Part II
Mathematical Background

Chapter 3

Stochastic Processes and Ito's Lemma

THIS CHAPTER and the next provide the mathematical tools—stochastic calculus, dynamic programming, and contingent claims analysis—that will be used throughout the rest of this book. With these tools, we can study investment decisions using a continuous-time approach, which is both intuitively appealing and quite powerful. In addition, the concepts and techniques that we introduce here are becoming widely used in a number of areas of economics and finance, and so are worth learning even apart from their application to investment problems.

This chapter begins with a discussion of stochastic processes. We will begin with simple discrete-time processes, and then turn to the Wiener process (or Brownian motion), an important continuous-time process that is a fundamental building block for many of the models that we will develop in this book. We will explain the meaning and properties of the Wiener process, and show how it can be derived as the continuous limit of a discrete-time random walk. We will then see how the Wiener process can be generalized to a broad class of continuous-time stochastic processes, called Ito processes. Ito processes can be used to represent the dynamics of the value of a project, output prices, input costs, and other variables that evolve stochastically over time and that affect the decision to invest.

As we will see, these processes do not have a time derivative in the conventional sense, and as a result, cannot always be manipulated using the ordinary rules of calculus. To work with these processes, we must make use of Ito's Lemma. This lemma, sometimes called the Fundamental Theorem of stochastic calculus, is an important result that will allow us to differentiate and integrate functions of stochastic processes. We will provide a heuristic derivation of Ito's Lemma and then, through a variety of examples, show how it can be used to perform simple operations on functions of Wiener processes. We will also show how it can be used to derive and solve stochastic differential equations. Next, we will introduce jump processes—processes that make infrequent but discrete jumps, rather than fluctuate continuously—and show how they can be analyzed using a version of Ito's Lemma. Finally, in the Appendix to this chapter we introduce the Kolmogorov equations, which describe the dynamics of the probability density function for a stochastic process, and show how they can be applied.

1 Stochastic Processes

A stochastic process is a variable that evolves over time in a way that is at least in part random. The temperature in downtown Boston is an example; its variation through time is partly deterministic (rising during the day and falling at night, and rising towards summer and falling towards winter), and partly random and unpredictable.[1] The price of IBM stock is another example: it fluctuates randomly, but over the long haul has had a positive expected rate of growth that compensated investors for risk in holding the stock.

Somewhat more formally, a stochastic process is defined by a probability law for the evolution x_t of a variable x over time t. Thus, for given times $t_1 < t_2 < t_3$, etc., we are given, or can calculate, the probability that the corresponding values x_1, x_2, x_3, etc., lie in some specified range, for example

$$\text{prob}(a_1 < x_1 \leq b_1,\ a_2 < x_2 \leq b_2,\ \ldots).$$

When time t_1 arrives and we observe the actual value x_1, we can condition the probability of future events on this information.[2]

[1] One might argue that the randomness is a reflection of the limitations of meteorology, and that in principle it could be eliminated if we could build sufficiently complete and accurate meteorological models. Perhaps, but from an operational point of view, next week's temperature is indeed a random variable.

[2] In this book we will not attempt any detailed or rigorous treatment of stochastic processes, offering instead the minimal explanations and intuitions that suffice for our applications. For detailed and general treatments, see Cox and Miller (1965), Feller (1971), and Karlin and Taylor (1975).

The temperature in Boston and the price of IBM stock are processes that differ in an important respect. The temperature in Boston is a *stationary* process. This means, roughly, that the statistical properties of this variable are constant over long periods of time.[3] For example, although the expected temperature tomorrow may depend in part on today's temperature, the expectation and variance of the temperature on January 1 of next year is largely independent of today's temperature, and is equal to the expectation and variance of the temperature on January 1 two years from now, three years from now, etc. The price of IBM stock, on the other hand, is a *nonstationary* process. The expected value of this price can grow without bound, and, as we will soon see, the variance of price T years from now increases with T.

The temperature in Boston and the price of IBM stock are both *continuous-time* stochastic processes, in the sense that the time index t is a continuous variable. (Even though we might only measure the temperature or stock price at particular points in time, these variables vary continuously through time.) Although we will work mostly with continuous-time processes in this book, it is easiest to begin with some examples of *discrete-time* processes, that is, variables whose values can change only at discrete points in time. Similarly, the set of all logically conceivable values for x_t (often called the states) can be continuous or discrete. Our definition above is general enough to allow all these possibilities.

One of the simplest examples of a stochastic process is the *discrete-time discrete-state random walk*. Here, x_t is a random variable that begins at a known value x_0, and at times $t = 1, 2, 3, \ldots$, takes a jump of size 1 either up or down, each with probability $\frac{1}{2}$. Since the jumps are independent of each other, we can describe the dynamics of x_t with the following equation:

$$x_t = x_{t-1} + \epsilon_t, \tag{1}$$

where ϵ_t is a random variable with probability distribution

$$\text{prob}(\epsilon_t = 1) = \text{prob}(\epsilon_t = -1) = \tfrac{1}{2} \qquad (t = 1, 2, \ldots).$$

We call x_t a discrete-state process because it can only take on discrete values. For example, set $x_0 = 0$. Then for odd values of t, possible values of x_t are $(-t, \ldots, -1, 1, \ldots, t)$, and for even values of t, possible values of x_t are $(-t, \ldots, -2, 0, 2, \ldots, t)$. The probability distribution for x_t is found

[3] This ignores the very long-run possibilities of global warming or cooling.

from the binomial distribution. For t steps, the probability that there are n downward jumps and $t - n$ upward jumps is

$$\binom{t}{n} 2^{-t}.$$

Therefore, the probability that x_t will take on the value $t - 2n$ at time t is

$$\text{prob}(x_t = t - 2n) = \binom{t}{n} 2^{-t}. \tag{2}$$

We will use this probability distribution in the next section when we derive the Wiener process as the continuous limit of the discrete-time random walk. At this point, however, note that the range of possible values that x_t can take on increases with t, as does the variance of x_t. Hence x_t is a nonstationary process.

Because the probability of an upward or downward jump is $\frac{1}{2}$, at time $t = 0$ the expected value of x_t is zero for all t. (Likewise, at time t, the expected value of x_T for $T > t$ is x_t.) One way to generalize this process is by changing the probabilities for an upward or downward jump. Let p be the probability of an upward jump and $q = (1 - p)$ the probability of a downward jump, with $p > q$. Now we have a *random walk with drift*; at time $t = 0$, the expected value of x_t for $t > 0$ is greater than zero, and is increasing with t.

Another way to generalize this process is to let the size of the jump at each time t be a continuous random variable. For example, we might let the size of each jump be normally distributed with mean zero and standard deviation σ. Then, we refer to x_t as a *discrete-time continuous-state stochastic process*.

Another example of a discrete-time continuous-state stochastic process is the *first-order autoregressive process*, abbreviated as AR(1). It is given by the equation

$$x_t = \delta + \rho x_{t-1} + \zeta_t, \tag{3}$$

where δ and ρ are constants, with $-1 < \rho < 1$, and ζ_t is a normally distributed random variable with zero mean. This process is stationary, and x_t has the long-run expected value $\delta/(1 - \rho)$, irrespective of its current value. [This long-run expected value is found by setting $x_t = x_{t-1} = x$ in equation (3) and solving for x.] The AR(1) process is also referred to as a mean-reverting process, because x_t tends to revert back to this long-run expected value. We will examine a continuous-time version of this process later in this chapter.

Both the random walk (with discrete or continuous states, and with drift or without) and the AR(1) process satisfy the *Markov property*, and are therefore called *Markov processes*. This property is that the probability distribution

for x_{t+1} depends only on x_t, and not additionally on what happened before time t. For example, in the case of the simple random walk given by equation (1), if $x_t = 6$, then x_{t+1} can equal 5 or 7, each with probability $\frac{1}{2}$. The values of x_{t-1}, x_{t-2}, etc., are irrelevant once we know x_t. The Markov property is important because it can greatly simplify the analysis of a stochastic process. We will see this shortly as we turn to continuous-time processes.

2 The Wiener Process

A Wiener process—also called a *Brownian motion*—is a continuous-time stochastic process with three important properties.[4] First, it is a *Markov process*. As explained above, this means that the probability distribution for all future values of the process depends only on its current value, and is unaffected by past values of the process or by any other current information. As a result, the current value of the process is all one needs to make a best forecast of its future value. Second, the Wiener process has *independent increments*. This means that the probability distribution for the change in the process over any time interval is independent of any other (nonoverlapping) time interval. Third, changes in the process over any finite interval of time are *normally distributed*, with a variance that increases linearly with the time interval.

The Markov property is particularly important. Again, it implies that only current information is useful for forecasting the future path of the process. Stock prices are often modelled as Markov processes, on the grounds that public information is quickly incorporated in the current price of the stock, so that the past pattern of prices has no forecasting value. (This is called the weak form of market efficiency. If it did not hold, investors could in principle "beat the market" through technical analysis, that is, by using the past pattern of prices to forecast the future.) The fact that a Wiener process has independent increments means that we can think of it as a continuous-time version of a random walk, a point that we will return to below.

The three conditions discussed above—the Markov property, independent increments, and changes that are normally distributed—may seem quite restrictive, and might suggest that there are very few real-world variables

[4]In 1827, the botanist Robert Brown first observed and described the motion of small particles suspended in a liquid, resulting from the apparent successive and random impacts of neighboring particles; hence the term Brownian motion. In 1905, Albert Einstein proposed a mathematical theory of Brownian motion, which was further developed and made more rigorous by Norbert Wiener in 1923.

that can be realistically modelled with Wiener processes. For example, while it probably seems reasonable that stock prices satisfy the Markov property and have independent increments, it is not reasonable to assume that price changes are normally distributed; after all, we know that the price of a stock can never fall below zero. It is more reasonable to assume that changes in stock prices are *lognormally* distributed, that is, that changes in the logarithm of the price are normally distributed.[5] But this just means modelling the logarithm of price as a Wiener process, rather than the price itself. As we will see, through the use of suitable transformations, the Wiener process can be used as a building block to model an extremely broad range of variables that vary continuously (or almost continuously) and stochastically through time.

It is useful to restate the properties of a Wiener process somewhat more formally. If $z(t)$ is a Wiener process, then any change in z, Δz, corresponding to a time interval Δt, satisfies the following conditions:

1. The relationship between Δz and Δt is given by

$$\Delta z = \epsilon_t \sqrt{\Delta t},$$

where ϵ_t is a normally distributed random variable with a mean of zero and a standard deviation of 1.

2. The random variable ϵ_t is serially uncorrelated, that is, $\mathcal{E}[\epsilon_t \epsilon_s] = 0$ for $t \neq s$. Thus the values of Δz for any two different intervals of time are independent. [Thus $z(t)$ follows a Markov process with independent increments.]

Let us examine what these two conditions imply for the change in z over some finite interval of time T. We can break this interval up into n units of length Δt each, with $n = T/\Delta t$. Then the change in z over this interval is given by

$$z(s+T) - z(s) = \sum_{i=1}^{n} \epsilon_i \sqrt{\Delta t}. \tag{4}$$

The ϵ_i's are independent of each other. Therefore we can apply the Central Limit Theorem to their sum, and say that the change $z(s+T) - z(s)$ is normally distributed with mean zero and variance $n\,\Delta t = T$. This last point, which follows from the fact that Δz depends on $\sqrt{\Delta t}$ and not on Δt, is particularly important; *the variance of the change in a Wiener process grows linearly with the time horizon.*

[5] We always use natural logarithms, that is, those with base e.

We will make considerable use of this property later. Also, note that the Wiener process is nonstationary. Over the long run its variance will go to infinity.

By letting Δt become infinitesimally small, we can represent the increment of a Wiener process, dz, in continuous time as

$$dz = \epsilon_t \sqrt{dt}. \tag{5}$$

Since ϵ_t has zero mean and unit standard deviation, $\mathcal{E}(dz) = 0$, and $\mathcal{V}[dz] = \mathcal{E}[(dz)^2] = dt$. Note, however, that a Wiener process has no time derivative in a conventional sense; $\Delta z/\Delta t = \epsilon_t (\Delta t)^{-1/2}$, which becomes infinite as Δt approaches zero.

At times we may want to work with two or more Wiener processes, and we will be interested in their covariances. Suppose that $z_1(t)$ and $z_2(t)$ are Wiener processes. Then we can write $\mathcal{E}(dz_1\, dz_2) = \rho_{12}\, dt$, where ρ_{12} is the *coefficient of correlation* between the two processes. Because a Wiener process has a variance and standard deviation per unit of time equal to 1 ($\mathcal{E}[(dz)^2]/dt = 1$), ρ_{12} is also the covariance per unit of time for the two processes.[6]

2.A Brownian Motion with Drift

We mentioned earlier that the Wiener process can easily be generalized into more complex processes. The simplest generalization of equation (5) is the *Brownian motion with drift*:

$$dx = \alpha\, dt + \sigma\, dz, \tag{6}$$

where dz is the increment of a Wiener process as defined above. In equation (6), α is called the drift parameter, and σ the variance parameter. Note that over any time interval Δt, the change in x, denoted by Δx, is normally distributed, and has expected value $\mathcal{E}(\Delta x) = \alpha\,\Delta t$ and variance $\mathcal{V}(\Delta x) = \sigma^2 \Delta t$.

Figure 3.1 shows three sample paths of equation (6), with trend $\alpha = 0.2$ per year, and standard deviation $\sigma = 1.0$ per year. Although the graph is shown in annual terms (over the time period 1950 to 2000), each sample path was generated by taking a time interval, Δt, of one month, and then calculating a trajectory for $x(t)$ using the equation

$$x_t = x_{t-1} + 0.01667 + 0.2887\,\epsilon_t, \tag{7}$$

[6] Recall that if X and Y are random variables, their coefficient of correlation is $\rho_{XY} = \mathrm{Cov}(XY)/(\sigma_X \sigma_Y)$. In this case $\sigma_X = \sigma_Y = 1$.

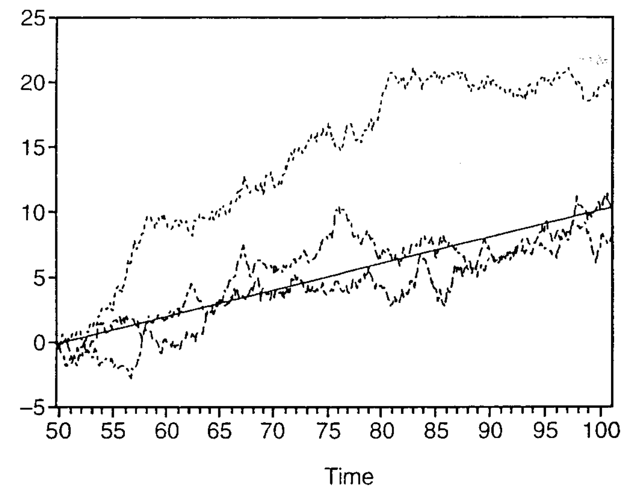

Figure 3.1. Sample Paths of Brownian Motion with Drift

with $x_{1950} = 0$. In equation (7), at each time t, ϵ_t is drawn from a normal distribution with zero mean and unit standard deviation. (Also note that the parameters α and σ have been put in monthly terms. A trend of 0.2 per year implies a trend of 0.0167 per month. A standard deviation of 1.0 per year implies a variance of 1.0 per year, and hence a variance of $\frac{1}{12} = 0.0833$ per month, so that the standard deviation in monthly terms is $\sqrt{0.0833} = 0.2887$.) Also shown is a trend line, that is, equation (7) with $\epsilon_t = 0$.

Figure 3.2 shows an optimal forecast of the same stochastic process. Here, a sample path was generated from 1950 to the end of 1974, again using equation (7), and then forecasts of $x(t)$ were constructed for 1975 to 2000. (For comparison, a realization, that is, continuation of the sample path, was also generated.) Recall that because of the Markov property, *only* the value of $x(t)$ for December 1974 is needed to construct this forecast. The forecasted value of x for a time T months beyond December 1974 is given by

$$\hat{x}_{1974+T} = x_{1974} + 0.01667\,T.$$

The graph also shows a 66-percent forecast confidence interval, that is, the forecasted trajectory for $x(t)$ plus or minus one standard deviation. (A 95-percent confidence interval would be given by the forecasted trajectory plus or minus 1.96 standard deviations.) Recall that since the variance of the

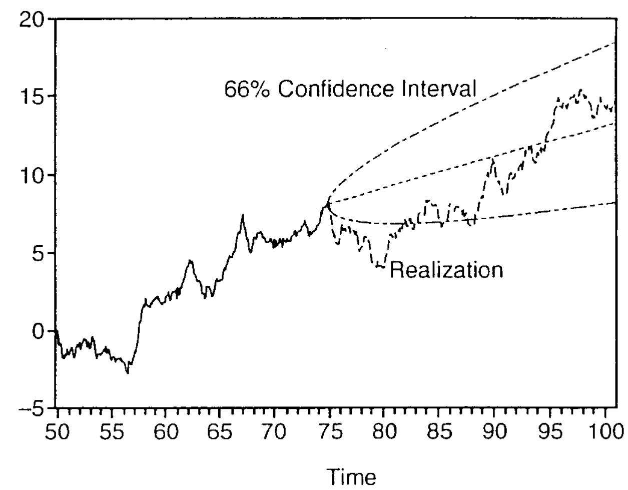

Figure 3.2. Optimal Forecast of Brownian Motion with Drift

Wiener process grows linearly with the time horizon, the standard deviation grows as the *square root* of the time horizon. Hence the 66-percent confidence interval for a forecast T months ahead is given by

$$x_{1974} + 0.01667T \pm 0.2887\sqrt{T}.$$

One can similarly construct 90- or 95-percent confidence intervals.

One can also observe from Figures 3.1 and 3.2 that, in the long run, the trend is the dominant determinant of Brownian motion, whereas in the short run, the volatility of the process dominates. Again, this is an implication of the fact that the mean of $(x_t - x_0)$ is $\alpha\, t$, and the standard deviation is $\sigma\sqrt{t}$; for large t, that is, the long run, $\sqrt{t} \ll t$, but for small t, the opposite holds. Another way to see this is to consider the probability that $x_t < x_0$ when $\alpha > 0$. This probability is very small for large t, but is about $\frac{1}{2}$ for small t.

The Brownian motion with drift is a fairly simple stochastic process, but it is important that it be clearly understood. At this point, the definition of a Wiener process and the properties of its generalization in equation (6) may seem somewhat arbitrary. Why, for example, should dx depend on the square root of dt, and not just on dt? And is it reasonable to expect changes in x over any finite interval to be normally distributed? One way to better motivate equation (6) and its properties is to show how it relates to a random walk in discrete time. We turn to this next.

2.B Random Walk Representation of Brownian Motion

Here we show how equation (6) can be derived as the continuous limit of a discrete-time random walk.[7] To do this, we will divide time up into discrete periods of length Δt, and we will assume that in each period the variable x either moves up or down by an amount Δh. Let the probability that it moves up be p, and the probability that it moves down be $q = 1 - p$. Figure 3.3 shows the possible values of x in each of three periods, assuming it begins at the point x_0. For each possible combination of t and x, the probability of it being reached is also shown. Note that from each period to the next, Δx is a random variable that can take on the values $\pm\Delta h$. Also note that x follows a Markov process with independent increments—the probability distribution for its future value depends only on where it is now, and the probability that it will move up or down in each period is independent of what happened in previous periods.

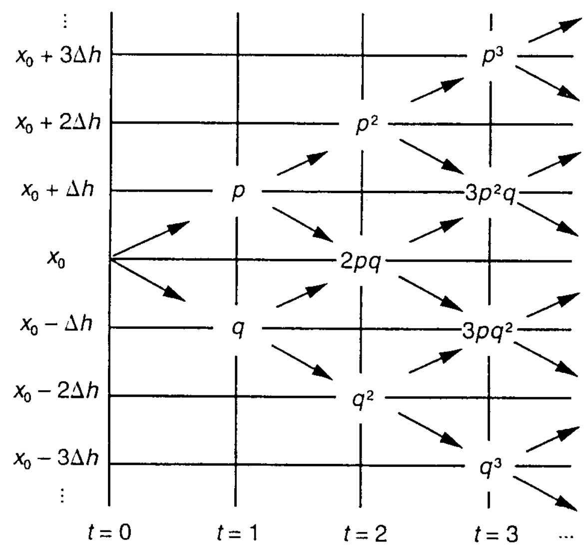

Figure 3.3. Random Walk Representation of Brownian Motion

[7] This approach is developed in detail by Cox and Miller (1965). Our exposition is based on Dixit (1993a).

Let us examine the distribution for future values of x. First, observe that the mean of Δx is $\mathcal{E}[\Delta x] = (p - q)\,\Delta h$. Also,

$$\mathcal{E}[(\Delta x)^2] = p\,(\Delta h)^2 + q\,(-\Delta h)^2 = (\Delta h)^2.$$

Thus the variance of Δx is

$$\mathcal{V}[\Delta x] = \mathcal{E}[(\Delta x)^2] - (\mathcal{E}[\Delta x]^2) = [1 - (p-q)^2](\Delta h)^2 = 4pq(\Delta h)^2. \qquad (8)$$

A time interval of length t has $n = t/\Delta t$ discrete steps. Since the successive steps of the random walk are independent, the cumulated change $(x_t - x_0)$ is a binomial random variable with mean

$$n\,(p-q)\,\Delta h = t\,(p-q)\,\Delta h/\Delta t,$$

and variance

$$n\,[1 - (p-q)^2]\,(\Delta h)^2 = 4pq\,t\,(\Delta h)^2/\Delta t.$$

To interpret this, consider a series of n independent trials, where a success in any one trial counts as 1 and occurs with probability p, while a failure counts as 0 and occurs with probability $q = 1 - p$. The number of successes in n independent trials has a binomial distribution with mean np and variance npq; see Feller (1968, pp. 223, 228). The expressions above are analogous. Now a success counts as Δh and a failure as $-\Delta h$, so the variance, for example, is $4(\Delta h)^2$ times that of the usual binomial expression.

So far the probabilities p and q and the increments Δh and Δt have been chosen arbitrarily, and shortly we will want to let Δt go to zero. As it does, we would like the mean and variance of $(x_t - x_0)$ to remain unchanged, and to be independent of the particular choice of p, q, Δh, and Δt. In addition, we would like to reach equation (6) in the limit. We can ensure that this will indeed be the case by setting

$$\Delta h = \sigma\sqrt{\Delta t}, \qquad (9)$$

and

$$p = \tfrac{1}{2}\left[1 + \frac{\alpha}{\sigma}\sqrt{\Delta t}\right], \qquad q = \tfrac{1}{2}\left[1 - \frac{\alpha}{\sigma}\sqrt{\Delta t}\right]. \qquad (10)$$

Then

$$p - q = \frac{\alpha}{\sigma}\,\sqrt{\Delta t} = \frac{\alpha}{\sigma^2}\,\Delta h.$$

Substitute these expressions for Δh and $p - q$ into the formulas above for the mean and variance of $(x_t - x_0)$, and let Δt approach zero. For any finite

t, the number of steps, n, then goes to infinity, and the binomial distribution converges to a normal distribution, with mean

$$t\,\frac{\alpha}{\sigma^2}\,\Delta h\,\frac{\Delta h}{\Delta t} = \alpha t$$

and variance

$$t\left[1-\left(\frac{\alpha}{\sigma}\right)^2 \Delta t\right]\frac{\sigma^2 \Delta t}{\Delta t} \rightarrow \sigma^2 t.$$

These are exactly the values we need for Brownian motion; α is the drift, and σ^2 the variance, per unit time. In the limit as $\Delta t \rightarrow 0$, both the mean and variance of $(x_t - x_0)$ are independent of Δh and Δt.

We see, then, that Brownian motion is the limit of a random walk, when the time interval and step length go to zero together while preserving the relationship of equation (9). Furthermore, this relationship between Δh and Δt is not an arbitrary one; it is the only way to make the variance of $(x_t - x_0)$ depend on t and not on the number of steps. Thus it should now be clear why dx in equation (6) depends (via dz) on the square root of dt and not on dt. And it should also be clear why changes in x over finite periods are normally distributed; as the number of steps becomes very large, the binomial distribution approaches a normal distribution.

An interesting property of Brownian motion is that as $\Delta t \rightarrow 0$, the total distance travelled over any finite interval of time becomes infinite. This follows from the relationship between Δh and Δt. Since $|\Delta x| = \Delta h$ with certainty, $\mathcal{E}(|\Delta x|) = \Delta h$. Then the total expected length of the path over a time interval of length t is

$$n\,\Delta h = t\,\frac{\Delta h}{\Delta t} = \frac{t\sigma}{\sqrt{\Delta t}},$$

which goes to infinity as Δt goes to zero. Likewise, $\Delta x/\Delta t \rightarrow \pm\infty$, depending on whether $\Delta x = +\Delta h$ or $-\Delta h$. Thus the sample paths of a Brownian motion must have many ups and downs, and look very jagged, and such paths will not be differentiable. The derivative dx/dt will not exist, and we cannot speak of $\mathcal{E}(dx/dt)$. However, $\mathcal{E}[dx]$ will in general exist, and so will $(1/dt)\,\mathcal{E}[dx]$.

3 Generalized Brownian Motion—Ito Processes

The Wiener process can serve as a building block to model a broad range of stochastic variables. We will examine a number of examples, all of which are

special cases of the following generalization of the simple Brownian motion with drift that we studied in the previous section:

$$dx = a(x,t)\,dt + b(x,t)\,dz, \tag{11}$$

where, again, dz is the increment of a Wiener process, and $a(x,t)$ and $b(x,t)$ are known (nonrandom) functions. The new feature is that the drift and variance coefficients are functions of the current state and time. The continuous-time stochastic process $x(t)$ represented by equation (11) is called an *Ito process*.

Consider the mean and variance of the increments of this process. Since $\mathcal{E}(dz) = 0$, $\mathcal{E}(dx) = a(x,t)\,dt$. The variance of dx is equal to $\mathcal{E}[dx^2] - (\mathcal{E}[dx]^2)$, which contains terms in dt, in $(dt)^2$, and in $(dt)(dz)$, which is of order $(dt)^{3/2}$. For dt infinitesimally small, terms in $(dt)^2$ and $(dt)^{3/2}$ can be ignored, and to order dt the variance is

$$\mathcal{V}[dx] = b^2(x,t)\,dt.$$

We refer to $a(x,t)$ as the expected instantaneous *drift rate* of the Ito process, and to $b^2(x,t)$ as the instantaneous *variance rate*.

3.A Geometric Brownian Motion

An important special case of equation (11) is the *geometric Brownian motion with drift*. Here, $a(x,t) = \alpha x$, and $b(x,t) = \sigma x$, where α and σ are constants. In this case equation (11) becomes

$$dx = \alpha\,x\,dt + \sigma\,x\,dz. \tag{12}$$

From our discussion of the simple Brownian motion of equation (6), we know that percentage changes in x, $\Delta x/x$, are normally distributed. Since these are changes in the natural logarithm of x, absolute changes in x, Δx, are *lognormally* distributed.

The relation between x and its logarithm is somewhat more complicated in this context. In the next section we will show that if $x(t)$ is given by equation (12), then $F(x) = \log x$ is the following simple Brownian motion with drift:

$$dF = (\alpha - \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dz, \tag{13}$$

so that over a finite time interval t, the change in the logarithm of x is normally distributed with mean $(\alpha - \frac{1}{2}\sigma^2)\,t$ and variance $\sigma^2 t$. As for x itself, it can be shown that if currently $x(0) = x_0$, the expected value of $x(t)$ is given by

$$\mathcal{E}[x(t)] = x_0\,e^{\alpha t}.$$

and the variance of $x(t)$ is given by[8]

$$\mathcal{V}[x(t)] = x_0^2\, e^{2\alpha t}\, (e^{\sigma^2 t} - 1).$$

This result for the expectation of a geometric Brownian motion can be used to calculate the expected present discounted value of $x(t)$ over some period of time. For example, note that

$$\mathcal{E}\left[\int_0^\infty x(t)\, e^{-rt}\, dt\right] = \int_0^\infty x_0\, e^{-(r-\alpha)t}\, dt = x_0/(r-\alpha), \tag{14}$$

provided the discount rate r exceeds the growth rate α. This will prove useful in later chapters when we will need to calculate the discounted present value of a profit flow that follows a geometric Brownian motion.

Geometric Brownian motion is frequently used to model securities prices, as well as interest rates, wage rates, output prices, and other economic and financial variables. Figure 3.4 shows three sample paths of equation (12), with a drift rate of $\alpha = 0.09$, that is, 9 percent per year, and $\sigma = 0.2$, that is, 20 percent per year. These particular numbers were chosen because they are approximately equal to the annual expected rate of growth and standard deviation of the New York Stock Exchange Index, expressed in real (constant-dollar) terms. As with Figure 3.1, the sample paths were generated by taking a time interval, Δt, of one month. Then, $x(t)$ is calculated using the equation

$$x_t = 1.0075\, x_{t-1} + 0.0577\, x_{t-1}\, \epsilon_t, \tag{15}$$

with $x_{1950} = 100$. (Again, at each time t, ϵ_t is drawn from a normal distribution with zero mean and unit standard deviation.) Also shown is the trend line, that is, equation (15) with $\epsilon_t = 0$. Note that in one of these sample paths the "stock market" outperformed its expected rate of growth, but in the other two sample paths it clearly underperformed.

Figure 3.5 shows an optimal forecast of this process. As before, a sample path was generated from 1950 to the end of 1974, and then forecasts of $x(t)$ were constructed for 1975 to 2000. For comparison, a realization, that is, a continuation of the sample path, is also shown. Once again, because of the Markov property, only the value of $x(t)$ for December 1974 is needed to construct the forecast. The forecasted value of x is given by

$$\hat{x}_{1974+T} = (1.0075)^T\, x_{1974},$$

[8]See Aitchison and Brown (1957) for a detailed discussion of the lognormal distribution and its properties.

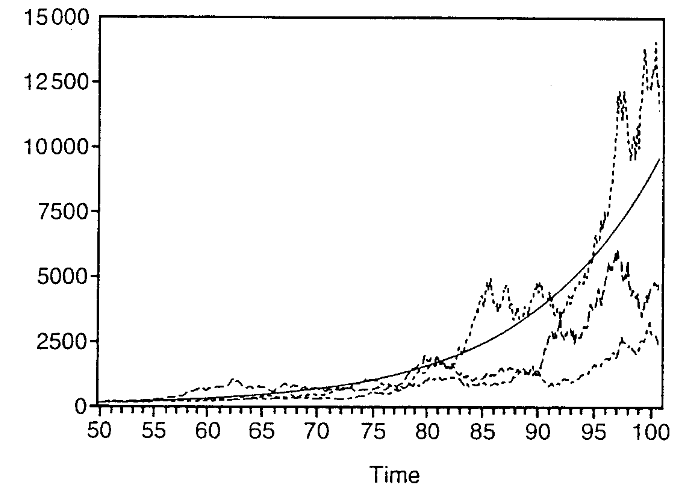

Figure 3.4. *Sample Paths of Geometric Brownian Motion*

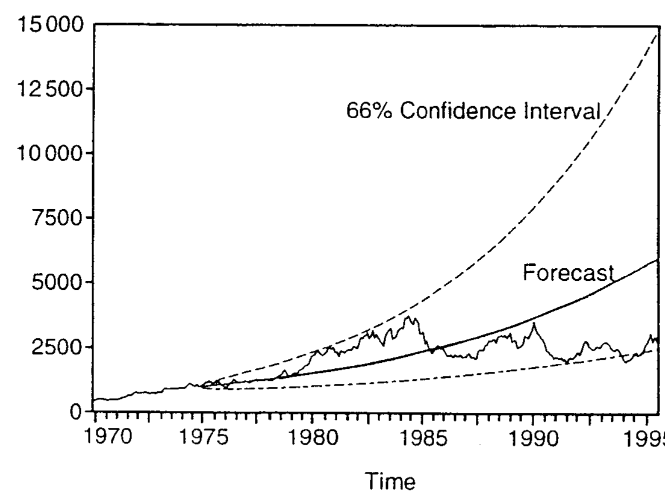

Figure 3.5. *Optimal Forecast of Geometric Brownian Motion*

where T is measured in months starting in January 1975. Also shown in the figure is a 66-percent forecast confidence interval. Since the standard deviation of percentage changes in x grows with the square root of the time horizon, the upper and lower bounds of this confidence interval are given by

$$(1.0075)^T (1.0577)^{\sqrt{T}} x_{1974} \quad \text{and} \quad (1.0075)^T (1.0577)^{-\sqrt{T}} x_{1974}.$$

Note that this confidence interval becomes quite wide. In this particular realization, the "stock market" underperformed its forecast.

3.B Mean-Reverting Processes

As the sample paths in Figures 3.1 and 3.4 illustrate, Brownian motions tend to wander far from their starting points. This is realistic for some economic variables—for example, speculative asset prices—but not for others. Consider, for example, the prices of raw commodities such as copper or oil. Although such prices are often modelled as geometric Brownian motions, one could argue that they should somehow be related to long-run marginal production costs. In other words, while in the short run the price of oil might fluctuate randomly up and down (in response to wars or revolutions in oil-producing countries, or in response to the strengthening or weakening of the OPEC cartel), in the longer run it ought to be drawn back towards the marginal cost of producing oil. Thus one might argue that the price of oil should be modelled as a *mean-reverting process*.

The simplest mean-reverting process—also known as an *Ornstein-Uhlenbeck process*—is the following:

$$dx = \eta\,(\bar{x} - x)\,dt + \sigma\,dz. \tag{16}$$

Here, η is the speed of reversion, and $\bar{x}$ is the "normal" level of x, that is, the level to which x tends to revert. (If x is a commodity price, then $\bar{x}$ might be the long-run marginal cost of production of this commodity.) Note that the expected change in x depends on the difference between x and $\bar{x}$. If x is greater (less) than $\bar{x}$, it is more likely to fall (rise) over the next short interval of time. Hence this process, although satisfying the Markov property, does not have independent increments.

If the value of x is currently x_0 and x follows equation (16), then its expected value at any future time t is

$$\mathcal{E}[x_t] = \bar{x} + (x_0 - \bar{x})\,e^{-\eta t}. \tag{17}$$

Also, the variance of $(x_t - \bar{x})$ is

$$\mathcal{V}[x_t - \bar{x}] = \frac{\sigma^2}{2\eta}\,(1 - e^{-2\eta t}). \tag{18}$$

[See the Appendix to this chapter for a derivation of equations (17) and (18).] Observe from these equations that the expected value of x_t converges to $\bar{x}$ as t becomes large, and the variance converges to $\sigma^2/2\eta$. Also, as $\eta \to \infty$, $\mathcal{V}[x_t] \to 0$, which means that x can never deviate from $\bar{x}$, even momentarily. Finally, as $\eta \to 0$, x becomes a simple Brownian motion, and $\mathcal{V}[x_t] \to \sigma^2 t$.

Figure 3.6 shows four sample paths of equation (16) for different values of η. In each case, $\sigma = 0.05$ in *monthly* terms, $\bar{x} = 1$, and $x(t)$ begins at $x_0 = 1$. The first is for $\eta = 0$, which corresponds to a simple Brownian motion without drift. Note that $x(t)$ tends to drift far from its initial value of 1. Sample paths are also shown for $\eta = 0.01$, 0.02, and 0.5. Note that the larger η is, the less $x(t)$ tends to drift away from $\bar{x}$. When $\eta = 0.5$, $x(t)$ makes only small and short-lived excursions away from $\bar{x}$.

Figure 3.7 shows an optimal forecast of this process, for $\eta = 0.02$. In this case a sample path was generated from 1950 to the end of 1980, and then forecasts of $x(t)$ were constructed for 1981 to 2000. For comparison, a realization, that is, a continuation of the sample path, is also shown, as is a

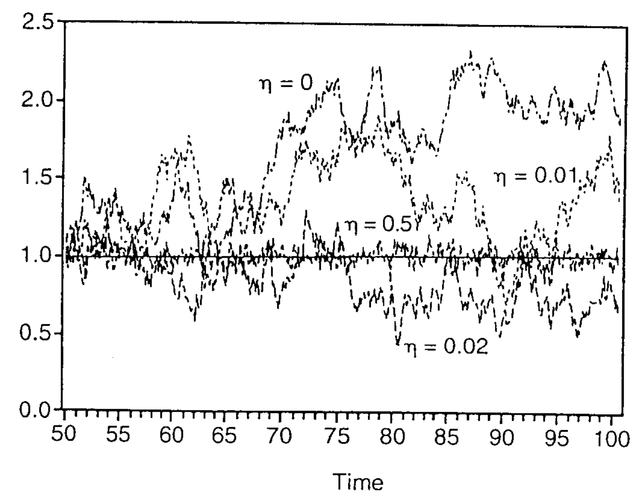

Figure 3.6. Sample Paths of Mean-Reverting Process: $dx = \eta(\bar{x} - x)dt + \sigma dz$

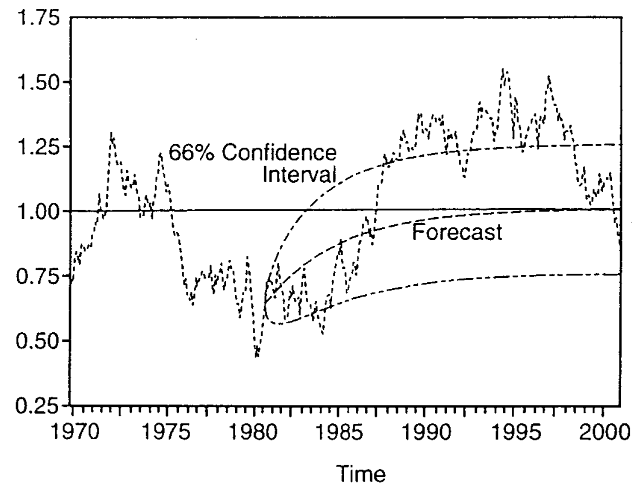

Figure 3.7. Optimal Forecast of Mean-Reverting Process

66-percent confidence interval for the forecast. Note that after four or five years, the variance of the forecast converges to $\sigma^2/2\eta = 0.0025/0.04 = 0.065$, so the 66-percent confidence interval ($\pm$ 1 standard deviation) converges to the forecast ± 0.25.

Equation (16) is the continuous-time version of the first-order autoregressive process in discrete time. Specifically, equation (16) is the limiting case as $\Delta t \to 0$ of the following AR(1) process:

$$x_t - x_{t-1} = \bar{x}\,(1 - e^{-\eta}) + (e^{-\eta} - 1)\,x_{t-1} + \epsilon_t, \tag{19}$$

where ϵ_t is normally distributed with mean zero and standard deviation σ_ϵ, and

$$\sigma_\epsilon^2 = \frac{\sigma^2}{2\eta}\,(1 - e^{-2\eta}).$$

Thus one could estimate the parameters of equation (16) using discrete-time data (the only data ever available) by running the regression

$$x_t - x_{t-1} = a + b\,x_{t-1} + \epsilon_t,$$

and then calculating $\bar{x} = -\hat{a}/\hat{b}$, $\hat{\eta} = -\log(1 + \hat{b})$, and

$$\hat{\sigma} = \hat{\sigma}_\epsilon \sqrt{\frac{\log(1 + \hat{b})}{(1 + \hat{b})^2 - 1}},$$

where $\hat{\sigma}_\epsilon$ is the standard error of the regression.

It is easy to generalize equation (16). For example, one might expect $x(t)$ to revert to $\bar{x}$ as in (16), but the variance rate to grow with x. Then one could use the following process:

$$dx = \eta\,(\bar{x} - x)\,dt + \sigma\,x\,dz. \tag{20}$$

Alternatively, proportional changes in a variable might be modelled as a simple mean-reverting process. This is equivalent to describing $x(t)$ by the process

$$dx = \eta\,x\,(\bar{x} - x)\,dt + \sigma\,x\,dz. \tag{21}$$

We will examine the implications of different mean-reverting processes for investment decisions later in this book.

Before closing this subsection, let us return to a question we asked earlier: are the prices of raw commodities and other goods best modelled as geometric Brownian motions or as mean-reverting processes? One way to answer this is to examine the data for the price variable in question, and in particular to estimate equation (19) and test whether the coefficient of x_{t-1} on the right-hand side is significantly different from zero. There are two problems with this. First, under the null hypothesis that this coefficient is indeed zero (so that x_t follows a random walk), its ordinary least squares estimator is biased towards zero, so one cannot use a standard t-test to determine whether the estimate is significantly different from zero. However, there are alternative tests, called *unit root tests*, that can easily be applied instead.[9] Second and more serious, it usually requires many years of data to determine with any degree of confidence whether a variable is indeed mean reverting.

As an illustration, Figures 3.8 and 3.9 show the prices of crude oil and copper, in constant 1967 dollars, over the past 120 years.[10] A cursory look at these figures suggests that these prices are mean reverting, but that the rate of mean reversion is very slow. This is indeed confirmed by running unit root tests on the data. Running these tests on the full 120 years of data, one can

[9]The tests were originally developed by Dickey and Fuller (1981), and have since been extended and refined. See Chapter 15 of Pindyck and Rubinfeld (1991) for an introduction to these tests.

[10]The data for 1870 to 1973 are from Manthy (1978); data after 1973 are from publications of the U.S. Energy Information Energy and U.S. Bureau of Mines. Prices were deflated by the Wholesale Price Index (now the Producer Price Index).

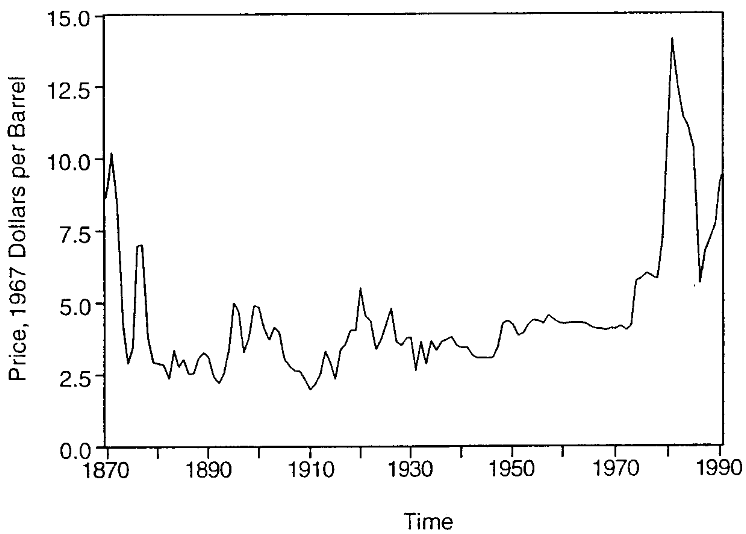

Figure 3.8. *Price of Crude Oil in 1967 Dollars per Barrel*

easily reject the random walk hypothesis; that is, the data confirm that the prices are mean reverting. However, if one performs unit root tests using data for only the past 30 or 40 years, one fails to reject the random walk hypothesis. This seems to be the case for many other economic variables as well; using 30 or so years of data, it is difficult to statistically distinguish between a random walk and a mean-reverting process.

As a result, one must often rely on theoretical considerations (for example, intuition concerning the operation of equilibrating mechanisms) more than statistical tests when deciding whether or not to model a price or other variable as a mean-reverting process. Another criterion that may enter into this modelling decision is analytical tractability. As we will see in later chapters, valuing a project and solving for an optimal investment rule is often much simpler when the underlying stochastic variables are modelled as geometric Brownian motions.[11]

[11] It is usually impossible to obtain analytical solutions for optimal investment rules when the underlying stochastic variables are modelled as mean reverting, so that one must instead use numerical solution techniques. For a strong argument *against* modelling the prices of oil and other exhaustible resources as geometric Brownian motions, see Lund (1991b).

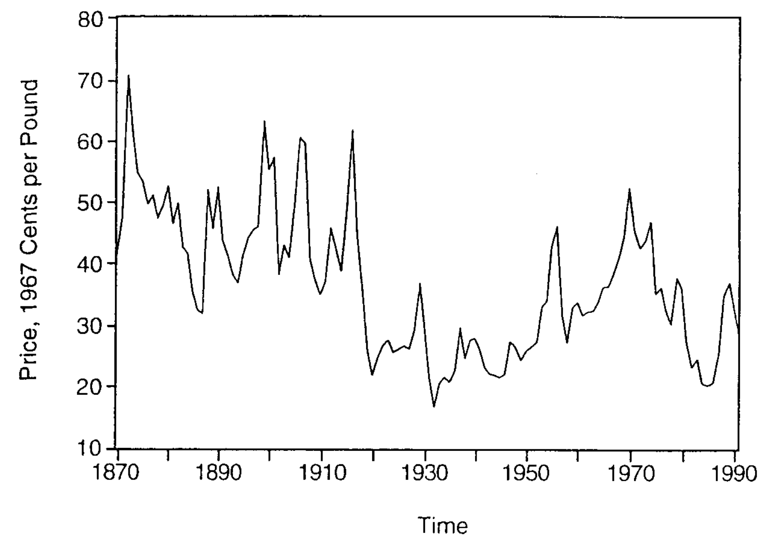

Figure 3.9. *Price of Copper in 1967 Cents per Pound*

4 Ito's Lemma

We have seen that the Ito process of equation (11) is continuous in time, but is not differentiable. However, we will often need to work with functions of Ito processes, and we will want to take the differentials of such functions. For example, we might describe the value of an option to invest in a copper mine as a function of the price of copper, which in turn might be represented by a geometric Brownian motion. In this case we would want to determine the stochastic process that the value of the option follows. To do this, and in general to differentiate or integrate functions of Ito processes, we will need to make use of *Ito's Lemma*.

Ito's Lemma is easiest to understand as a Taylor series expansion. Suppose that $x(t)$ follows the process of equation (11), and consider a function $F(x, t)$ that is at least twice differentiable in x and once in t. We would like to find the total differential of this function, dF. The usual rules of calculus define this differential in terms of first-order changes in x and t:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial t} dt.$$

But suppose that we also include higher-order terms for changes in x:

$$dF = \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial t}\,dt + \tfrac{1}{2}\frac{\partial^2 F}{\partial x^2}\,(dx)^2 + \tfrac{1}{6}\frac{\partial^3 F}{\partial x^3}\,(dx)^3 + \cdots . \qquad (22)$$

In ordinary calculus, these higher-order terms all vanish in the limit. To see whether that is also the case here, expand the third and fourth terms on the right-hand side of equation (22). First, substitute equation (11) for dx to determine $(dx)^2$:

$$(dx)^2 = a^2(x,t)\,(dt)^2 + 2a(x,t)\,b(x,t)\,(dt)^{3/2} + b^2(x,t)\,dt. \qquad (23)$$

Terms in $(dt)^{3/2}$ and $(dt)^2$ go to zero faster than dt as it becomes infinitesimally small, so we can ignore these terms and write

$$(dx)^2 = b^2(x,t)\,dt.$$

As for the fourth term on the right-hand side of equation (22), every term in the expansion of $(dx)^3$ will include dt raised to a power greater than 1, and so will go to zero faster than dt in the limit. This is likewise the case for any higher-order terms, such as $(dx)^4$, etc. Hence Ito's Lemma gives the differential dF as

$$dF = \frac{\partial F}{\partial t}\,dt + \frac{\partial F}{\partial x}\,dx + \tfrac{1}{2}\frac{\partial^2 F}{\partial x^2}\,(dx)^2. \qquad (24)$$

We can also write this in expanded form by substituting equation (11) for dx:

$$dF = \left[\frac{\partial F}{\partial t} + a(x,t)\frac{\partial F}{\partial x} + \tfrac{1}{2}b^2(x,t)\frac{\partial^2 F}{\partial x^2}\right]dt + b(x,t)\frac{\partial F}{\partial x}\,dz. \qquad (25)$$

Compared to the chain rule for differentiation in ordinary calculus, equation (25) has one extra term. For some intuition, suppose for simplicity that the drift rate $a(x,t) = 0$, and that $\partial F/\partial t = 0$. Now $\mathcal{E}(dx) = 0$, but $\mathcal{E}(dF) \neq 0$. This is just an implication of Jensen's Inequality. $\mathcal{E}(dF)$ will be positive if F is a convex function of x (that is, $\partial^2 F/\partial x^2 > 0$), and negative if F is a concave function of x (that is, $\partial^2 F/\partial x^2 < 0$). For Ito processes, dx behaves like $\sqrt{dt}$, and $(dx)^2$ like dt, so the effect of convexity or concavity is of order

dt and cannot be ignored when writing the differential of F. The extra term in equation (25) captures exactly this effect.

We can easily extend this Taylor series expansion to functions of several Ito processes. For example, suppose that $F = F(x_1, \ldots, x_m, t)$ is a function of time and of the m Ito processes $x_1, \ldots, x_m$, where

$$dx_i = a_i(x_1, \ldots, x_m, t)\,dt + b_i(x_1, \ldots, x_m, t)\,dz_i, \qquad i = 1, \ldots, m, \qquad (26)$$

with $\mathcal{E}(dz_i\,dz_j) = \rho_{ij}\,dt$. Then Ito's Lemma gives the differential dF as

$$dF = \frac{\partial F}{\partial t}\,dt + \sum_i \frac{\partial F}{\partial x_i}\,dx_i + \tfrac{1}{2}\sum_i\sum_j \frac{\partial^2 F}{\partial x_i \partial x_j}\,dx_i\,dx_j. \qquad (27)$$

Again, we can substitute (26) for dx_i and write this in expanded form as

$$\begin{aligned} dF = \Bigg[\frac{\partial F}{\partial t} &+ \sum_i a_i(x_1, \ldots, t)\frac{\partial F}{\partial x_i} + \tfrac{1}{2}\sum_i b_i^2(x_1, \ldots, t)\frac{\partial^2 F}{\partial x_i^2} \\ &+ \tfrac{1}{2}\sum_{i \neq j} \rho_{ij}\,b_i(x_1, \ldots, t)\,b_j(x_1, \ldots, t)\frac{\partial^2 F}{\partial x_i \partial x_j}\Bigg]dt \\ &+ \sum_i b_i(x_1, \ldots, t)\frac{\partial F}{\partial x_i}\,dz_i. \end{aligned} \qquad (28)$$

Example: **Geometric Brownian Motion.** Let us return to the geometric Brownian motion of equation (12). We will use Ito's Lemma to show that the process followed by $F(x) = \log x$ is indeed given by equation (13). Since $\partial F/\partial t = 0$, $\partial F/\partial x = 1/x$, and $\partial^2 F/\partial x^2 = -1/x^2$, we have from (24):

$$\begin{aligned} dF &= \frac{1}{x}\,dx - \frac{1}{2x^2}\,(dx)^2 \\ &= \alpha\,dt + \sigma\,dz - \tfrac{1}{2}\sigma^2\,dt = (\alpha - \tfrac{1}{2}\sigma^2)\,dt + \sigma\,dz. \end{aligned} \qquad (29)$$

Hence over any finite time interval T, the change in $\log x$ is normally distributed with mean $(\alpha - \frac{1}{2}\sigma^2)\,T$ and variance $\sigma^2 T$.

Why is the drift rate of $F(x) = \log x$ less than α? The reason is that $\log x$ is a concave function of x, so with x uncertain, the expected value of $\log x$ changes by less than the logarithm of the expected value of x. (Once again this is just a consequence of Jensen's Inequality.) Uncertainty over x is greater the longer the time horizon, so the expected value of $\log x$ is reduced by an amount that increases with time; hence the drift rate is reduced.

***Example:* Correlated Brownian Motions.** As a second example, consider the function $F(x, y) = xy$, where x and y each follow geometric Brownian motions:

$$dx = \alpha_x\, x\, dt + \sigma_x\, x\, dz_x,$$

$$dy = \alpha_y\, y\, dt + \sigma_y\, y\, dz_y,$$

with $\mathcal{E}[dz_x\, dz_y] = \rho\, dt$. We will find the process followed by $F(x, y)$, and the process followed by $G = \log F$.

Since $\partial^2 F/\partial x^2 = \partial^2 F/\partial y^2 = 0$ and $\partial^2 F/\partial x \partial y = 1$, we have from (27):

$$dF = x\, dy + y\, dx + dx\, dy. \tag{30}$$

Now substitute for dx and dy and rearrange:

$$dF = (\alpha_x + \alpha_y + \rho\, \sigma_x\, \sigma_y)\, F\, dt + (\sigma_x\, dz_x + \sigma_y\, dz_y)\, F. \tag{31}$$

Hence F also follows a geometric Brownian motion. What about $G = \log F$? Going through the same steps as in the previous example, we find that

$$dG = (\alpha_x + \alpha_y - \tfrac{1}{2}\sigma_x^2 - \tfrac{1}{2}\sigma_y^2)\, dt + \sigma_x\, dz_x + \sigma_y\, dz_y. \tag{32}$$

From equation (32) we see that over any time interval T, the change in $\log F$ is normally distributed with mean $(\alpha_x + \alpha_y - \frac{1}{2}\sigma_x^2 - \frac{1}{2}\sigma_y^2)\, T$ and variance $(\sigma_x^2 + \sigma_y^2 + 2\rho\, \sigma_x\, \sigma_y)\, T$.

***Example:* Present Discounted Value.** Suppose $F(x) = x^\theta$, where x follows the geometric Brownian motion of equation (12). We will show how to calculate the expected present discounted value

$$\mathcal{E}\left[\int_0^\infty F(x(t))\, e^{-rt}\, dt\right].$$

First, use Ito's Lemma to write

$$\begin{aligned} dF &= \theta\, x^{\theta-1}\, [\alpha\, x\, dt + \sigma\, x\, dz] + \tfrac{1}{2}\, \theta(\theta - 1)\, x^{\theta-2}\, \sigma^2\, x^2\, dt \\ &= [\theta\, \alpha + \tfrac{1}{2}\, \theta(\theta - 1)\, \sigma^2]\, F\, dt + \theta\, \sigma\, F\, dz. \end{aligned} \tag{33}$$

Observe from equation (33) that F follows a geometric Brownian motion. Hence we can use equation (14) for the expectation of a geometric Brownian motion to write

$$\mathcal{E}[F(x, t)] = F(x_0)\, \exp[(\theta\, \alpha + \tfrac{1}{2}\, \theta\, (\theta - 1)\, \sigma^2) t], \tag{34}$$

and the present discounted value is

$$x_0^\theta / [r - \theta\, \alpha - \tfrac{1}{2}\, \theta(\theta - 1)\, \sigma^2],$$

provided the denominator is positive.

5 Barriers and Long-Run Distribution

Suppose x starts at x_0, and follows the simple Brownian motion of equation (6). Left to itself, at time t it will have a normal distribution with mean $(x_0 + \alpha\, t)$ and variance $\sigma^2 t$. However, now we constrain the process of x not to cross an upper reflecting barrier at $\overline{x}$. Using our random walk representation, this means that starting at $\overline{x} - \Delta h$, if x tries to take an upward step it is moved right back to $\overline{x} - \Delta h$. (A step down to $\overline{x} - 2\, \Delta h$ is allowed to occur without interference.) Similarly, we place a lower reflecting barrier at $\underline{x}$.

In economic applications, such barriers often occur because of equilibrating mechanisms in the market. For example, if x is the price of a commodity, it is subject to an upper barrier where new firms enter, and a lower barrier where incumbent firms will exit. We will encounter such models in Chapters 8 and 9.

Here we want to consider what happens when x follows such a process for a long time. The influence of the initial x_0 will disappear when one of the barriers is hit, because of the Markov property. As the motion goes back and forth trapped between the two barriers, we expect it will settle down to a stationary long-run process. We want to find the probability density of this distribution, which we denote by $\phi(x)$.

Once again we use the random walk representation. Consider any three adjacent points, say $x - \Delta h$, x, and $x + \Delta h$. In any one small time interval Δt, the probability mass $\phi(x - \Delta h)$ moves up with probability p, and the mass $\phi(x + \Delta h)$ moves down with probability q, where p and q are given by equation (10). These moves constitute the probability mass at x after the end of the interval. If the probability distribution is to be stationary, this must be just $\phi(x)$. Thus

$$\phi(x) = p\, \phi(x - \Delta h) + q\, \phi(x + \Delta h). \tag{35}$$

Now we substitute for p and q, expand the right-hand side by Taylor's theorem, and collect terms together. This gives

$$\begin{aligned} \phi(x) &= \tfrac{1}{2}\, [1 + (\alpha/\sigma^2)\, \Delta h]\, [\phi(x) - \Delta h\, \phi'(x) + \tfrac{1}{2}\, (\Delta h)^2\, \phi''(x) + \cdots] \\ &\quad + \tfrac{1}{2}\, [1 - (\alpha/\sigma^2)\, \Delta h]\, [\phi(x) + \Delta h\, \phi'(x) + \tfrac{1}{2}\, (\Delta h)^2\, \phi''(x) + \cdots] \\ &= \phi(x) - (\alpha/\sigma^2)\, (\Delta h)^2\, \phi'(x) + \tfrac{1}{2}\, (\Delta h)^2\, \phi''(x) + \cdots . \end{aligned}$$

The omitted terms all go to zero faster than $(\Delta h)^2$ (or Δt). So we cancel $\phi(x)$ from both sides, divide by $(\Delta h)^2$, and take limits as $\Delta h \to 0$. This gives the differential equation

$$\phi''(x) = \gamma\, \phi'(x), \tag{36}$$

where $\gamma = 2\,\alpha/\sigma^2$.

The general solution of this equation is easily seen to be

$$\phi(x) = A\, e^{\gamma x} + B,$$

where A and B are constants to be determined. For that we must consider what happens at the barriers $\bar{x}$ and $\underline{x}$. Consider the upper barrier. The mass of probability $\phi(\bar{x} - \Delta h)$ moves up with probability p, but is reflected right back to $\bar{x} - \Delta h$. Therefore the equation balancing the probability masses is modified to

$$\phi(\bar{x} - \Delta h) = p\,\phi(\bar{x} - \Delta h) + p\,\phi(\bar{x} - 2\,\Delta h).$$

Expanding this and simplifying as before, we find

$$\phi'(\bar{x}) = \gamma\, \phi(\bar{x}).$$

Substituting the general solution into this, we find $B = 0$. The same conclusion could have been obtained by considering the lower barrier $\underline{x}$.

Finally, the constant A must be chosen to ensure that the whole probability mass between $\underline{x}$ and $\bar{x}$ must sum to one. This gives

$$\phi(x) = \gamma\ \exp(\gamma\, x) \,/\, [\exp(\gamma\, \bar{x}) - \exp(\gamma\, \underline{x})]. \tag{37}$$

The long-run stationary probability density is thus a simple exponential. It is natural that if the x process has a positive drift rate ($\alpha > 0$ and therefore $\gamma > 0$), the exponential should be rising toward the upper barrier, and if it has a negative drift rate, the density should be falling to the right. If $\gamma > 0$, we can in fact let the lower barrier $\underline{x}$ go to $-\infty$ and consider a process with only an upper barrier at $\bar{x}$; it will have a long-run stationary distribution whose density falls exponentially to the left of $\bar{x}$. Likewise, if $\gamma < 0$, we can let the upper barrier $\bar{x}$ go to ∞.

We will have occasion to make use of this distribution at several points in Chapters 8 and 9. On a couple of occasions we will generalize the argument to allow a jump process of "sudden death." More substantial extensions include the case where x follows a general Ito process, and examination of the actual dynamics of the probability distribution of x rather than only its long-run stationary state. Both these extensions require the development of a differential equation for the probability density—the Kolmogorov equation—and we treat this in an Appendix to this chapter.

6 Jump Processes

So far we have considered only diffusion processes, that is, stochastic processes that are everywhere continuous. Often, however, it is more realistic to model an economic variable as a process that makes infrequent but discrete jumps. An example would be entry by a new competitor in a market with few firms, so that price suddenly drops. Likewise, one might model the value of a patent as subject to unpredictable but sizable drops in response to competitors' success in developing related patents. Or, one might view the price of oil as a mixed Brownian motion–jump process; during normal times the price fluctuates continuously, but the price can also take large jumps or falls if a war or revolution begins or ends. In this section, we discuss *Poisson (jump) processes*, and we introduce a version of Ito's Lemma that will help us to work with them.

A Poisson process is a process subject to jumps of fixed or random size, for which the arrival times follow a Poisson distribution. We call these jumps "events." Letting λ denote the *mean arrival rate* of an event, during a time interval of infinitesimal length dt, the probability that an event will occur is given by $\lambda\, dt$, and the probability that an event will not occur is given by $1 - \lambda\, dt$. The event is a jump of size u, which can itself be a random variable.

Let q denote a Poisson process by analogy with the Wiener process; in other words

$$dq = \begin{cases} 0 & \text{with probability } 1 - \lambda\, dt, \\ u & \text{with probability } \lambda\, dt. \end{cases}$$

Then we write the stochastic process for the variable x as a Poisson differential equation, which corresponds to the Ito process of equation (11) as follows:

$$dx = f(x,t)\, dt + g(x,t)\, dq, \tag{38}$$

where $f(x,t)$ and $g(x,t)$ are known (nonrandom) functions.

Suppose that $H(x,t)$ is some (differentiable) function of x and t, and that we would like to derive an expression for the expected change in H, that is, $\mathcal{E}(dH)$. To do this, expand dH as follows:

$$\begin{aligned} dH &= \frac{\partial H}{\partial t}\, dt + \frac{\partial H}{\partial x}\, dx \\ &= \frac{\partial H}{\partial t}\, dt + \frac{\partial H}{\partial x}\left[f(x,t)\, dt + g(x,t)\, dq\right]. \end{aligned} \tag{39}$$

(Note that higher-order terms go to zero faster than dt because, unlike for the Ito process, dx does not depend on $\sqrt{dt}$.) Thus changes in x cause changes in H in two ways. First, $H(x,t)$ will change continuously and deterministically in response to the drift in x. Second, there is a possibility that a Poisson event will occur; if it does, x will change by the random amount $u\,g(x,t)$, and $H(x,t)$ will change accordingly. Since the probability that a Poisson event will occur over the interval dt is $\lambda\,dt$, we have

$$\mathcal{E}\left[\frac{\partial H}{\partial x}\,g(x,t)\,dq\right] = \mathcal{E}_u\{\lambda\,[H(x+u\,g(x,t),t) - H(x,t)]\}\,dt, \qquad (40)$$

where the expectation on the right-hand side of the equation is with respect to the size of the jump u. Hence the expectation of the differential of H is given by

$$\mathcal{E}[dH] = \left[\frac{\partial H}{\partial t} + f(x,t)\,\frac{\partial H}{\partial x}\right] dt + \mathcal{E}_u\{\lambda[H(x+g(x,t)\,u,t) - H(x,t)]\}\,dt. \qquad (41)$$

We can use equation (41) in the same way that we used Ito's Lemma when working with continuous processes.

Sometimes we meet a combination of an Ito process and a jump process. The former goes on all the time; the latter occurs infrequently. Then the appropriate version of Ito's Lemma also combines the two effects. Thus, if

$$dx = a(x,t)\,dt + b(x,t)\,dz + g(x,t)\,dq,$$

then the expected value of the change in the function $H(x,t)$ is given by

$$\mathcal{E}[dH] = \left[\frac{\partial H}{\partial t} + a(x,t)\,\frac{\partial H}{\partial x} + \tfrac{1}{2}\,b^2(x,t)\,\frac{\partial^2 H}{\partial x^2}\right] dt + \mathcal{E}_u\{\lambda[H(x+g(x,t)\,u,t) - H(x,t)]\}\,dt. \qquad (42)$$

Note that the second-order derivative is relevant only for the variance contributed by the continuous part of the process. The jump part contributes the last term on the right-hand side involving a difference in values of H at discretely different points.

***Example:* Present Value of Wages.** Suppose an individual who lives forever receives a wage $W(t)$ which rises by a constant amount ϵ at random points in time. If λ is the mean arrival rate of raises, we can write the differential equation for the individual's wage as

$$dW = \epsilon\,dq, \qquad (43)$$

with $u = 1$ with probability 1. What is the present value of the individual's expected earnings stream?

We want to find

$$V(W) = \mathcal{E}\int_0^\infty W(t)e^{-\rho t}\,dt.$$

We can treat V as an asset and equate the normal return on it at rate ρ to the sum of the dividend (current wage) and the expected capital gain:

$$\rho\,V\,dt = W(t)\,dt + \mathcal{E}(dV).$$

In this case, $\mathcal{E}(dV) = (dV/dW)\,\lambda\,\epsilon\,dt = (\lambda\,\epsilon/\rho)\,dt$, so that

$$V(W) = \frac{W}{\rho} + \frac{\lambda\,\epsilon}{\rho^2}.$$

Thus V is equivalent to a perpetuity that pays out forever the current wage W plus the capitalized value of the average raise per unit time.

***Example:* Value of a Machine.** Suppose a machine produces a constant flow of profit π as long as it operates. It requires no maintenance, but at some point in time it will break down and will have to be discarded. If λ is the arrival rate of a breakdown, and ρ is the discount rate, what is the machine's value?

The value of the machine follows the process

$$dV = -V\,dq,$$

where an "event" is $u = 1$ with probability 1. Then the asset return equation becomes

$$\rho\,V\,dt = \pi\,dt + \mathcal{E}(dV) = \pi\,dt - \lambda\,V\,dt.$$

Thus,

$$V = \frac{\pi}{\rho + \lambda}.$$

Hence we can treat the profit flow as a perpetuity, and value it by increasing the discount rate by an amount λ. This is a very general idea: if a profit flow can stop when a Poisson event with arrival rate λ occurs, then we can calculate the expected present value of the stream as if it never stops, but adding λ to the discount rate. We will come across this in many applications in later chapters.

7 Guide to the Literature

Our treatment of stochastic processes and Ito's Lemma has been at an introductory and heuristic level. For a more in-depth development of stochastic processes and their properties, see Cox and Miller (1965), Feller (1971), and Karlin and Taylor (1975, 1981). Cox and Miller (1965) and Karlin and Taylor (1981) provide particularly nice treatments of the Kolmogorov equations. Also, see Chapter 3 of Merton (1990) for a discussion of the use of continuous and jump processes in economic and financial modelling. For a much more rigorous approach to stochastic processes, see Karatzas and Shreve (1988).

For more detailed but still introductory discussions of Ito's Lemma and its application, see Merton (1971), Chow (1979), Malliaris and Brock (1982), and Hull (1989). [Merton (1971) also provides a nice discussion of Poisson processes, with examples.] For more rigorous treatments, see Kushner (1967), Arnold (1974), Dothan (1990), and Chapter 4 of Harrison (1985). Also, for a more detailed discussion of the calculation of expected present values, see Dixit (1993a).

Appendix

A The Kolmogorov Equations

At times we will want to answer questions of the following sort: If $x(t)$ follows a particular stochastic process and its current value is x_0, what is the probability that it will be within a certain range at time t later? Or, what is the probability that $x(t)$ will have reached a point x_1 within a time $t \leq T$? To answer questions like these we will need to describe the probability distribution for x and its evolution over time. This can be done using the *Kolmogorov equations*.

We will derive the Kolmogorov equation for the simple Brownian motion with drift of equation (6), using the discrete-time random walk representation that we introduced in Section 2.B of this chapter. Recall that we broke a time interval of length t into $n = t/\Delta t$ discrete steps, and in each step, with probability p, x would increase by an amount Δh, and with probability $q = 1 - p$, it would decrease by Δh. Finally, to keep the variance of $(x_t - x_0)$ independent of the particular choice of Δt, we set $\Delta h = \sigma \sqrt{\Delta t}$.

Let $\phi(x_0, t_0; x, t)$ be the probability density function for $x(t)$, given that at an earlier time t_0 we have $x(0) = x_0$. Thus

$$\text{Prob}[a \leq x(t) \leq b \mid x(t_0) = x_0] = \int_a^b \phi(x_0, t_0; u, t)\, du.$$

Over the interval of time $t - \Delta t$ to t, the process can reach the point x in one of two ways, by increasing from the point $x - \Delta h$, or by decreasing from the point $x + \Delta h$. Hence

$$\phi(x_0, t_0; x, t) = p\,\phi(x_0, t_0; x - \Delta h, t - \Delta t) + q\,\phi(x_0, t_0; x + \Delta h, t - \Delta t). \quad (44)$$

We recognize this as a dynamic generalization of the stationary-state computation in the text; see equation (35).

Now expand $\phi(x_0, t_0; x - \Delta h, t - \Delta t)$ in a Taylor series around $\phi(x_0, t_0; x, t)$:

$$\phi(x_0, t_0; x - \Delta h, t - \Delta t) = \phi(x_0, t_0; x, t) - \Delta t\,\frac{\partial \phi}{\partial t} - \Delta h\,\frac{\partial \phi}{\partial x} + \tfrac{1}{2}(\Delta h)^2\,\frac{\partial^2 \phi}{\partial x^2} + \cdots.$$

Note that third- and higher-order terms are of order $(\Delta t)^{3/2}$, $(\Delta t)^2$, etc., and hence will go to zero faster than Δt. Expand $\phi(x_0, t_0; x + \Delta h, t - \Delta t)$ likewise, and substitute these expressions into equation (44):

$$\phi(x_0, t_0; x, t) = (p + q)\,\phi(x_0, t_0; x, t) - (p + q)\,\Delta t\,\frac{\partial \phi}{\partial t} - (p - q)\,\Delta h\,\frac{\partial \phi}{\partial x} + \tfrac{1}{2}(p + q)(\Delta h)^2\,\frac{\partial^2 \phi}{\partial x^2}.$$

Finally, we use $p + q = 1$, and from equation (10), $p - q = (\alpha/\sigma)\sqrt{\Delta t}$. We also substitute $\Delta h = \sigma\sqrt{\Delta t}$, divide through by Δt, and rearrange:

$$\tfrac{1}{2}\sigma^2\,\frac{\partial^2}{\partial x^2}\,\phi(x_0, t_0; x, t) - \alpha\,\frac{\partial}{\partial x}\,\phi(x_0, t_0; x, t) = \frac{\partial}{\partial t}\phi(x_0, t_0; x, t). \quad (45)$$

Equation (45) is called the *Kolmogorov forward equation* for the Brownian motion with drift, and it describes the evolution over time of the probability density function $\phi(x_0, t_0; x, t)$. In a similar manner, one can derive the Kolmogorov forward equation for the general Ito process of equation (11):[12]

$$\tfrac{1}{2}\,\frac{\partial^2}{\partial x^2}[b^2(x, t)\,\phi(x_0, t_0; x, t)] - \frac{\partial}{\partial x}[a(x, t)\,\phi(x_0, t_0; x, t)] = \frac{\partial}{\partial t}\phi(x_0, t_0; x, t). \quad (46)$$

Equations (45) and (46) are called "forward" equations because they have as boundary conditions the initial value x_0 at time t_0, and are solved forward for the density function for future values of x. One can likewise describe the evolution of the density function backward in time, that is, taking $x(t)$ at time t as the boundary conditions and solving for the density function for previous

[12]See Karlin and Taylor (1981) for derivations and more detailed discussions of the Kolmogorov forward and backward equations.

values of x_0 at time $t_0 < t$. The *Kolmogorov backward equation* for the Ito process is

$$\tfrac{1}{2}\, b^2(x_0, t_0)\, \frac{\partial^2}{\partial x_0^2}\phi(x_0, t_0; x, t) + a(x_0, t_0)\, \frac{\partial}{\partial x_0}\phi(x_0, t_0; x, t) = -\frac{\partial}{\partial t_0}\phi(x_0, t_0; x, t). \tag{47}$$

We will use the Kolmogorov equations later in this book, but at this point it is useful to consider a couple of examples.

***Example:* Ornstein-Uhlenbeck Process.** In this example we will return to the simple Ornstein-Uhlenbeck (mean-reverting) process of equation (16). For simplicity, set $\bar{x} = 0$, so that the equation becomes

$$dx = -\eta\, x\, dt + \sigma\, dz. \tag{48}$$

In the text we asserted that equations (17) and (18) give the mean and variance of $x(t)$. We can use the Kolmogorov forward equation to prove this claim.

Write the moment-generating function for $x(t)$ as

$$M(\theta, t) \equiv \mathcal{E}(e^{-\theta x}) = \int_{-\infty}^{\infty} \phi(x_0, t_0; x, t)\, e^{-\theta x}\, dx. \tag{49}$$

Then,

$$\frac{\partial M}{\partial t} = \int_{-\infty}^{\infty} \frac{\partial \phi}{\partial t}\, e^{-\theta x}\, dx. \tag{50}$$

The Kolmogorov forward equation for this process is

$$\frac{\partial \phi}{\partial t} = \tfrac{1}{2}\sigma^2 \frac{\partial^2 \phi}{\partial x^2} - \eta\, x\, \frac{\partial \phi}{\partial x} + \eta\, \phi. \tag{51}$$

Substitute this for $\partial\phi/\partial t$ in equation (50), and integrate by parts to get the following equation for $M(\theta, t)$:

$$\tfrac{1}{2}\sigma^2 \theta^2 M - \eta\, \theta\, \frac{\partial M}{\partial \theta} = \frac{\partial M}{\partial t}. \tag{52}$$

This partial differential equation must be solved subject to the boundary conditions:

$$M(0, t) = 1, \quad -M_\theta(0, 0) = x_0, \quad \text{and} \quad \mathcal{V}[x(0)] = M_{\theta\theta}(0, 0) - x_0^2 = 0.$$

The reader can verify that the equation has the following solution:

$$M(\theta, t) = e^{\sigma^2\theta^2/4\eta}\left[1 - x_0\, \theta\, e^{-\eta t} + \left(\tfrac{1}{2}x_0^2 - \frac{\sigma^2}{4\eta}\right)\theta^2 e^{-2\eta t}\right]. \tag{53}$$

Using the fact that $\mathcal{E}(x_t) = -M_\theta(0, t)$ and $\mathcal{E}(x_t^2) = M_{\theta\theta}(0, t)$, the reader can verify equations (17) and (18).

***Example:* The Steady-State Distribution for a Renewable Resource.** The Kolomogorov equations are partial differential equations, and so are usually difficult to solve. However, we are often interested in the long-run steady-state characteristics of stochastic variables. Not all stochastic processes will have probability distributions that converge to some steady-state function (the geometric Brownian motion, for example, will not, but the Ornstein-Uhlenbeck process will). However, if a steady-state distribution exists, in most cases it can be readily found using the Kolmogorov forward equation, which then reduces to an ordinary differential equation.

As a simple first exercise, we begin by rederiving the negative exponential distribution for a long-run stationary Brownian motion between reflecting barriers directly from equation (45). This distribution is independent of the initial x_0, t_0, and the current t, so we get an ordinary differential equation for the function $\phi(x)$:

$$\tfrac{1}{2}\sigma^2 \phi''(x) - \alpha\, \phi'(x) = 0.$$

This is the same as equation (36) in the text, and the solution proceeds as there.

Now consider a more complicated stochastic process, which could be used to describe the evolution of a renewable resource stock, $x(t)$, subject to some rate of harvesting $q(x)$ that might depend on x:

$$dx = [f(x) - q(x)]\, dt + \sigma(x)\, dz. \tag{54}$$

Here, $f(x)$ is the *growth function* for the resource, and is concave, with

$$f(x_{\min}) = f(x_{\max}) = 0, \quad \text{and} \quad f(x) > 0 \quad \text{for} \quad x_{\min} < x < x_{\max}.$$

Much of the literature on renewable resources deals with deterministic versions of equation (54), and compares the socially optimal $q(x)$ with that resulting from a competitive market. However, as biologists and population ecologists have long recognized, most renewable resource stocks evolve stochastically, and (54) is a natural representation.[13] Although $x(t)$ will fluctuate stochastically, it is of interest to study the probability density for x in steady-state equilibrium.

In steady-state equilibrium, ϕ does not depend on x_0, t_0, or t in the forward equation (46). Write the stationary density function as $\phi_\infty(x)$. Then the equation becomes (after integrating once):

$$\tfrac{1}{2}\frac{d}{dx}[\sigma^2(x)\,\phi_\infty(x)] = [f(x) - q(x)]\,\phi_\infty(x).$$

This can be rewritten as

$$\frac{d[\sigma^2(x)\,\phi_\infty(x)]}{\sigma^2(x)\,\phi_\infty(x)} = \frac{2}{\sigma^2(x)}\,[f(x) - q(x)]\,dx. \tag{55}$$

This can be integrated to give the following equation for the steady-state density:[14]

$$\phi_\infty(x) = \frac{m}{\sigma^2(x)}\,\exp\left[2\int^x \frac{f(v) - q(v)}{\sigma^2(v)}\,dv\right], \tag{56}$$

where m is a constant of integration, chosen so that $\int_0^\infty \phi_\infty(x)dx = 1$.

As an example, suppose that $f(x)$ is the logistic function, that is,

$$f(x) = \alpha\, x\,(1 - x/K),$$

where K is the "carrying capacity" of the resource stock. Also, suppose that there is no harvesting, that is, $q(x) = 0$, and that $\sigma(x) = \sigma x$. Then equation (56) yields the following steady-state density that applies when $\sigma^2 < 2\alpha$:

$$\phi_\infty(x) = (2\alpha/\sigma^2 K)^{2\alpha/\sigma^2 - 1}\, x^{2\alpha/\sigma^2 - 2}\, e^{-2\alpha x/\sigma^2 K} \,/\, \Gamma(2\alpha/\sigma^2 - 1), \tag{57}$$

where Γ denotes the gamma function. From this we can determine that the expected value of x in steady-state equilibrium is

$$\mathcal{E}(x_\infty) = K\left(1 - \frac{\sigma^2}{2\alpha}\right). \tag{58}$$

Note that stochastic fluctuations reduce the steady-state expected value of x, and as σ^2 approaches 2α, $\mathcal{E}(x_\infty)$ approaches zero. Also, if $\sigma^2 \geq 2\alpha$, stochastic fluctuations will drive the resource stock to extinction, that is, ϕ_∞ collapses and $x(t) \to 0$ with probability 1. [For a more detailed discussion of this and related models of renewable resources, as well as a derivation of the optimal stochastic harvesting rule $q^*(x)$, see Pindyck (1984).]

[13]See, for example, Beddington and May (1977) and Goel and Richter-Dyn (1974). For a good overview of renewable resource economics in a deterministic context, see Clark (1976).

[14]Merton (1975) also provides a derivation of this equation, and shows how it can be applied to a neoclassical model of growth with a stochastically evolving population.

Chapter 4

Dynamic Optimization under Uncertainty

TIME PLAYS a particularly important role for investment decisions. The payoffs to a firm's investment made today accrue as a stream over the future, and are affected by uncertainty as well as by other decisions that the firm or its rivals will make later. The firm must look ahead to all these developments when making its current decision. As we emphasized in Chapter 2, one aspect of this future is an opportunity to make the same decision later; therefore the option of postponement should be included in today's menu of choices. The mathematical techniques we employ to model investment decisions must be capable of handling all these considerations.

In this chapter we develop two such techniques: dynamic programming and contingent claims analysis. They are in fact closely related to each other, and lead to identical results in many applications. However, they make different assumptions about financial markets, and the discount rates that firms use to value future cash flows.

Dynamic programming is a very general tool for dynamic optimization, and is particularly useful in treating uncertainty. It breaks a whole sequence of decisions into just two components: the immediate decision, and a valuation function that encapsulates the consequences of all subsequent decisions, starting with the position that results from the immediate decision. If the planning horizon is finite, the very last decision at its end has nothing following it, and can therefore be found using standard static optimization methods. This solution then provides the valuation function appropriate to the penultimate