount* future consumption relative to consumption in the present. We do this via a parameter $0 < \delta < 1$ representing the *pure rate of time preference*. In our examples, we'll set $\delta = 0.95$, meaning a risk-free interest rate of about 5 percent makes the individual barely indifferent between a little additional utility now and later.

The utility function, of course, must have positive marginal utility $u'(c_t) > 0$ for any value of c_t; more is better. To display *risk aversion*, the marginal utility of consumption also has to decline as the level of consumption rises, that is, utility is a *concave* function of consumption, $u''(c_t) < 0$. With such a utility function, the prospect of consuming either two meatballs or no meatballs, with a 50-50 chance of each, will be less attractive than consuming one meatball for sure. More is still better, but much more is not quite that much better. The *Arrow-Pratt measure of risk aversion* is a standard measure equal to $-cu''(c)/u'(c)$.

Example 2.2 A simple example is *logarithmic utility*:

$$u(c) = \log(c)$$

Marginal utility is

$$u'(c) = \frac{1}{c}$$

which is positive but declining as c gets larger, that is, $u''(c_t) < 0$. The Arrow-Pratt measure of risk aversion for logarithmic utility is

$$-c\frac{u''(c)}{u'(c)} = c\frac{1}{c^2}c = 1$$

In a two-period model, consumption today is a quantity c_0 that is known with certainty once savings decisions are taken. Consumption tomorrow is a random quantity $c_1^{(i)}, i = 1, \ldots, I$. The expected utility function is

$$u(c_0) + \delta\mathbf{E}\left[u(c_1)\right] = u(c_0) + \delta\sum_i^I \pi_i u(c_1^{(i)})$$

where $\mathbf{E}[x]$ represents the mathematical expected value of the random variable x. The future payoff of a security is a random variable, that is, it varies by state. It can be a relatively simple one, such as a two-state model in which a coin toss determines whether the security will have a low or a high payoff, or a relatively complex one, for example, a normally distributed variate with a payoff anywhere on $(-\infty, +\infty)$, but with small-magnitude outcomes much more likely than extreme ones.

Let's flesh out the simpler example, a two-state, two-period model with two securities, one risky and one risk-free. The two periods are labeled 0 and 1. The payoff on one unit of the security at $t = 1$ can be represented as a random variable y_1, and the payoff in each state by $y_1^{(i)}, i = 1, 2$. We're giving the payoffs a subscript "1" because, while we're only looking at one risky security at the moment, in most contexts we study several risky securities or risk factors. The risk-free security has a payoff of 1 in each state. The prices at time 0 of the risky and risk-free securities are denoted S_1 and S_f.

We also need to stipulate what resources are available to the individual. The simplest assumption is that in each period and in each state, a certain "endowment" of the consumption good is settled upon him like manna. This is called the *endowment process*, denoted $(e_0, e_1, \ldots)$, and with the $e_1, e_2, \ldots$ random variables. There is no uncertainty about consumption in the present, only about the future. The endowment process is the model's representation of the output fundamentals of the economy. In this highly stylized model, there is no concept of investment in real productive capacity. The time-t endowment in each state is denoted $e_t^{(i)}, i = 1, 2$.

We can compute returns from the asset prices. Any risky asset has a state-dependent arithmetic rate of return r_n, with realizations

$$r_n^{(i)} = \frac{y_n^{(i)}}{S_n} - 1 \qquad n = 1, 2, \ldots; i = 1, \ldots, I$$

$$r_f = \frac{1}{S_f} - 1$$

The second equation describes the relationship between the price S_f of the risk-free asset and its return r_f. (We are sticking to arithmetic rates in the examples of this section because we are dealing with a discrete-time problem. We can also simplify the subscripts because both the date [0] and the period [1] are unambiguous.)

As we will see shortly, we can always calculate a risk-free rate, even if there isn't really a tradeable risk-free asset. So we can define the *excess returns* of asset n as

$$r_n - r_f \qquad n = 1, \ldots$$

The excess return $r_n - r_f$ is a random variable, with realizations $r_n^{(i)} - r_f$ in different states. The *expected return* of an asset is defined as $\mu_n = \mathbf{E}[r_n]$.

The expected value of the excess return of asset n

$$\mathbf{E}\left[r_n - r_f\right] = \mu_n - r_f \qquad n = 1, \ldots, N$$

is the *risk premium* of asset n, the expected spread between the equilibrium return on asset n and the risk-free rate.

Example 2.3 In the two-state, two-period version, the individual receives a known amount e_0 of the consumption good at $t = 0$ and a random amount e_1 at $t = 1$. We can summarize the setup of the example in this table:

State	$e_1^{(i)}$	$y_1^{(i)}$	π_i
$i = 1$	1.25	1.50	0.90
$i = 2$	0.50	0.25	0.10

In State 2, consumption is very low, but it also has a low probability. The risky security behaves like "nature," but is even more generous in the high-probability state and even stingier in the low-probability state.

2.2.2 Optimal Portfolio Selection

The next step is to set up the individual's optimization problem. The individual is to maximize the expected present value of utility $u(c_0) + \delta \mathbf{E}[u(c_1)]$, subject to budget constraints

$$c_0 = e_0 - x_f S_f - x_1 S_1$$
$$c_1 = e_1 + x_f + x_1 y_1$$

The choice variables are x_1 and x_f, the amounts of the risky and risk-free securities purchased. The individual is a price taker in this problem; that is, he responds to asset prices as given by the market and can't influence them. We will see shortly how the equilibrium prices are determined.

To solve the problem, substitute the constraints into the utility function to get a simple, rather than constrained, optimization problem

$$\max_{\{x_1, x_f\}} u(e_0 - x_f S_f - x_1 S_1) + \delta \mathbf{E}\left[u(e_1 + x_f + x_1 y_1)\right]$$

The first-order conditions for solving this problem are:

$$
\begin{aligned}
S_1 u'(e_0 - x_f S_f - x_1 S_1) &= \delta \mathbf{E}\left[y_1 u'(e_1 + x_f + x_1 y_1)\right] \\
&= \delta \sum_i^s \pi_i y_1^{(i)} u'\left(e_1^{(i)} + x_f + x_1 y_1^{(i)}\right) \\
S_f u'(e_0 - x_f S_f - x_1 S_1) &= \delta \mathbf{E}\left[u'(e_1 + x_f + x_1 y_1)\right] \\
&= \delta \sum_i^s \pi_i u'\left(e_1^{(i)} + x_f + x_1 y_1^{(i)}\right)
\end{aligned}
$$

Each of these first-order conditions has the structure

$$
S_n = \delta \frac{\mathbf{E}\left[y_n u'(c_1)\right]}{u'(c_0)} \qquad n = 1,\ f
$$

setting the asset price equal to the trade-off ratio between present and expected future satisfaction. This ratio is called the *marginal rate of substitution* or *transformation*.

Example 2.4 If utility is logarithmic, the individual solves

$$
\max_{\{x_1, x_f\}} \log(e_0 - x_f S_f - x_1 S_1) + \delta \mathbf{E}\left[\log(e_1 + x_f + x_1 y_1)\right]
$$

leading to the first-order conditions

$$
\begin{aligned}
\delta \mathbf{E}\left[y(e_1 + x_f + x_1 y_1)^{-1}\right] &= S_1(e_0 - x_f S_f - x_1 S_1)^{-1} \\
\delta \mathbf{E}\left[(e_1 + x_f + x_1 y_1)^{-1}\right] &= S_f(e_0 - x_f S_f - x_1 S_1)^{-1}
\end{aligned}
$$

Substituting the parameters of our example, we have

$$
\begin{gathered}
0.95[0.9 \cdot 1.50(1.25 + x_f 1.50 + x_1)^{-1} + 0.1 \cdot 0.25(0.50 + x_f + x_1 0.25)^{-1}] \\
= S_1(1 - x_f S_f - x_1 S_1)^{-1} \\
0.95[0.9(1.25 + x_f 1.50 + x_1)^{-1} + 0.1(0.50 + x_f + x_1 0.25)^{-1}] \\
= S_f(1 - x_f S_f - x_1 S_1)^{-1}
\end{gathered}
$$

We can solve these two equations numerically for the asset demand functions $x_f(S_f, S_1)$ and $x_1(S_f, S_1)$. For example, for $S_f = 0.875$ and $S_1 = 1.05$—which are not necessarily the equilibrium prices—we have

$x_f(0.875, 1.05) = -0.066$ and $x_1(0.875, 1.05) = 0.062$. At those prices, the individual will borrow at the (arithmetic) risk-free rate of $0.875^{-1} - 1 = 0.1429$ to invest in more of the risky asset. The latter, at the stipulated prices, would have an expected return of

$$0.9\left(\frac{1.50}{1.05} - 1\right) + 0.1\left(\frac{0.25}{1.05} - 1\right) = 0.3095$$

and a risk premium of 16.67 percent.

The two asset demand functions are each downward-sloping in its own price and upward-sloping in the price of the other asset. A rise in the price of each asset decreases demand for it, but increases demand for the other asset.

2.2.3 Equilibrium Asset Prices and Returns

The final step is to nail down the equilibrium asset prices S_f and S_1. In the model environment we have set up, there is an easy way to do this. Since we have specified the individual's endowment, we can dispense with setting up a model of the production and supply side of the economy. Instead, we cast the individual as a representative agent, imagining him as either the one individual in the economy or as one of myriad identical individuals.

Equilibrium prevails if the net demand for each asset is zero. The supply of the consumption good is already set by the endowment process, and there are no positive quantities of any other assets. The securities in our example are shadow entities, and their prices must adjust so the representative agent desires to hold exactly zero amounts of them.

In this model, therefore, we know the individual will eat exactly his endowment, that is,

$$c_t^{(i)} = e_t^{(i)} \qquad t = 1, \ldots; i = 1, \ldots, I$$

so we know not only the endowment process, but the *consumption* process. This leads to a solution of the equilibrium problem. It also provides a simple approach to asset pricing that has had enormous influence on academic and practitioner work.

Each asset price is a similar function of the same set of arguments, namely, the marginal rate of substitution between utility now and utility later. For the two-security, two-period problem, the first-order conditions are

$$\delta \mathbf{E}\left[y_1 u'(e_1)\right] = \delta \sum_i^I \pi_i y_1^{(i)} u'\left(e_1^{(i)}\right) = S_1 u'(e_0)$$

$$\delta \mathbf{E}\left[u'(e_1)\right] = \delta \sum_i^I \pi_i u'\left(e_1^{(i)}\right) = S_f u'(e_0)$$

or

$$S_1 = \delta \frac{\mathbf{E}\left[y_1 u'(e_1)\right]}{u'(e_0)}$$

$$S_f = \delta \frac{\mathbf{E}\left[u'(e_1)\right]}{u'(e_0)}$$

Example 2.5 In our example, the first-order conditions simplify in equilibrium to discounted expected values of the marginal utilities of the future payoffs of each security:

$$S_1 = \delta e_0 \mathbf{E}\left[y_1 e_1^{-1}\right] = 0.95\left(0.9\frac{1.50}{1.25} + 0.1\frac{0.25}{0.50}\right) = 1.0735$$

$$S_f = \delta e_0 \mathbf{E}\left[e_1^{-1}\right] = 0.95\left(0.9\frac{1}{1.25} + 0.1\frac{1}{0.50}\right) = 0.874$$

When we first looked at the first-order conditions, these trade-offs were seen to depend on the individual's portfolio choices; given the prices S_1 and S_f, we know what the household will choose. But in the no-trade equilibrium of the representative agent model, given agent preferences and the endowment process, choices and prices mutually adjust. Thus, given future consumption, state-by-state, we know what the shadow asset prices must be.

This leads to a set of pricing arguments that varies by state, and combines information about both pure time discounting and risk aversion. It is called the *stochastic discount factor* (SDF), *pricing kernel*, *pricing functional*, or the *state prices*, depending on which of its characteristics you want to emphasize. We'll denote it κ. State prices are a random variable

$$\kappa_i = \delta \frac{u'\left(c_1^{(i)}\right)}{u'(c_0)} \qquad i = 1, \ldots, I$$

The SDF is a state-by-state list of marginal rates of transformation of present for future consumption, in equilibrium and given the current and future endowments.

The first-order conditions can be rewritten so as to display the asset prices as simple linear functions of κ, the payoffs, and their probabilities:

$$S_1 = \mathbf{E}\left[y_1\kappa\right] = \sum_i^I \pi_i y_1^{(i)} \kappa_i$$

$$S_f = \mathbf{E}\left[\kappa\right] = \sum_i^I \pi_i \kappa_i$$

The state-price approach simplifies asset pricing because the state prices come directly out of the pure intertemporal optimization problem, in which there are no assets, only a consumption process over time. First we identify the "pricing calculator"; information about individual asset payoffs can then be added in a later step. The pricing formulas hold for any riskless or risky asset n, and can be computed as long as we know its payoffs $y_n^{(i)}$, $i = 1, \ldots, I$:

$$S_n = \mathbf{E}\left[y_n\kappa\right] = \sum_i^I \pi_i y_n^{(i)} \kappa_i \qquad n = 1, 2, \ldots$$

The payoffs are different, but the state price density κ is the same for all n. That's what we mean by "shadow" assets. They don't really have to exist for us to price them this way, and we don't need a complete list of assets in advance of pricing any individual one of them. In particular, there does not have to actually be a riskless asset for us to know what its price would be if such a thing were to exist. That is convenient, because, as we see in Section 2.5, it is hard to find an asset that is truly riskless.

The pricing formula can be expressed equivalently in terms of returns as

$$1 = \mathbf{E}\left[(1 + r_n)\kappa\right] = \sum_i^I \pi_i \left(1 + r_n^{(i)}\right) \kappa_i$$

$$1 = \mathbf{E}\left[(1 + r_f)\kappa\right] = (1 + r_f) \sum_i^I \pi_i \kappa_i$$

The risk-free rate is higher when future consumption is higher on average and less uncertain. The price of the risk-free asset is the expected value of the SDF, and is related to r_f, the risk-free rate of return, by

$$S_f = \frac{1}{1+r_f} = \mathbf{E}[\kappa] = \delta \mathbf{E}\left[\frac{u'(e_1)}{u'(e_0)}\right]$$
$$= \sum_i^I \pi_i \kappa_i = \delta \frac{1}{u'(c_0)} \sum_i^I \pi_i u'\left(c_1^{(i)}\right)$$

Example 2.6 The state prices in our example are

State	κ_i
$i = 1$	0.76
$i = 2$	1.90

The state price for state 1, $\kappa_1 = 0.76$, is much smaller than $\kappa_2 = 1.90$, because consumption is higher in state 1, lowering its marginal utility and thus the value of an additional unit of it.

The risk-free return is $(0.874)^{-1} - 1$ or 14.4 percent, and the returns and excess returns of the risky asset in each state are

State	$r_1^{(i)}$	$r_1^{(i)} - r_f$
$i = 1$	39.7	25.3
$i = 2$	−76.7	−91.1

The expected risky return and excess return are 28.1 and 13.7 percent.

The risk premium is just a different way of expressing the price of a risky asset, but has an illuminating relationship to the stochastic discount factor. Subtract the equilibrium pricing condition of risky asset n from that of the risk-free asset:

$$\mathbf{E}\left[(r_n - r_f)\kappa\right] = 0$$

The expected value of the excess return on any risky asset times the SDF equals zero. In other words, at equilibrium prices, you are exactly compensated for the risk you take.

We'll transform this expression using the *covariance*, which measures the extent to which fluctuations in two random variables coincide. For any random variables x and y, the covariance is defined as

$$\mathrm{Cov}(x, y) = \mathbf{E}\left[(x - \mathbf{E}[x])(y - \mathbf{E}[y])\right]$$

Next, use the relationship $\mathrm{Cov}(x, y) = \mathbf{E}[xy] - \mathbf{E}[x]\,\mathbf{E}[y]$ for any random variables x and y to get

$$\mathbf{E}\left[(r_n - r_f)\kappa\right] = \mathrm{Cov}(r_n - r_f, \kappa) + \mathbf{E}\left[r_n - r_f\right]\mathbf{E}[\kappa] = 0$$

implying

$$\mathbf{E}\left[r_n - r_f\right]\mathbf{E}[\kappa] = -\,\mathrm{Cov}(r_n - r_f, \kappa)$$

Now, r_f is not random, but known at $t = 0$, so $\mathrm{Cov}(r_n - r_f, \kappa) = \mathrm{Cov}(r_n, \kappa)$ and

$$\mathbf{E}\left[r_n - r_f\right] = -\frac{\mathrm{Cov}(r_n, \kappa)}{\mathbf{E}[\kappa]} = -\frac{\mathrm{Cov}(r_n, \kappa)}{\mathrm{Var}(\kappa)}\frac{\mathrm{Var}(\kappa)}{\mathbf{E}[\kappa]}$$

The first term, $\frac{\mathrm{Cov}(r_n,\kappa)}{\mathrm{Var}(\kappa)}$, is the regression coefficient of the risky security's returns on the SDF and is called the *beta* of the risky security to the SDF κ. The beta is often interpreted as the variation in the risky security's return that can be "explained by" or "predicted from" the SDF's variation across states. The beta depends on the return distribution of the specific security as well as—via the SDF—preferences and the endowment. It is positive when the risky security has high payoffs in high-SDF states, that is, states in which consumption is low and its marginal utility is high.

The second term,

$$\frac{\mathrm{Var}(\kappa)}{\mathbf{E}[\kappa]} = (1 + r_f)\,\mathrm{Var}(\kappa)$$

is a measure of aggregate risk called the *price of risk*. It is the same for all assets and is driven by preferences (via δ and the type of utility function) and the variance of future consumption across states. The variance of the SDF can be high either because the dispersion of future consumption, that is, the fundamental risk in the economy, is high, or because risk aversion is high, pulling apart the marginal rates of transformation for a given distribution of future consumption.

If the beta is negative, that is, the risky security tends to pay off more when the SDF is low and the endowment is plentiful, and the SDF also has high dispersion, then prices of risky assets will be driven lower, and their risk premiums higher; hence the minus sign.

Example 2.7 Continuing our example, we have $\text{Var}(\kappa) = 0.650$ and $\text{Cov}(r_n, \kappa) = -0.7125$, so the beta is

$$\frac{\text{Cov}(r_n, \kappa)}{\text{Var}(\kappa)} = -1.0965$$

and the price of risk is

$$\frac{\text{Var}(\kappa)}{\mathbf{E}[\kappa]} = 0.7435$$

The risky security has a positive risk premium, since it has a high payoff in the high-consumption state, when you need that high payoff the least, and vice versa.

2.2.4 Risk-Neutral Probabilities

There is an alternative point of view on this approach to pricing via utility and endowments. Instead of taking the expected value of the asset's payoffs, weighted by the stochastic discount factor, we'll ask an "inverse" question: What probabilities do we need to be able to derive the equilibrium asset price, but now as an expected present value discounted at *the risk-free rate*? These new probabilities are called *risk-neutral probabilities* $\tilde{\pi}_i, i = 1, \ldots, I$, in contrast to the real-life *subjective* or *physical probabilities* we have used up until now to compute expected values. Substituting the risk-neutral for the physical probabilities, but leaving the state space and the set of outcomes unchanged, leads to the *risk-neutral probability measure*.

In the two-period model, the risk neutral probabilities are defined as

$$\tilde{\pi}_i = \frac{\pi_i \kappa_i}{\sum_i^I \pi_i \kappa_i} = \frac{\pi_i \kappa_i}{1 + r_f} \qquad i = 1, \ldots, I$$

This gives us a way of expressing security prices as a straightforward expected present value of the payoffs, weighted using the risk-neutral probabilities, discounting by the risk-free rate. The standard pricing equation can be rewritten

$$S_n = \mathbf{E}[y_n \kappa] = \frac{\tilde{\mathrm{E}}[y_n]}{1+r_f} = \frac{1}{1+r_f} \sum_i^I \tilde{\pi}_i y_n^{(i)} \qquad n = 1, 2, \ldots$$

The notation $\tilde{\mathrm{E}}[x]$ denotes the expected value of x, using risk-neutral rather than physical probabilities.

To provide some intuition on risk-neutral probabilities, think for a moment about how asset pricing works in our model. The assets are traded in the market, and the price of each claim will depend on the payoffs of the asset in different states, the endowment or "market" payoff in each state, the probability of each state, and the desirability of getting an additional payoff in that state via investment in the asset. If there are many traders or agents, "desirability" is aggregated through the market. In a representative agent model, it depends on endowments; the marginal utility of future consumption is higher in states in which the endowment is low.

Desirability is important. If two claims have the same payoff in two equiprobable states, they won't have the same price if consumption goods are scarcer in one state than in the other. The impact of return distributions is blended with that of the market's risk preferences.

If there is no arbitrage, the prices of the assets therefore imply a probability distribution of states. It may be quite different from the "true" or physical distribution, but *all the claims will be priced consistently with it.* Conversely, if this is not true, that is, if the prices of two claims reflect different risk-neutral probabilities of some states, then an arbitrage profit is available.

We can also see that the risk-neutral expected return is equal to the risk-free rate:

$$\begin{aligned} 1 &= \mathbf{E}[(1+r_n)\kappa] = \frac{\tilde{\mathrm{E}}[(1+r_n)]}{1+r_f} \\ \Rightarrow \tilde{\mathrm{E}}[r_n] &= r_f \end{aligned}$$

Suppose we have a set of securities that each provide a payoff of \$1 in one of the I future states, and 0 otherwise. These securities are called *elementary* or *contingent claims* or *state price securities*. There is an elementary claim for each state $i = 1, \ldots, I$, with a price ε_i.

Even without knowing the individual prices of the elementary claims, we can make some statements about them. If you had a portfolio of 1 unit of each of the elementary claims, you would receive 1 unit of the consumption good regardless of the future state, so it would be equivalent to having a risk-free bond. The sum of the prices of the elementary claims must therefore equal the price of a risk free-bond $S_f = \frac{1}{1+r_f}$.

To find the individual prices of the elementary claims, write out the pricing formula for a security that pays off one unit in one state, say i:

$$\varepsilon_i = \pi_1\kappa_1 \cdot 0 + \cdots + \pi_i\kappa_i \cdot 1 + \cdots + \pi_I\kappa_I \cdot 0 = \pi_i\kappa_i \qquad i = 1, \ldots, I$$

But then the risk-neutral probabilities are also equal to the present values of the elementary claims prices:

$$\tilde{\pi}_i = (1 + r_f)\pi_i\kappa_i = (1 + r_f)\varepsilon_i \qquad i = 1, \ldots, I$$

These properties also provide an equivalent way of valuing any asset as the price of a bundle of elementary claims.

Example 2.8 In our two-state log utility example, the risk-neutral probabilities and elementary claim prices are

State	π_i	$\tilde{\pi}_i$	ε_i
$i = 1$	0.90	0.78	0.68
$i = 2$	0.10	0.22	0.19

The risk-neutral probability for "bad" state 2 is more than double the physical probability, reflecting the desire of the risk-averse individual to insure a higher level of consumption in that state.

Risk-neutral probabilities combine information about the physical probabilities with information about preferences. They are important for two reasons. First, they provide an alternative pricing approach for securities that can be hedged by other securities, plain-vanilla options providing the classic example. Second, as we see in Chapters 10 and 14, we can not only compute asset prices using risk-neutral probabilities. We can also use observed asset prices to estimate risk-neutral probabilities. A risk-neutral probability distribution can be extracted from forward and option prices, and can shed a great deal of light on market sentiment regarding risk.

2.3 THE STANDARD ASSET DISTRIBUTION MODEL

The standard risk measurement model we describe in this section is focused not so much on the asset price itself, or even its mean, as on its potential range of variation; we are interested in the volatility, not the trend. The standard model says little on the trend of asset prices.

The starting point for the model is the *geometric Brownian motion* or *diffusion* model of the behavior over time of an asset price or risk factor S_t. This model is also the basis for the Black-Scholes option pricing model, and is generally a point of departure for analysis of asset return behavior.

In this standard model, returns are normally distributed. The standard model can be useful when overall or "main body" risk is important, as opposed to extreme moves. The model can also serve as a building block, or a teaching tool. But there is extensive evidence, which we review in Chapter 10, that returns are not normally distributed.

First, we will show how this model is built, so that we can understand its implications. We are interested in price fluctuations, so it is the asset *returns* that we ultimately need to understand. We start with a model of asset returns, and show how to derive the geometric Brownian motion model of the asset price from it. Finally, we see how to measure return volatility using the model.

To model the variability of risk factors over time, we look for a stochastic process that, for a suitable choice of parameters, will mimic the behavior seen in real-world risk factors. A stochastic process is an indexed set of random variables; the index is taken to be time. If we could model an observable risk factor exactly using a particular stochastic process, the historical time series of risk factor values would still not itself *be* the stochastic process. Rather, it would be a *sample path* of the stochastic process. The number of possible sample paths depends on the sample space of the stochastic process.

2.3.1 Random Walks and Wiener Processes

We'll set out a model of r_t, the logarithmic asset return, and then build on that to derive a model of asset price behavior. The geometric Brownian motion model builds on a stochastic process called the *random walk*. A random walk is a random variable that

- Is a function of "time" t, which runs from zero (or some arbitrary positive number) to infinity
- Starts at (0, 0)
- Adds increments to its level at time steps of length Δt
- Moves up or down at each time step by an amount $\sqrt{\Delta t}$
- Moves up with probability π and down with probability $1 - \pi$

Let's denote the steps or increments to the random walk by Y_i and its position or state after n steps by

$$X_n = \sum_{i=1}^{n} Y_i$$

If we chop t into finer and finer steps, that is, make Δt smaller, n will be larger. The Y_i are independent and identically distributed. The number of moves up ($Y_i > 0$) is therefore a binomially distributed random variable with parameters n and π. Denoting by k the number of up-moves in the first n steps, the state or position of the random walk X_n can be written as

$$X_n = [k - (n - k)]\Delta t = (2k - n)\Delta t$$

since the number of down-moves ($Y_i < 0$) in the first n steps is $n - k$.

There are 2^n possible paths $(Y_1, \ldots, Y_n)$, each with a probability $\frac{1}{2^n}$. There are $2n + 1$ possible values $(-n\Delta t, \ldots, -\Delta t, 0, \Delta t, \ldots, n\Delta t)$ of the position X_n, with the probability of each equal to the that of the corresponding binomially distributed number of up-moves. Apart from the extreme "straight up" and "straight down" paths that lead to $X_n = -n\Delta t$ and $X_n = -n\Delta t$, there is more than one way to reach each possible value of X_n.

It's convenient to set $\pi = \frac{1}{2}$. The mean of each step is then

$$\mathbf{E}[Y_i] = \mathbf{E}\left[\frac{1}{2}\sqrt{\Delta t} + \frac{1}{2}\left(-\sqrt{\Delta t}\right)\right] = 0 \quad i = 1, \ldots, n$$

and its variance is

$$\text{Var}[Y_i] = \mathbf{E}\left[Y_i^2\right] - (\mathbf{E}[Y_i])^2 = \frac{1}{2}\left(\sqrt{\Delta t}\right)^2 + \frac{1}{2}\left(-\sqrt{\Delta t}\right)^2 = \Delta t \quad i = 1, \ldots, n$$

From these facts, together with the independence of the steps, we can compute the variance of the position after n steps:

$$\mathbf{E}[X_n] = 0$$
$$\text{Var}[X_n] = n\Delta t$$

The time elapsed after n steps is $t = n\Delta t$, so one property we can see right away is that the variance of the state of a random walk is equal to the time elapsed t, and thus its standard deviation is equal to $\sqrt{t}$. This is an

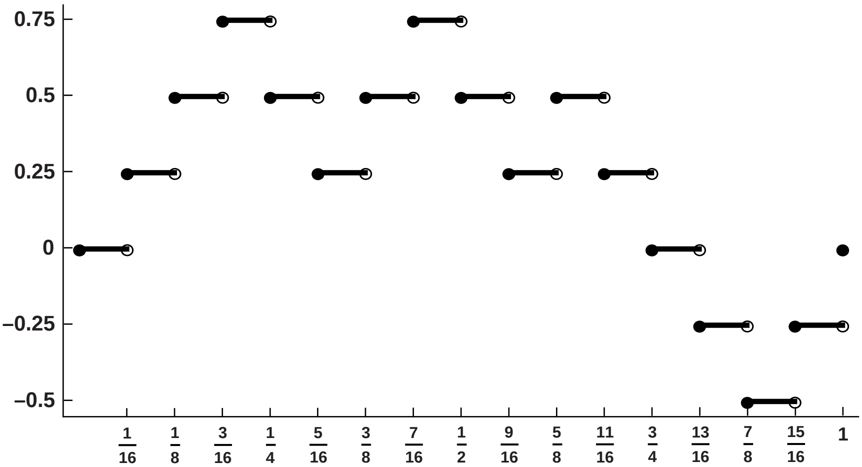

FIGURE 2.2 Sample Path of a Random Walk
Plot shows the first 16 positions of one simulation of a random walk. The total time elapsed in the first 16 steps is set to $t = 1$, so the time interval between adjacent steps is $\Delta t = \frac{1}{16}$ and the magnitude of the increments is $|Y_i| = \frac{1}{\sqrt{\Delta t}} = \frac{1}{\sqrt{16}} = \frac{1}{4}, i = 1, \ldots$ The solid dots show the value the position takes on at the start of each time interval, and the open circles show the points of discontinuity.

important relationship between volatility and the passage of time that we will see repeatedly in different forms.

The random walk is a discrete time or discontinuous stochastic process, as illustrated in Figure 2.2. Any value of t will do, so let's set $t = 1$, so that $\Delta t = \frac{1}{n}$. But now let the number of the time steps go to infinity by holding $t = n\Delta t = 1$ constant, letting $n \to \infty$ and $\Delta t \to 0$. As the number of time steps grows, the mean and variance of the position X_n remain unchanged at 0 and 1, respectively.

The magnitude of the low-probability extreme states $X_n = \pm n\Delta t = \pm 1$ remains constant, but the probability of reaching them decreases as the number of time steps increases, even though the total time elapsed and the variance of the state X_n are *not* increasing. Figure 2.3 displays the distribution of the terminal position X_n for an increasing fineness of the time steps between time 0 and 1.

But computing the probabilities under the binomial distribution of the number of up moves, and thus of X_n, becomes difficult as the number of time steps grows. It is also not necessary. Figure 2.3 suggests that the distribution of the terminal position X_n converges to the normal. In fact, we can invoke the central limit theorem to show that the distribution of X_n converges to a

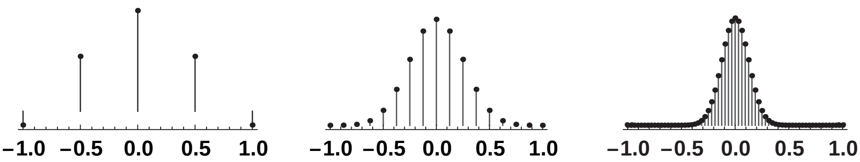

FIGURE 2.3 Convergence of a Random Walk to a Brownian Motion
The plots display the probability density of X_n, the terminal position of a random walk over an interval of length 1, with $n = 4, 16, 64$ time steps and with the probability of a step in either direction set to $\pi = \frac{1}{2}$.

standard normal random variable. The mean and variance haven't changed, but we have an easier distribution than the binomial to work with.

The stochastic process to which the state X_n of the random walk converges is called a *Brownian motion* or *Wiener process* W_t. The value of W_t is a zero-mean normal variate with a variance equal to t:

$$W_t \sim N(0, \sqrt{t})$$

We have defined an increment to a Brownian motion as the limit of a random walk as its time steps shrink to zero, but this is technically not necessary. A Brownian motion is defined uniquely by the following properties:

1. It starts at $W_0 = 0$.
2. For any $t \geq 0$ and $\tau > 0$, the increment to the process over the period $t + \tau$ is normally distributed:

$$W_{t+\tau} - W_t \sim N(0, \sqrt{\tau})$$

   Each such increment can be thought of as composed of infinitely many random walk steps.
3. The increment $W_{t+\tau} - W_t$ is independent of the history of the process up to time t. In other words, the random variables W_t and $W_{t+\tau} - W_t$ are independent. This is called the *martingale property*. In fact, if we chop all of history into intervals such that $0 \leq t_1 < t_2 < \cdots < t_n < \infty$, the increments $W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_n} - W_{t_{n-1}}$ are all independent.
4. Every sample path is continuous.

All other properties of Brownian motion can be derived from this definition. The last one is a major difference from the random walk, which, as we saw, is discontinuous.

Some of Brownian motion's properties are initially counterintuitive, such as the fact that its *total variation process*, the expected value of the total distance a Brownian motion moves over any interval, no matter how short, is infinite. To see this, recall that the magnitude of the random walk step is $|\sqrt{\Delta t}| = \sqrt{\Delta t}$ with certainty, so that is also its expected magnitude. But then the total distance traveled over the interval t is, with certainty, equal to

$$n\sqrt{\Delta t} = \frac{t}{\Delta t}\sqrt{\Delta t} = \frac{t}{\sqrt{\Delta t}}$$

and its limit as $\Delta t \to 0$ is infinity. This is one aspect of Brownian motion's "urge to wander." It is constantly vibrating, so if you take the path it has traveled, even over a very short time, and uncurl it, it is infinitely long, and composed of infinitely many tiny, sharp, up or down moves. For this reason, although continuous, Brownian motion is not smooth. In fact, it cannot be differentiated: $\frac{\partial W_t}{\partial t}$ does not exist.

Yet its *quadratic variation process*, the expected value of the sum of the *squared* distances it moves over any interval, no matter how many, is *finite* and equal to the length of the interval. The quadratic variation of a Brownian motion is another form that the relationship between volatility and the passage of time takes. Although it has a strong urge to wander "in the small"—infinite total variation—it wanders in a contained way "in the large."

To gain intuition about these properties, let's look at Figure 2.4, which illustrates the convergence of a random walk to a Brownian motion. It shows a *single* sample path of a Brownian motion, sampled at progressively finer subintervals of the interval $(0, 1)$. It starts in the upper left hand panel by dividing $(0, 1)$ into 16 subintervals, and ends in the lower right with 1,024 subintervals. In other words, we are approximating the continuous Brownian motion with a sequence of discrete random walks.

Because we are plotting a single sample path, the value of W_t must be the same in each panel for any point in time, such as $\frac{1}{16}$. For example, in the upper right panel, the interval $(0, 1)$ is divided into 64 subintervals. The values of the Brownian motion at times $\frac{1}{16}, \frac{1}{8}, \ldots, \frac{15}{16}, 1$ must be the same as in the upper left panel. But we can't just generate the 64-step version and connect every fourth value with a straight line to get the 16-step version. This would violate the requirement that the random walk move up or down at each time step.

Rather, we generate the 16-step version, and interpolate a 4-step discrete *Brownian bridge* between neighboring points. A Brownian bridge is just a Brownian motion that is constrained to end at a certain point at a certain

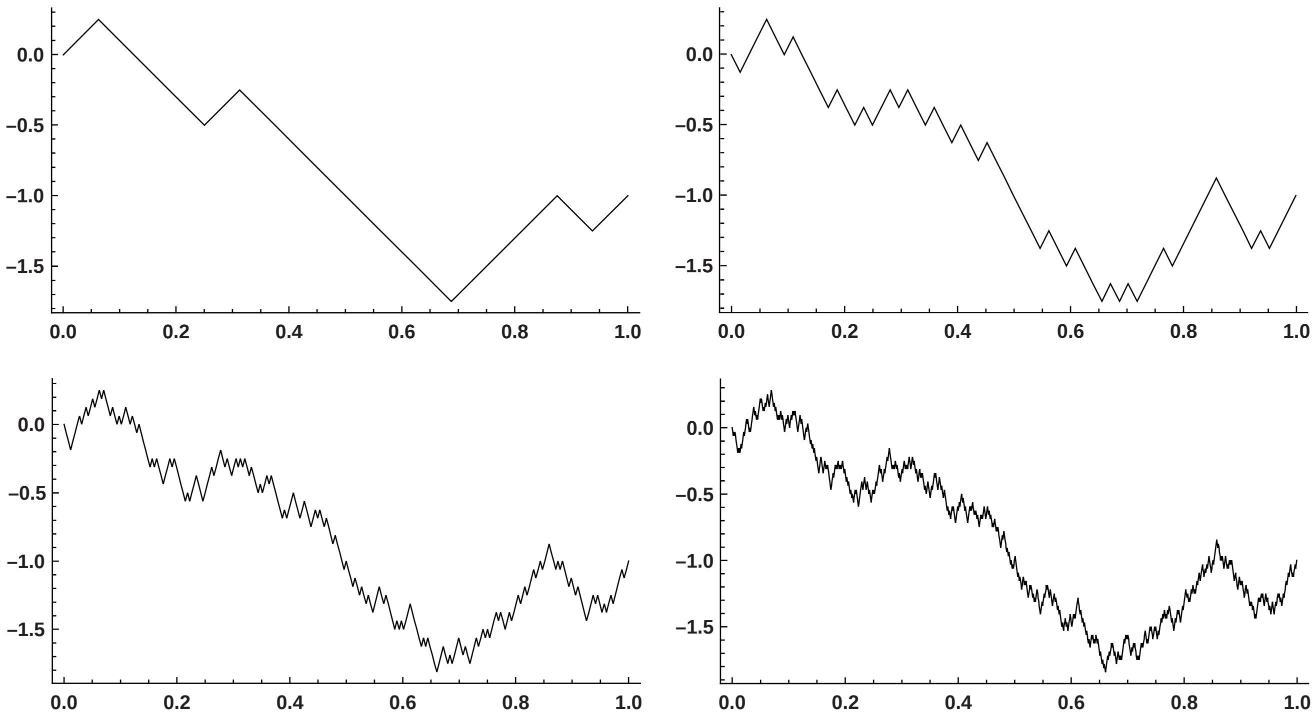

FIGURE 2.4 Convergence of a Random Walk to a Brownian Motion
Random walks with 16, 64, 256, and 1,024 steps, but the same random seed. The graphs are constructed by filling in the intermediate steps with discrete Brownian bridges.

time. In simulation, one uses a Brownian bridge to "thicken up" Brownian motion sample points in a way that is true to the Brownian motion's tendency to wander, while still getting it to the next, already-known, sample point. Let's take a closer look at the construction of a Brownian bridge in order to better understand the variation properties of Brownian motion.

Denote up-moves in the random walk representations by a 1 and down-moves by a 0. The 16-step random walk can then be represented as a random sequence of 0s and 1s. In our example, we are increasing the fineness of the random walk by subdividing each time interval into four equal subintervals in each step of the convergence to a Brownian motion. Since each time step is divided into four, and $2 = \sqrt{4}$, the magnitude of the up- and down-moves is divided by two. If the next step in the coarser random walk is a down-move, there are then four different ways the finer random walk could get there in four steps, with each increment just half the size of the coarser random walk's increments. They can be represented by the 4-dimensional identity matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The four possible bridging paths if there is an up-move are

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Note that the required bridge does not depend on the last step or on the position, but only on the next step. This is an expression of the martingale property, or "memorylessness," of Brownian motion. To generate the finer random walk, at each step of the coarser random walk, depending on whether it is an up or down move, we randomly pick one of the four rows of the appropriate one of these matrices to generate the intermediate four steps.

We can now see how increasing the fineness of the random walk increases the total distance traveled. Each time we refine the random walk by dividing the time steps by a factor of four, multiplying the number of time steps by four, and connecting the steps with a Brownian bridge, we double the distance traveled compared to one step of the coarser random walk. As the random walk becomes finer and finer, and converges to a Brownian motion, the distance traveled—the total variation—approaches infinity. But the quadratic variation has remained at unity.

2.3.2 Geometric Brownian Motion

The Wiener process W_t, with a few changes, provides our model of *logarithmic returns* on the asset:

- We add a *drift* μ, expressed as a decimal per time period t. We can do this by adding a constant $\mu \Delta t$ to each step of the random walk as it converges to W_t.
- We scale the random walk by a *volatility* term σ by multiplying each time step of the original random walk by σ. The volatility will be assumed for now to be a constant parameter, not dependent on the current level of the asset price and not dependent on time.

The convergence results of the last subsection remain the same, but the value of the Brownian motion is now a normally distributed random variable with mean μt and variance $\sigma^2 t$:

$$W_t \sim N(\mu t, \sigma \sqrt{t})$$

This completes the model of *increments to the logarithm* of the asset price or risk factor. A standard way to write this is as a *stochastic differential equation* (SDE):

$$d \log(S_t) = \mu dt + \sigma dW_t$$

The random value of an increment to the log of the asset price over some time interval $(t, t + \tau)$ can be expressed as

$$\log(S_{t+\tau}) = \log(S_t) + \mu\tau + \sigma(W_{t+\tau} - W_t)$$

So we now have a model of log returns:

$$r_{t,\tau} = \mu\tau + \sigma(W_{t+\tau} - W_t)$$

Logarithmic returns are therefore normally distributed:

$$r_{t,\tau} \sim N(\mu\tau, \sigma\sqrt{\tau})$$

Based on this model, we build a model of the *increments to the level* of the asset price. We can't simply take exponents to characterize the level of the asset price itself. Rather, we must apply Itô's Lemma, using the relationship $S_t = e^{\log(S_t)}$. We get the SDE

$$dS_t = \left(\mu + \frac{1}{2}\sigma^2\right) S_t dt + \sigma S_t dW_t \tag{2.1}$$

Intuitively, this makes perfect sense. If the logarithm of the asset price moves randomly up or down in proportion to a volatility factor σ, the random increases will accelerate the exponential growth more than the equally probable random decreases will decelerate them. The reason is that the positive returns increase the asset price level, the base for the next vibration, while the negative returns decrease the base. So the exponent of the process, the asset price level, will grow on average just a bit faster than μ, quite apart from its random fluctuations. Itô's Lemma tells us exactly how much: $\frac{1}{2}\sigma^2$ per time unit.

Another way to see this is to go back to the discrete-time model for a moment and imagine an asset price moving up or down each period by the same proportional amount σ. The asset price starts at some arbitrary positive level S_0, and

$$S_t = \begin{cases} S_{t-1}(1+\sigma) & \text{for} \quad t-1 \quad \text{odd} \\ S_{t-1}(1-\sigma) & \text{for} \quad t-1 \quad \text{even} \end{cases}$$

The volatility of logarithmic changes in the asset is approximately σ if σ is low. For any $\sigma > 0$, S_t will eventually go to zero; how quickly depends on how large is σ. If, however, we adjust the process to

$$S_t = \begin{cases} S_{t-1}\left(1+\sigma+\frac{\sigma^2}{2}\right) & \text{for} \quad t-1 \quad \text{odd} \\ S_{t-1}\left(1-\sigma+\frac{\sigma^2}{2}\right) & \text{for} \quad t-1 \quad \text{even} \end{cases}$$

it will stay very close to S_0 almost indefinitely.

The solution to the SDE, Equation (2.1), for the asset price level is

$$S_t = S_0 e^{\mu dt + \sigma dW_t} \tag{2.2}$$

Figure 2.5 illustrates the behavior of an asset price following geometric Brownian motion, and Figure 2.6 illustrates the behavior of log returns. The

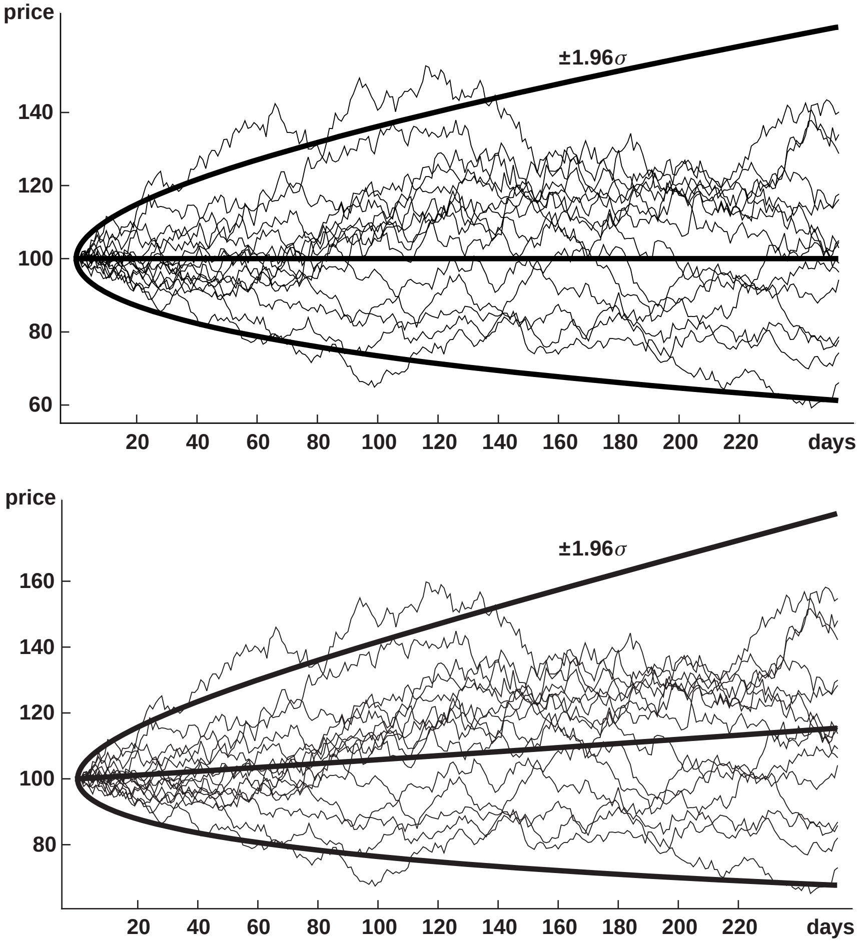

FIGURE 2.5 Geometric Brownian Motion: Asset Price Level
Upper panel (*zero drift*) Fifteen simulations of the price level over time of an asset following a geometric Brownian motion process with $\mu = 0$ and $\sigma = 0.25$ at an annual rate. The initial price of the asset is $S_0 = 100$. The hyperbola plots the 95 percent confidence interval over time: at any point in time, 95 percent of the simulated path should be within it.
Lower panel (*positive drift*) This panel is identical to the one above but with $\mu = 0.20$ at an annual rate.

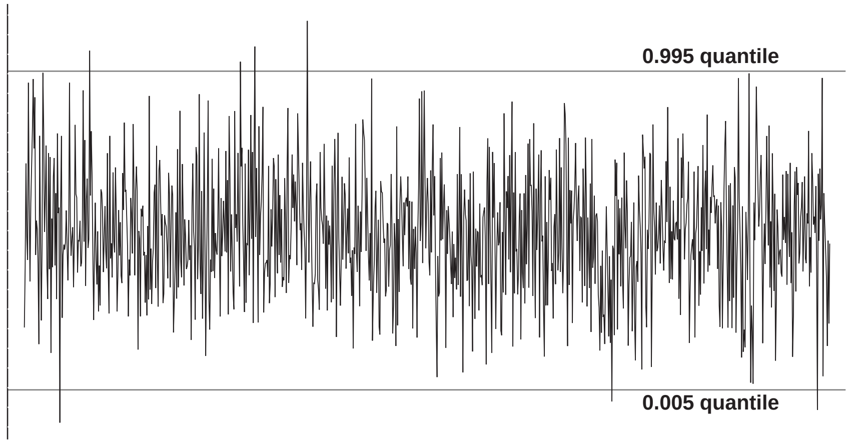

FIGURE 2.6 Geometric Brownian Motion: Daily Returns
Simulation of 1,000 sequential steps of a geometric Brownian motion process with $\mu = 0$ and $\sigma = 0.25$ at an annual rate. The horizontal grid lines mark the 99 percent confidence interval. There should be about five occurrences of returns above and below the grid lines. In most samples of 1,000, however, this will not be exactly the case. This becomes important when we try to assess the quality of model-based risk measures.

asset price has these properties if it follows a geometric Brownian motion process:

- The asset price is equally likely to move up or down.
- The further into the future we look, the likelier it is that the asset price will be far from its current level.
- The higher the volatility, the likelier it is that the asset price will be far from its current level within a short time.
- The time series of daily logarithmic asset *returns* are independent draws from a normal distribution.

Example 2.9 (Distribution of Asset Price Under Standard Model) Using the stochastic process illustrated in Figure 2.5, with $S_0 = 100$, $\mu = 0$, and $\sigma = 0.125$, the expected value of the asset price after 128 business days ($t = 0.5$) is 100, and the 95 percent confidence interval for $S_{0.5}$ is (83.7659, 118.451).

2.3.3 Asset Return Volatility

As we have seen, in the standard model, logarithmic returns on an asset that follows a geometric Brownian motion with a drift μ and volatility σ

are normally distributed as $N(\mu\tau, \sigma\sqrt{\tau})$. Expected return and variance are proportional to the time interval over which the returns are measured:

$$E[r_{t,\tau}] = \mu\tau$$
$$\mathrm{Var}(r_{t,\tau}) = E[(r_{t,\tau} - E[r_{t,\tau}])^2]) = \sigma^2\tau$$

The return volatility or standard deviation is proportional to the square root of the interval:

$$\sqrt{\mathrm{Var}(r_{t,\tau})} = \sigma\sqrt{\tau}$$

This is called the *square-root-of-time rule*. It is commonly used to convert volatilities and covariances from one interval, such as daily, to another, such as annual, and vice versa. In computing volatility for shorter periods than one year, we usually take only trading days rather than calendar days into account. Once holidays and weekends are accounted for, a typical year in the United States and Europe has a bit more than 250 trading days (52 five-day weeks, less 8 to 10 holidays). A convenient rule of thumb is to assume 256 trading days. Converting from annual to daily volatility is then a simple division by $\sqrt{256} = 16$.

Our notation for return doesn't indicate the period over which returns are measured. In the standard model, this doesn't matter, because the square-root-of-time tells us how to move back and forth between measurement intervals. In a different model, the measurement period can be very important. For example, in a mean-reverting volatility model, measuring return volatility over a short period such as one day might lead to a wild overestimate of volatility over a longer period such as one year.

2.4 PORTFOLIO RISK IN THE STANDARD MODEL

The standard model, as we've presented it so far, is a world with just one risk factor: The only source of uncertainty is the randomness of the future endowment. But the world may have more than one "ultimate" source of risk, and portfolios typically contain exposures to many securities. So we need to think about how the risks of the individual securities or risk factors relate to one another. First, of course, we introduce even more notation. We have a portfolio with x_n units, priced at S_n, of each of the N assets in the portfolio, $n = 1, \ldots, N$. The units might be the number of shares of stock or the bond par value or the number of currency units.

We need a model of the joint or multivariate probability distribution of the N returns. The natural extension of the standard model we just outlined is to stipulate that risk factors follow a multidimensional geometric Brownian motion process. The portfolio's logarithmic returns are then jointly normally distributed. In the joint normal model, each of the N assets has a log return distribution

$$r_{t,\tau,n} \sim N(\mu_n \tau, \sigma_n \sqrt{\tau}) \qquad n = 1, \ldots, N$$

In a normal model, we can discuss "risk" and "volatility" nearly interchangeably. We therefore have to repeat our warning that returns are not, in fact, joint normal, so that the theories and tools laid out here should be thought of as expository devices or approximations.

2.4.1 Beta and Market Risk

Suppose we want to know how one asset's return varies with that of another asset or portfolio. The return covariance measures the variance of one asset's return that is associated with another asset's return variance. It is expressed in squared return units:

$$\sigma_{mn} = \mathbf{E}\left[(r_m - \mathbf{E}\left[r_m\right])(r_n - \mathbf{E}\left[r_n\right])\right] \qquad m, n = 1, \ldots$$

where r_m and r_n now represent returns on two different assets as random variables rather than at a point in time.

The return correlation also measures the extent to which the asset returns move together. While the covariance will be affected by typical size of the returns of the two assets, and is hard to compare across asset pairs with different volatilities, correlation is always on $(-1, 1)$ and therefore easy to compare and interpret. The correlation ρ_{mn} is related to volatility $\sigma_n = \sqrt{\mathbf{E}\left[(r_n - \mu_n)^2\right]}$ and covariance by

$$\rho_{mn} = \frac{\sigma_{mn}}{\sigma_m \sigma_n} \qquad m, n = 1, \ldots$$

We previously introduced beta as a measure of the comovement of a security's returns with aggregate economic risk. Beta is also used as a statistical measure of the comovement of the returns of two different assets. The beta of asset m to asset n is defined as the ratio of the covariance of asset m's and asset n's excess returns to the variance of asset n's excess return:

$$\beta_{mn} = \frac{\sigma_{mn}}{\sigma_n^2}$$

The beta is often interpreted as the variation in asset m return that can be "explained by" or "predicted from" asset n's variation. Typically, asset

n is a portfolio representing a major asset class such as stocks or bonds, or a major subclass, such as a national or industry index. Beta is then used as a relative risk measure; it can be viewed as telling you what the risk is of, say, a single stock relative to the stock market as a whole.

The standard way to estimate beta is to run a regression between the two assets' excess returns. The regression equation is:

$$r_{mt} - r_f = \alpha + \beta_{mn}(r_{nt} - r_f) + u_t$$

In the joint normal model, the error terms u_t in this equation are also normally distributed. The direction of the regression matters: in general $\beta_{mn} \neq \beta_{nm}$.

When we look at the beta of one asset's returns to another's or to an index, we are interested not only in "how much beta" there is—is the beta near unity, or negative, well above unity, or near zero?—but also the extent to which variation in the asset's return is attributable to variation in the index return. To measure this, we can look at the *R-squared* or R^2 of the regression, which is equal to the square of the correlation between the dependent and independent variables. Let $\hat{\rho}_{x,y}$ denote the sample correlation coefficient between x and y:

$$R^2 \equiv \hat{\rho}^2_{x,y}$$

We have $0 \leq R^2 \leq 1$. An R^2 close to 1 indicates that the index returns do much to explain movements in the asset. The R^2 of the regression is often interpreted as a measure of the explanatory power of beta.

The beta coefficient is related to the correlation by

$$\rho_{mn} = \beta_{mn}\frac{\sigma_n}{\sigma_m} \qquad \Leftrightarrow \qquad \beta_{mn} = \rho_{mn}\frac{\sigma_m}{\sigma_n}$$

So the beta is close to the correlation if the two variables are about the same in variability. If the dependent variable, say an individual stock, has a low volatility compared to the independent variable, say, a stock index, then even if the correlation is close to unity, the beta will be low.

Example 2.10 (Descriptive Statistics for EUR-USD and USD-JPY) For EUR-USD and JPY-USD price returns, November 1, 2005, to October 31, 2006, we have the following descriptive statistics. An estimate of their joint return distribution, under the assumption the pair of currency returns is a bivariate normal, is illustrated in Figure 2.7. A scatter plot of the returns and the estimated regression line of USD-JPY on EUR-USD are displayed in Figure 2.8. The ellipse plotted in Figure 2.8 contains 95 percent of the

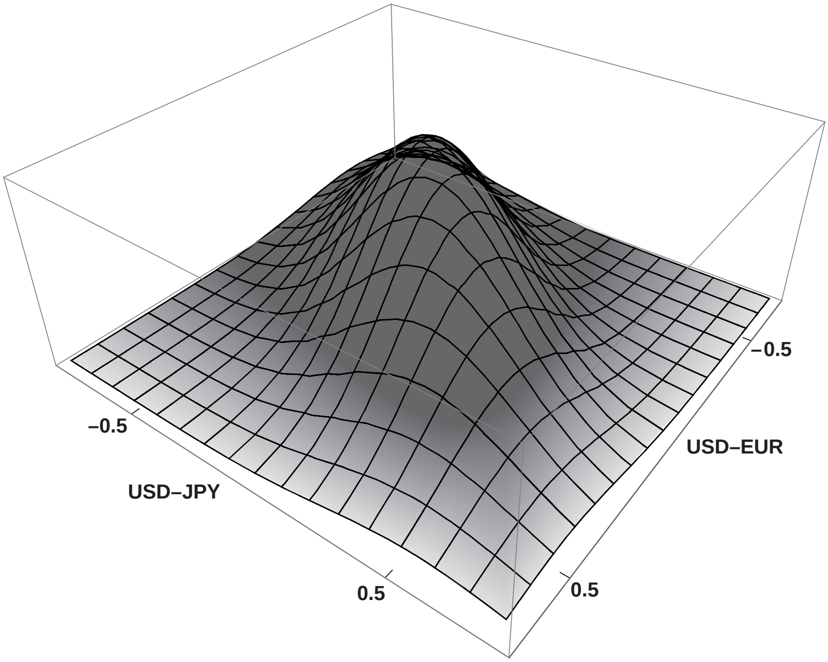

FIGURE 2.7 Joint Distribution of EUR and JPY Returns
Estimated joint normal density of logarithmic daily EUR-USD and JPY-USD price returns, January 2, 1996, to October 6, 2006. Prices are expressed as the dollar price of the non-USD currency. A slice parallel to x, y-plane corresponds to confidence region for the two returns.
Source: Bloomberg Financial L.P.

observed returns. The volatilities are, using $\sqrt{252} = 15.8745$ as the square-root-of-time factor:

	EUR-USD	JPY-USD
Daily return volatility	0.61	0.68
Annual return volatility	9.69	10.79

The statistics relating the two currencies to one another are

Covariance of daily returns	14.80×10^{-5}
Correlation of daily returns	0.357
Beta of EUR-USD to JPY-USD	0.320
Beta of JPY-USD to EUR-USD	0.398

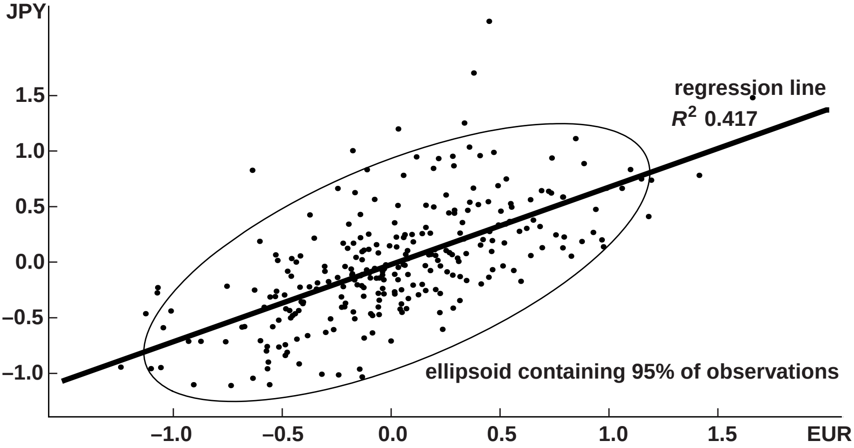

FIGURE 2.8 Correlation and Beta
Scatter plot of logarithmic daily EUR-USD and JPY-USD price returns, Nov. 1, 2005, to Oct. 31, 2006, and regression of USD-JPY on EUR-USD. Exchange rates are expressed as the dollar price of the non-USD currency.
Source: Bloomberg Financial L.P.

Conversely, one can easily have a high beta, but a low correlation and thus a low R^2, if the dependent variable has a very high volatility compared to the index. This will happen if an individual stock's return has only a loose association with the market's return, but, when the individual stock price moves, it moves a lot.

Figure 2.9 illustrates this possibility. In constructing the graph, we model the dependent and independent variables—the asset and index returns—as jointly normally distributed variates with zero mean. Once we have a model, we can state the population α, β, and R^2. The population α is zero, while the population β is given by the covariance matrix—that is, the volatilities and correlation—of the asset returns.

Suppose we fix the index return volatility at 15 percent. There are many ways to get, say, $\beta = 1$. If we set $\sigma_y = 2\sigma_x$, then we must have $\rho_{x,y} = 0.5$ and $R^2 = (0.5)^2 = 0.25$:

$$1 = \frac{2\sigma_x}{\sigma_x} 0.5$$

Similarly, if $\sigma_y = 4\sigma_x$, that is, the dependent variable is highly volatile, the correlation must be only $\rho_{x,y} = 0.25$ and $R^2 = (0.25)^2 = 0.0625$. The slope of the line doesn't change, but the goodness-of-fit is much lower.

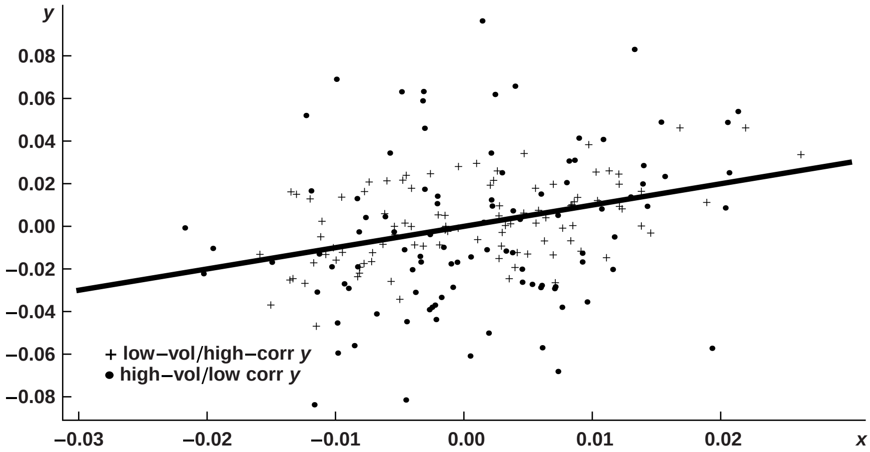

FIGURE 2.9 Volatility and Beta
Scatter plot of daily returns of joint realizations of a low-volatility and high volatility dependent and independent variates, and the common population regression line. The dependent variable volatility is set at $\sigma_x = 0.15$ per annum. The high-volatility dependent variable has $\sigma_x = 0.60$ and $\rho_{x,y} = 0.25$, and is represented by points. The low-volatility dependent variable has $\sigma_x = 0.30$ and $\rho_{x,y} = 0.50$, and is represented by $+$'s.
The population regression line has a slope $\beta = 1$ and intercept $\alpha = 0$ in both cases. If we did enough simulations, the sample regression lines would very likely coincide almost exactly with the population regression line. However, the high-volatility observations are vertically strewn very widely around the regression line compared to the low-volatility simulations.

Beta is often used in the context of the *capital asset pricing model* (CAPM), an equilibrium model framed in terms of security returns rather than consumption and endowments. The CAPM makes a strong assumption, mean-variance optimization, about investor preferences. In this model, a household's utility is a function of portfolio return and volatility rather than of consumption, and it seeks to maximize return for a given level of volatility. Like the consumption-based model developed above, the CAPM's major results can be expressed in terms of beta. The equilibrium return r_n of any security n can be expressed as

$$\mathbf{E}\left[r_n - r_f\right] = \beta_n \mathbf{E}\left[r_{\text{mkt}} - r_f\right] \qquad n = 1, 2, \ldots$$

where r_{mkt} represents the *market portfolio*. The market portfolio can be interpreted as an aggregate measure of risk, or as an aggregate of all the risky securities outstanding. In practical applications, it is represented by a broad market index such as the S&P 500 stock market index or a global stock market index such as the MSCI.

The beta relationship states that the expected excess return on any asset is proportional to the excess return of the market portfolio. The ratio of single-asset to market excess returns depends on security n's beta. This relationship is similar to that derived above in the standard asset pricing model. In both models, the risk premium of any individual security is a function of two components: a beta, which depends on the risk characteristics of the security itself, and a second component, the risk premium of the market portfolio, which is the same for all securities. In both models, there is single risk factor, whether the price of risk derived from the SDF, or the risk premium of the market portfolio, that embodies equilibrium prices, given preferences and fundamentals.

There are, however, important differences between β_n and $\frac{\mathrm{Cov}(r_n,\kappa)}{\mathrm{Var}(\kappa)}$. The CAPM beta is *positive* for securities that pay off in prosperous times, when the market portfolio is high, while $\frac{\mathrm{Cov}(r_n,\kappa)}{\mathrm{Var}(\kappa)}$ is positive for securities that pay off in lean times when consumption is low. That is why the CAPM beta relationship omits the minus sign.

A further major result of the CAPM is that, in equilibrium, asset prices must have adjusted so that the market portfolio is the optimal mix of risky assets for everyone. Portfolio choice is not about choosing securities, since prices will adjust in equilibrium so that households are content to own the mix of securities outstanding in the market. Rather, it is about choosing the mix of the risky market portfolio and a riskless asset. The latter choice depends on preferences and may differ across households with different utility functions.

The CAPM is not completely satisfying, because it relies on the very specific mean-variance optimization assumption about preferences, and because the concept of a market portfolio is both nebulous and unobservable. The *arbitrage pricing theory* (APT) is one response. Rather than making strong assumptions about preferences, it assumes that returns are linearly related to a set of (possibly unobservable) factors, and that arbitrage opportunities are competed away, at least for the most part.

The APT starts with a linear factor model of security n's returns:

$$r_n = \mathbf{E}[r_n] + \beta_{n1} f_1 + \beta_{n2} f_2 + \cdots + \beta_{nK} f_K + \epsilon_n \qquad n = 1, 2, \ldots$$

The factors $f_1, \ldots, f_K$ are random variables, representing surprises or "news" such as a higher-than-expected inflation rate, or lower-than-expected profits in a particular industry. The β_{nk} are betas or sensitivities to the factors f_k. The model imposes these restrictions:

$$
\begin{aligned}
E[f_k] &= 0 \qquad \forall n = 1, \ldots, K \\
E[f_j f_k] &= 0 \qquad \forall j, k = 1, \ldots, K \\
E[\epsilon_n] &= 0 \qquad \forall n = 1, 2, \ldots \\
E[\epsilon_m \epsilon_n] &= 0 \qquad \forall m, n = 1, 2, \ldots
\end{aligned}
$$

The first two restrictions state that, on average, there are no surprises in the factors and that the factors are independent of one another. The last two restrictions state that security prices adjust, up to a random error, so that expected excess returns are just what the factor model says they should be.

This model of expected returns is similar to, but more plausible than, the CAPM's single ultimate source of risk, embodying all the fundamental risks in the economy. The assumption of no or limited arbitrage possibilities is more attractive than a specific hypothesis on preferences. Using the factor-based return model, together with the no-arbitrage assumption, the APT arrives at this representation of equilibrium returns:

$$
\mathbf{E}\left[r_n - r_f\right] \approx \beta_{n1}\lambda_1 + \beta_{n2}\lambda_2 + \cdots + \beta_{nK}\lambda_K \qquad n = 1, 2, \ldots
$$

where the β_{nk} are the same factor sensitivities as in the return model, and the λ_k are the expected excess returns or risk premiums on each of the K factors. The intuition behind the arbitrage pricing model is that asset and portfolio returns line up "more or less" with the returns on the factors that govern them, apart perhaps from random noise. We use a model like this to generate credit risk in Chapters 6 and 8

From a risk measurement standpoint, one limitation of the CAPM and arbitrage pricing models is that they explain return in terms of a "snapshot" of factors, while in practice returns may be generated by varying a portfolio over time. Gamma trading, which we discuss in Chapters 5 and 13, is the classic example of such a *dynamic strategy*. It is also very hard in practice to distinguish genuine alpha from harvesting of risk premiums.

2.4.2 Diversification

Diversification is reduction in risk from owning several different assets. We will explore diversification further in Chapter 13, when we discuss portfolio market and credit risk. For now, let's get a sense of the interaction between risk, volatility, and correlation in the simple model in which returns are jointly normally distributed.

It's easier at this point to work with the *weights* rather than the *number of units* of the assets. The market value of each position is $x_n S_n$. Each asset has a portfolio weight, related to the market values by

$$\omega_n = \frac{x_n S_n}{\sum_{n=1}^{N} x_n S_n} \qquad n = 1, \ldots, N$$

The volatility of a portfolio depends on the volatilities of the components and their correlations among one another. If a portfolio contains N assets, its volatility is

$$\sigma_p = \sum_{n=1}^{N} \omega_n^2 \sigma_n^2 + \sum_{n=1}^{N} \sum_{m=1}^{N} \omega_n \omega_m \sigma_n \sigma_m \rho_{nm}$$

Suppose we have a portfolio consisting of long positions in two assets. The assets themselves can also be portfolios in their own right, as long as we have good estimates—or at least estimates for which we are prepared to suspend disbelief—of their expected return, volatilities, and correlation. For concreteness, let's imagine the two assets are well-diversified portfolios of equities (asset 1) and bonds (asset 2).

Assume there is a fixed dollar amount that can be invested in the two assets, and that we can only take long positions in the portfolios, so it is unambiguous to speak of portfolio weights based on market values. Denote the expected returns by μ_n, the return volatilities by σ_n, $n = 1, 2$, the asset 1 weight by ω, and the returns correlation between the two assets by ρ. The portfolio return is

$$\mu_p = \omega \mu_1 + (1 - \omega) \mu_2$$

and the portfolio volatility is

$$\sigma_p = \sqrt{\omega^2 \sigma_1^2 + (1 - w)^2 \sigma_2^2 + 2\omega(1 - \omega) \sigma_1 \sigma_2 \rho}$$

In this simple setup with long positions only, there is (almost) always a diversification effect which reduces risk. Figure 2.10 illustrates the impact of the asset return volatilities and correlation and the allocation among the two assets on the diversification benefit:

- If correlation is +1 (perfect correlation), any portfolio that is long one or both assets has same risk. The diversification effect is zero, but you also don't add risk as the weight changes.

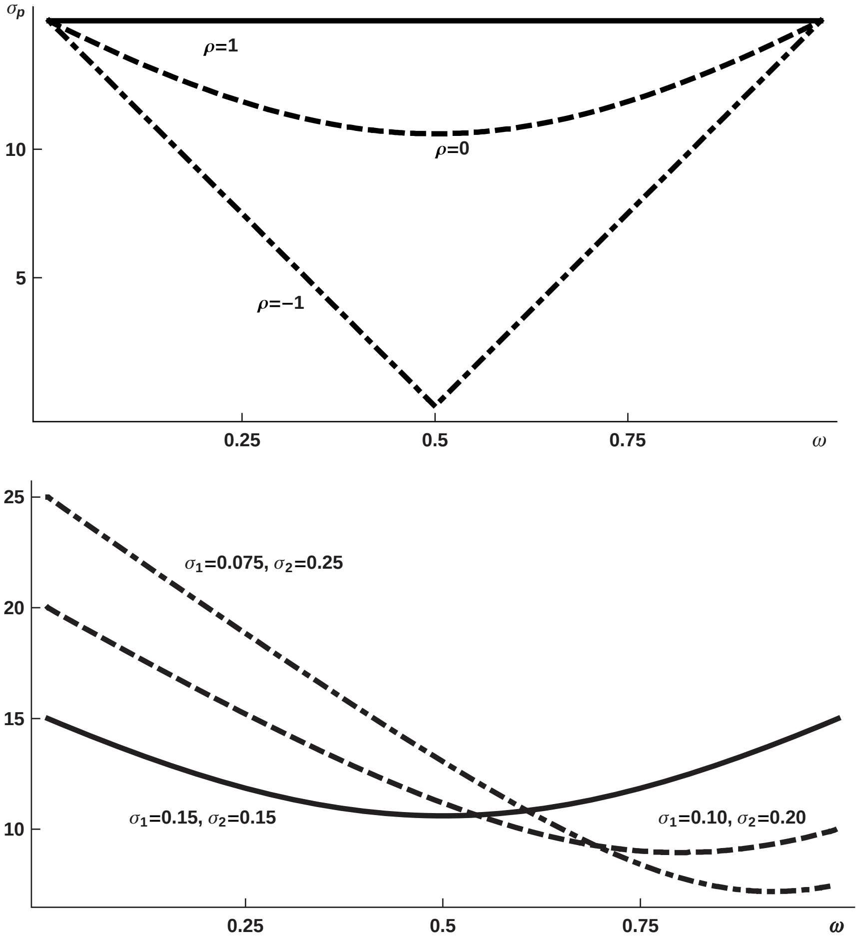

FIGURE 2.10 Diversification, Volatility, and Correlation
Upper panel: Volatility of a portfolio consisting of long positions in two assets with identical volatilities. Perfect correlation: any long-long portfolio has same risk. Zero correlation: one asset might have large loss while other does not. Diversification benefit in terms of vol. Greatest diversification benefit from negative (positive) correlation.
Lower panel: Volatility of a portfolio consisting of long positions in two assets with different volatilities and zero correlation. If the two volatilities are very different, the opportunity to diversify is more limited.

- If correlation is 0, one asset might have a loss while other does not. There is some diversification effect, unless one of the allocations and/or one of the volatilities is very large compared to the other.
- If correlation is –1 and the weights are the same ($\omega = 0.5$), the positions are a perfect offset, and portfolio risk is zero.
- If correlation is somewhere in between, there is a diversification benefit, but it may be small. If correlation is near zero then the diversification benefit will be small if the volatilities are very close to one another.

2.4.3 Efficiency

Not every diversified portfolio is *efficient*. An efficient portfolio is one in which the investments are combined in the right amounts, so that you cannot get extra return from a different combination without also getting some extra volatility. We distinguish between efficient portfolios and *minimum-variance portfolios*, which are portfolios with the lowest possible volatility for a given portfolio return. There generally are several minimum-variance portfolios with the same volatility, one of which has a higher portfolio return than the others. Only the latter is efficient. Figure 2.11 plots minimum-variance and efficient portfolios constructed from stocks and bonds for varying correlations between the two asset classes.

It is easy to plot these curves for just two assets; every combination of the two has minimum variance. If there is a third asset, things are more complicated. Substituting a high-volatility asset with a low or negative correlation to the other two, for a lower-volatility asset with a high correlation to the third asset, might reduce the portfolio volatility. So for more than two assets, mathematical programming techniques are required to plot the minimum-variance curve.

But once we identify any two minimum-variance portfolios, we "revert" to the two-asset case: We can plot the entire minimum-variance locus from just these two. So long as the stock and bond portfolios in our example are themselves minimum-variance portfolios, that is, well-diversified portfolios that can't be reshuffled to provide a higher return without also increasing volatility, the example is a good representation of the minimum-variance locus for the entire universe of stocks and bonds.

The efficient portfolios are the ones to the north and right of the *global minimum-variance portfolio*, the one combination of the assets that has the lowest volatility of all. Figure 2.10 displays an example of the efficient

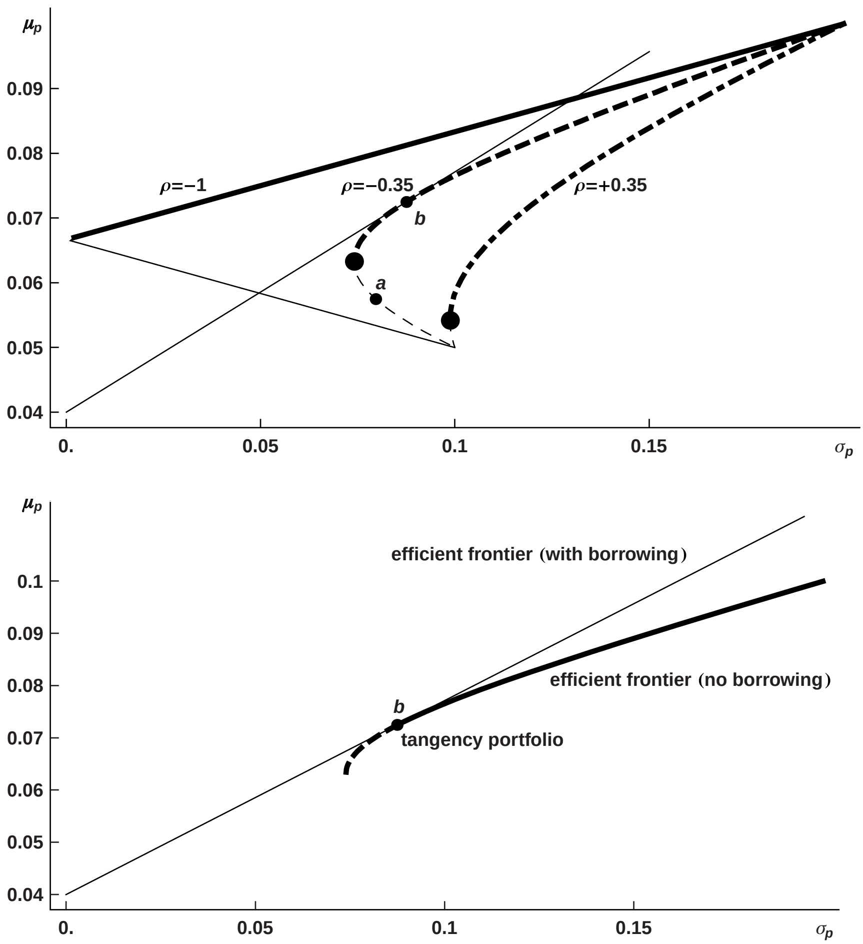

FIGURE 2.11 Minimum-Variance and Efficient Portfolios
Efficient frontier for portfolios that can be created from long positions in two assets, for different correlations between the two assets.
Upper panel: Each curve shows the entire locus of minimum-variance portfolios. The global minimum-variance portfolios are marked by black dots. The thick part of each curve shows the efficient minimum-variance portfolios.
Lower panel: Efficient frontier with a risk-free rate of 0.04 and correlation −0.35.

frontier for portfolios that can be created from long positions in two assets (or portfolios) with expected returns and volatilities given by

	Return	Volatility
Stock portfolio	0.10	0.10
Bond portfolio	0.05	0.20

and correlations as noted in the graphs. The global minimum-variance portfolios correspond to the minima of the plots in the upper panel of Figure 2.10 and are marked with large dots in the upper panel of Figure 2.11. For a given covariance matrix of returns, the asset weights at this unique point don't depend on expected returns, though the resulting expected return does.

How can a portfolio be minimum-variance but inefficient? The dot marked "a" identifies one. It contains more bonds and less stocks than its efficient twin directly north of it on the same minimum-variance locus. Since bonds have a lower expected return, the portfolio has sacrificed return for no reward in lower volatility.

The upper panel of Figure 2.11 shows only the return-volatility combinations available by mixing long positions in stocks and bonds. If we are allowed to go short bonds, that is, borrow bonds, sell what we borrow, and use the proceeds to buy more stocks, we can get higher returns and higher volatility. The locus of available combinations with shorting would be the continuation of the hyperbola plots in the upper panel of Figure 2.10, northeast of the point at which the allocation to stocks is 100 percent. With shorting permitted, extremely high expected returns can be obtained, albeit with extremely high volatility. In addition, the ratio of return to volatility deteriorates at higher returns.

The availability of a risk-free rate changes the problem of finding efficient portfolios. The new efficient frontier is now no longer the upper arm of the hyperbola, but rather the straight line through the point r_f on the y-axis that is tangent to the locus of efficient portfolios without a risk-free rate. As in the case in which there is no borrowing at the risk-free rate, every efficient portfolio is a combination of two other "portfolios," the risk-free investment and the combination of stocks and bonds defined by the tangency point.

The efficient frontier with a risk-free investment, for a stock-bond return correlation of –0.35, is displayed as a thin solid line in the upper panel of Figure 2.11. The lower panel of Figure 2.11 focuses on the case of a stock-bond return correlation of –0.35. Again, the locus of the efficient frontier depends on whether it is possible to short. If shorting is excluded, the efficient frontier is the straight line segment connecting the point $(0, r_f)$ to the tangency point and the segment of the upper arm of the hyperbola between the tangency point and the point at which the entire allocation is to stocks. If shorting is possible, the efficient frontier is the ray from $(0, r_f)$ through the tangency portfolio.

Expected returns on individual assets in an efficient portfolio have this relationship to one another:

$$\frac{\mu_n - r_f}{\sigma_{np}} = \frac{\mu_p - r_f}{\sigma_p^2} \qquad n = 1, \ldots, N$$

It states that the ratio of expected excess return of each asset to its return covariance with the portfolio is equal to the portfolio's excess return ratio to its variance. The right-hand side expression is the *Sharpe ratio* $\frac{\mu_p - r_f}{\sigma_p}$ of the portfolio. The tangency portfolio is the unique portfolio for which the Sharpe ratio is maximized.

This relationship can also be defined in terms of beta:

$$\mu_n - r_f = \frac{\sigma_{np}}{\sigma_p^2}(\mu_p - r_f) = \beta_{np}(\mu_p - r_f) \qquad n = 1, \ldots, N \tag{2.3}$$

If the portfolio is a broad index, representing the "market," the CAPM tells us it is not only minimum-variance but efficient (ignoring the Roll critique).

2.5 BENCHMARK INTEREST RATES

The term *risk-free rate* is something of a misnomer: there are no truly riskless returns in the real-world economy. The term is usually used to mean "free of default risk." There are two types of interest rates that are conventionally taken as proxies for a riskless return. Which of these types of risk-free rate is used depends on the context. In any case, we should be aware that these are conventions or benchmarks, used for pricing and analytics, but not to be taken literally as riskless returns.

Government Bond Interest Rates Government bonds are often considered riskless assets. The reason is that the issuer of the bonds, the central government, is also the issuer of the currency of denomination of the bonds, which is also generally the only legal tender.

This analysis, however, leaves out several risks:

- Longer-term government bills and bonds have price risk arising from *fluctuations in interest rates*, even if there is no credit risk. For example, the owner of a three month U.S. T-bill will suffer a capital loss if the three month rate rises. This risk is small for very short-term bills, but most government debt issues, including those of the United States, have initial maturities of at least a month. We discuss market risk due to interest rate volatility in Chapter 4.
- At the point of default, a government may have the alternative of issuing sufficient currency to meet its debt obligations. Government bonds are denominated in nominal currency units, that is, they are not routinely indexed for inflation. Investors therefore face *inflation risk*, the risk that the inflation rate in the currency of issue over the life of the bond will exceed the market expectation that was incorporated in the price of the bond at the time the investment is made.

  Many industrialized countries issue inflation-protected bonds, with coupons and principal payments that are indexed to inflation. These are generally longer-term issues, though, and have considerable interest-rate market risk.
- Foreign investors may face *exchange-rate risk*. The government can guarantee repayment in its own currency units, but cannot eliminate exchange-rate risk or guarantee the value of the bonds in terms of other currencies.
- Government issues may in fact have non-trivial *credit risk*. Bonds issued by most countries have higher yields than similar bonds issued by highly industrialized countries, such as the United States, the United Kingdom, and Germany. And even the latter countries' credit risk carries a positive price, as seen in prices of sovereign CDS. Government bonds are also subject to credit migration risk (see Chapter 6), the risk of an adverse change in credit rating, for example the August 5, 2011 downgrade by Standard and Poors of the U.S. long-term rating.

  Central government debt is more prone to default if it is denominated in foreign currency. If the sovereign entity is also the issuer of the currency in which the bond is denominated, it has the ability to redeem the bonds in freshly issued local currency, expanding the domestic money supply, no matter how large the debt. Whether it chooses to do so depends on the degree of central bank independence and its monetary policy choices.[2] A government might not, however, choose this course of

[2]The U.S. Treasury has issued few foreign currency–denominated liabilities. Almost all of these exceptions related to exchange rate policy and were obligations to other developed-country central banks.

action if redemption at par would entail an extreme inflation outcome. In contrast, U.S. dollar-denominated debt issued by other countries, or U.S. municipal bonds, cannot avoid default by issuing dollars.

There are, however, examples of countries defaulting on domestic currency–denominated obligations. For example, on August 17, 1998, Russia restructured its short-term ruble-denominated bonds in an attempt to avoid a massive expansion of its monetary supply and a devaluation of the ruble. Russia was forced nonetheless on September 2 to devalue.

Money Market Rates Short-term interest rates, those on obligations with a year or less to maturity, are called *money market rates*. Some money market rates serve as benchmarks for risk-free rates. The *Libor*, or London Interbank Offered Rate, curve are the rates paid for borrowing among the most creditworthy banks. There is a distinct Libor curve for each of ten major currencies. Libor rates are proxies for risk-free rates because of the central role the banking system plays. Many investors can obtain financing for their own trading activities at interest rates close to the Libor curve. We study the financing of investment positions in the contemporary financial system in Chapter 12. The British Bankers' Association (BBA) compiles and posts an estimate of the prevailing Libor rates for each currency at 11:00 every morning in London, based on a poll of major banks, making Libor a convenient reference point.

However, Libor rates incorporate a significant degree of credit risk. For major currencies, there is always a substantial spread between these rates and those on government bonds of that currency's issuing authority. For example, the spread between 10-year U.S. dollar swap par rates and the yield on the 10-year U.S. Treasury note is generally on the order of 50 bps and can occasionally be much higher (see Figure 14.15 and the accompanying discussion in Chapter 14).

Another type of money market rate that is also often used as an indicator of the general level of short-term rates is the repo rate, or rates payable on repurchase agreements, introduced in the previous chapter. We discuss repo in more detail in Chapter 12, but for now we can describe it as a short-term rate that is relatively free of credit risk. It, too, is often used as a proxy for risk-free rates.

The spread between Libor and other money market rates that serve as proxies fluctuates, and at times it is extremely unstable, as we see in Chapter 14. This has, for example, been the case during the subprime crisis. Even more than government bond yields, then, the Libor curve is not a risk-free rate, but rather a benchmark for interest rates.

FURTHER READING

Danthine and Donaldson (2005) is a textbook introduction to modern finance theory. Campbell (2000) is a review of the state of play and open questions in asset pricing and return behavior. Cochrane (2005), Dybvig and Ross (2003), and Ross (2005) are clear introductions to the state-price density approach to asset pricing. Drèze (1970) is an early paper identifying the risk-neutral probabilities embedded in equilibrium asset prices.

The first few chapters of Baxter and Rennie (1996) and Chapter 3 of Dixit and Pindyck (1994) are clear and accessible introductions to stochastic processes. Duffie (1996) is more advanced, but covers both the asset pricing and the return behavior material in this chapter, and is worth struggling with. Černý (2004) combines explication of models of return behavior and asset pricing with programming examples.

Elton and Gruber (2003) is a textbook covering the portfolio allocation material presented in this chapter. Huberman and Wang (2008) is an introduction to arbitrage pricing theory. The original presentation of the theory is Ross (1976).

Lo (1999) discusses risk management in the context of decision theory and behavioral economics. Barberis and Thaler (2003) is a survey of behavioral finance.

See Houweling and Vorst (2005) on the Libor curve as a risk-free benchmark.