

# 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.<sup>1</sup> 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

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

When time  $t_1$  arrives and we observe the actual value  $x_1$ , we can condition the probability of future events on this information.<sup>2</sup>

<sup>1</sup>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.

<sup>2</sup>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.<sup>3</sup> 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  $t = 1, 2, 3, \dots$ , 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:

$$x_t = x_{t-1} + \epsilon_t, \quad (1)$$

where  $\epsilon_t$  is a random variable with probability distribution

$$\text{prob}(\epsilon_t = 1) = \text{prob}(\epsilon_t = -1) = \frac{1}{2} \quad (t = 1, 2, \dots).$$

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  $(-t, \dots, -1, 1, \dots, t)$ , and for even values of  $t$ , possible values of  $x_t$  are  $(-t, \dots, -2, 0, 2, \dots, t)$ . The probability distribution for  $x_t$  is found

<sup>3</sup>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

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

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  $\sigma$ . 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

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

where  $\delta$  and  $\rho$  are constants, with  $-1 < \rho < 1$ , and  $\zeta_t$  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.<sup>4</sup> 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

<sup>4</sup>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.<sup>5</sup> 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  $z$ ,  $\Delta z$ , corresponding to a time interval  $\Delta t$ , satisfies the following conditions:

1. The relationship between  $\Delta z$  and  $\Delta t$  is given by

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

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

2. The random variable  $\epsilon_t$  is serially uncorrelated, that is,  $\mathcal{E}[\epsilon_t \epsilon_s] = 0$  for  $t \neq s$ . Thus the values of  $\Delta z$  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  $\Delta t$  each, with  $n = T/\Delta t$ . Then the change in  $z$  over this interval is given by

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

The  $\epsilon_i$ '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  $\Delta z$  depends on  $\sqrt{\Delta t}$  and not on  $\Delta t$ , is particularly important; *the variance of the change in a Wiener process grows linearly with the time horizon.*

<sup>5</sup>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  $\Delta t$  become infinitesimally small, we can represent the increment of a Wiener process,  $dz$ , in continuous time as

$$dz = \epsilon_t \sqrt{dt}. \quad (5)$$

Since  $\epsilon_t$  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  $\Delta t$  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  $\rho_{12}$  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 ( $\mathcal{E}[(dz)^2]/dt = 1$ ),  $\rho_{12}$  is also the covariance per unit of time for the two processes.<sup>6</sup>

## 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*:

$$dx = \alpha dt + \sigma dz, \quad (6)$$

where  $dz$  is the increment of a Wiener process as defined above. In equation (6),  $\alpha$  is called the drift parameter, and  $\sigma$  the variance parameter. Note that over any time interval  $\Delta t$ , the change in  $x$ , denoted by  $\Delta x$ , 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,  $\Delta t$ , of one month, and then calculating a trajectory for  $x(t)$  using the equation

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

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

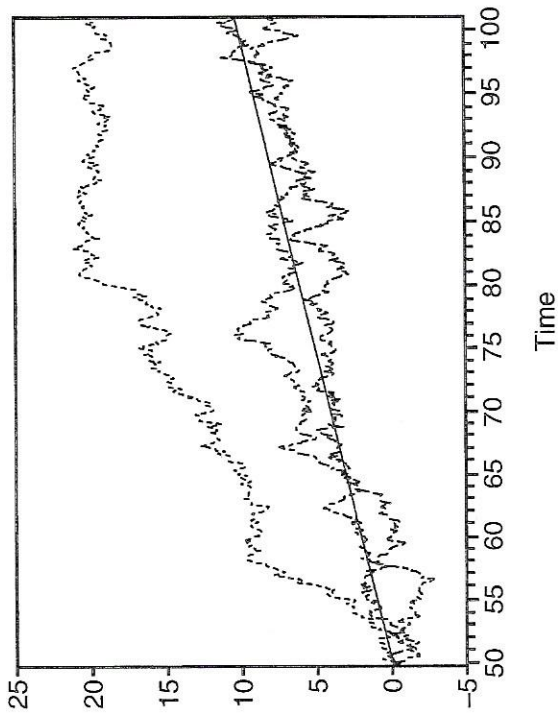


Figure 3.1. Sample Paths of Brownian Motion with Drift

with  $x_{1950} = 0$ . In equation (7), at each time  $t$ ,  $\epsilon_t$  is drawn from a normal distribution with zero mean and unit standard deviation. (Also note that the parameters  $\alpha$  and  $\sigma$  have been put in monthly terms. A trend of 0.2 per year implies a variance 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

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

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

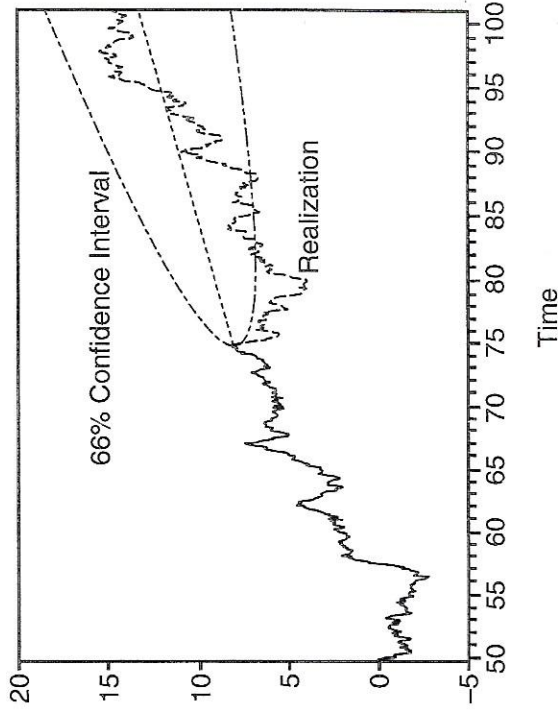


Figure 3.2. Optimal Forecast of Brownian Motion with Drift

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.<sup>7</sup> To do this, we will divide time up into discrete periods of length  $\Delta t$ , and we will assume that in each period the variable  $x$  either moves up or down by an amount  $\Delta h$ . 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,  $\Delta x$  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.

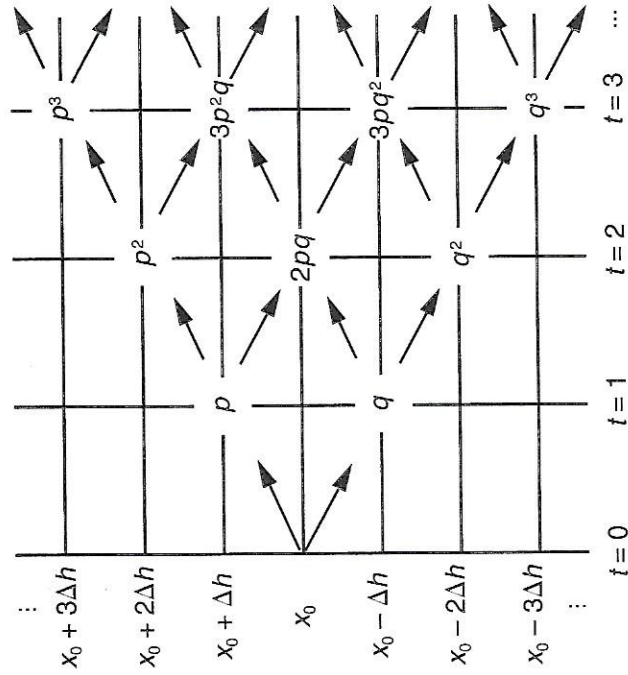


Figure 3.3. Random Walk Representation of Brownian Motion

<sup>7</sup>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  $\Delta x$  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  $\Delta x$  is

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

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  $\Delta h$  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  $\Delta h$  and  $\Delta t$  have been chosen arbitrarily, and shortly we will want to let  $\Delta t$  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  $p, q, \Delta h$ , and  $\Delta t$ . In addition, we would like to reach equation (6) in the limit. We can ensure that this will indeed be the case by setting

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

and

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

Then

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

Substitute these expressions for  $\Delta h$  and  $p - q$  into the formulas above for the mean and variance of  $(x_t - x_0)$ , and let  $\Delta t$  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;  $\alpha$  is the drift, and  $\sigma^2$  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  $\Delta h$  and  $\Delta t$ .

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  $\Delta h$  and  $\Delta t$  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  $\Delta h$  and  $\Delta t$ . 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  $\Delta t$  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:

$$dx = a(x, t) dt + b(x, t) dz, \quad (11)$$

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  $\alpha$  and  $\sigma$  are constants. In this case equation (11) becomes

$$dx = \alpha x dt + \sigma x dz. \quad (12)$$

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  $x$ ,  $\Delta x$ , 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:

$$dF = \left( \alpha - \frac{1}{2}\sigma^2 \right) dt + \sigma dz, \quad (13)$$

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

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



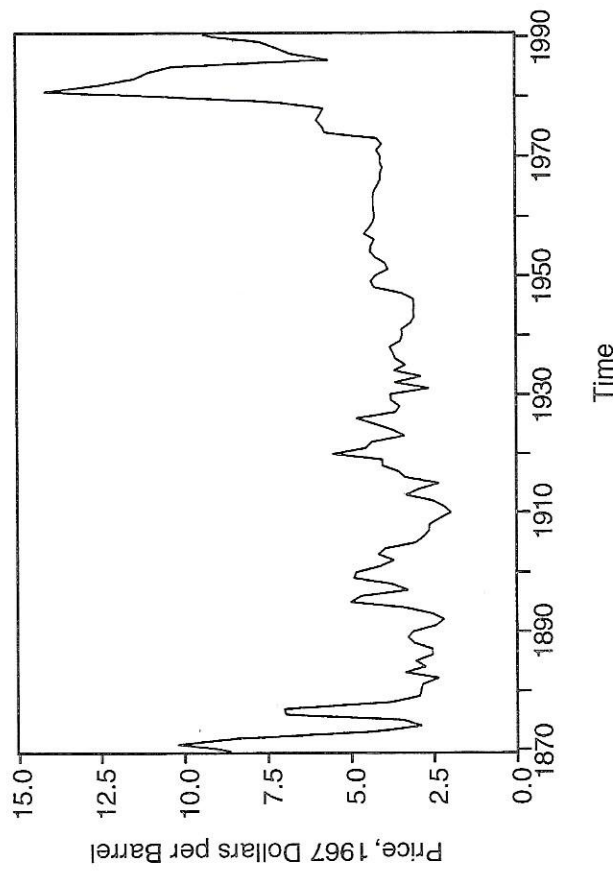


Figure 3.8. Price of Crude Oil in 1967 Dollars per Barrel

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.<sup>11</sup>

<sup>11</sup>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).

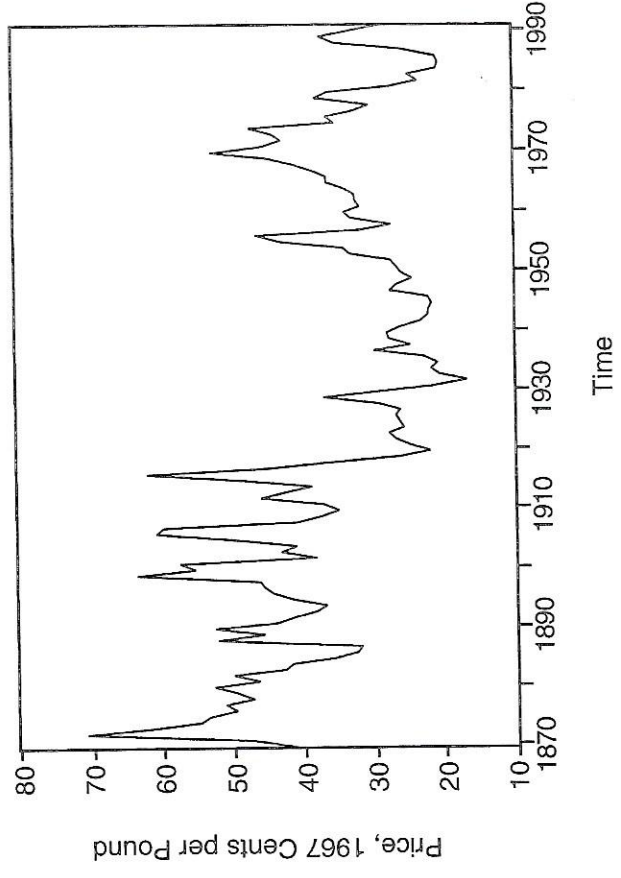


Figure 3.9. Price of Copper in 1967 Cents per Pound

### 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  $x$ :

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dx)^2 + \frac{1}{6} \frac{\partial^3 F}{\partial x^3} (dx)^3 + \dots \quad (22)$$

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  $(dx)^2$ :

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

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

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

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

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

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  $F = F(x_1, \dots, x_m, t)$  is a function of time and of the  $m$  Ito processes  $x_1, \dots, x_m$ , where

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

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

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

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

$$dF = \left[ \frac{\partial F}{\partial t} + \sum_i a_i(x_1, \dots, t) \frac{\partial F}{\partial x_i} + \frac{1}{2} \sum_i b_i^2(x_1, \dots, t) \frac{\partial^2 F}{\partial x_i^2} \right. \\ \left. + \sum_{i \neq j} \rho_{ij} b_{ij}(x_1, \dots, t) b_j(x_1, \dots, t) \frac{\partial^2 F}{\partial x_i \partial x_j} \right] dt \\ + \sum_i b_i(x_1, \dots, t) \frac{\partial F}{\partial x_i} dz_i. \quad (28)$$

**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):

$$dF = \frac{1}{x} dx - \frac{1}{2x^2} (dx)^2 \\ = \alpha dt + \sigma dz - \frac{1}{2} \sigma^2 dt = (\alpha - \frac{1}{2} \sigma^2) dt + \sigma dz. \quad (29)$$

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  $\alpha$ ? 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.