CHAPTER 1

Quantitative Measures of the Stock Market

1.1 PRICING FUTURE CASH FLOWS

Our first project in order to understand stock market risk, particularly downside risk, is to identify exactly what the stock market is and determine the motivation of its participants. Stock markets at their best provide a mechanism through which investors can be matched with firms that have a productive outlet for the investors' funds. It is a mechanism for allocating available financial funds into appropriate physical outlets. At the individual level the stock market can bring together buyers and sellers of investment instruments. At their worst, stock markets provide a platform for gamblers to bet for or against companies, or worse yet, manipulate company information for a profit. Each investor in the stock market has different aims, risk tolerance, and financial resources. Each firm has differing time horizons, scale of operations, along with many more unique characteristics including its location and employees.

So when it comes down to it, there need not be a physical entity that is *the* stock market. Of course, there are physical stock exchanges for a set of listed stocks such as the New York Stock Exchange. But any stock market is the combination of individuals. A trading floor is not a stock market without the individual investors, firms, brokers, specialists, and traders who all come together with their individual aims in mind to find another with complementary goals. For any routine stock trade, there is one individual whose goal it is to invest in the particular company's stock on the buy side. On the sell side, there is an individual who already owns the stock and wishes to liquidate all or part of the investment. With so much heterogeneity in the amalgam that is the stock market, our task of finding a common framework for all players seems

---

Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments,
by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7

intractable. However, we can find a number of features and commonalities which can be studied in a systematic manner.

1. The first among these commanalities is the time horizon. For any investor, whether saver or gambler, money is being invested in stock for some time horizon. For a young worker just beginning to save for his retirement through a mutual fund, this time horizon could be 30 years. For a day trader getting in and out of a stock position quickly, this time horizon could be hours, or even minutes. Whatever the time horizon, each investor parts with liquid assets for stock, intending to hold that stock for sale at a future date T. When we refer to prices, we will use the notation, P_T, where the subscript represents the time period for which the price applies. For example, if the time T is measured in years, P_0 denotes the current price (price today) and P_5 denotes the price five years from now.

2. The next commonality is that all investors expect a return on their investment. Since investors are parting with their money for a time, and giving up liquidity, they must be compensated. We will use r_T to represent the return earned on an investment of T years. Therefore r_1 would be the return earned on an investment after one year, r_5 on an investment after five years, and so on. Using our first two rules, we can derive a preliminary formula to price an asset with a future payment of P_T which returns exactly r_T percent per year for T years. We use capital T for the maturity date in this chapter. Lowercase t will be used as a variable denoting the current time period.

We start with the initial price, P_0, paid at the purchase date. After the first year, the investor would have the initial investment plus the return:

$$P_0(1+r_T). \tag{1.1.1}$$

For the second year, the return is compounded on the value at the end of the first year:

$$P_0(1+r_T)(1+r_T) \quad \text{or} \quad P_0(1+r_T)^2. \tag{1.1.2}$$

Thus the price that the investor must be paid in year T to give the required return is

$$P_0(1+r_T)(1+r_T)\ldots(1+r_T) = P_0(1+r_T)^T = P_T. \tag{1.1.3}$$

This is the formula to calculate a *future value* with compound interest each period. For example, interest compounded quarterly for two years would use the quarterly interest rate (annual rate divided by 4) and $T = 8$ periods.

Now let us find the fair price for this asset today, P_0, that will yield a return of exactly r_T every year for T years. Clearly, we would just need to divide through by $(1 + r_T)^T$ to obtain

$$\text{Present value} = P_{\mathrm{v}} = \frac{P_T}{(1+r_T)^T}. \tag{1.1.4}$$

This tells us that given a return r_T of periodic future payoffs, we can find the present price in order to yield the correct future price P_T at the end of the time horizon of length T. This formula is called the discounting, or present value, formula. The discounting formula is the basis of any pricing formula of a financial asset. Stocks and bonds, as well as financial derivatives such as options and futures, and hybrids between various financial instruments all start with the discounting formula to derive a price, since they all involve a time interval before the final payment is made.

Lest we get too comfortable with our solution of the price so quickly and easily, this misconception will be shattered with our last common feature of all investments.

3. All investments carry risk. In our discounting formula, there are only two parameters to plug in to find a price, namely the price at the end of time horizon or P_T and the return r_T. Unfortunately, for any stock the future payment P_T is not known with certainty. The price at which a stock is sold at time T depends on many events that happen in the holding duration of the stock. Company earnings, managerial actions, taxes, government regulations, or any of a large number of other random variables will affect the price P_T at which someone will be able to sell the stock.

With this step we have introduced uncertainty. What price P_0 should you pay for the stock today under such uncertainty? We know it is the discounted value of P_T, but without a crystal ball that can see into the future, P_0 is uncertain. There are a good many investors who feel this is where we should stop, and that stock prices have no fundamental value based on P_T. Many investors believe that the past trends and patterns in price data completely characterize most of the uncertainty of prices and try to predict P_T from data on past prices alone. These investors are called *technical analysts*, because they believe investor behavior is revealed by a time series of past prices. Some of these patterns will be incorporated in time series models in Section 4.1.

Another large group seeks to go deeper into the finances and prospects of the corporations to determine the fair value for the stock price represented by P_T. This group is called fundamental value investors, because they attempt to study intrinsic value of the firm. To deal with the fact that future prices are not known, fundamental value investors must base their value on risk, not uncertainty. By characterizing "what is not known" as risk, we are assuming that while we do not know exactly what will happen in the future, we do know what is possible, and the relative likelihoods of those possibilities. Instead of being lost in a random world, a study of risk lets us categorize occurrences and allows the randomness to be measured.

Using risk, we can derive a fundamental value of a firm's stock. As a stockholder, one has a claim of a firm's dividends, the paid out portion of net earnings. These dividends are random, and denoted as D^s for dividends in state s. This state is a member of a long list of possible occurrences. Each state represents a possibly distinct level of dividends, including extraordinarily high, average, zero, and bankruptcy. The probability of each state is denoted by π^s for $s = 1, 2, \ldots, S$, where S denotes the number of states considered. The more likely a state is, the higher is its probability. An investor can calculate the expected value of dividends that will be paid by summing each possible level of dividends multiplied by the corresponding probability:

$$E(D) = \sum_{s=1}^{S} \pi^s D^s, \quad \text{where} \sum_{s=1}^{S} \pi^s = 1, \tag{1.1.5}$$

where E is the expectations operator and Σ is the summation operator. Outcomes that are more likely are weighted by a higher probability and affect the expected value more. The expected value can also be thought of as the average value of dividends over several periods of investing, since those values with higher probabilities will occur more frequently than lower probability events. The sum of the probabilities must equal one to ensure that there is an exhaustive accounting of all possibilities.

Using the framework of risk and expected value, we can define the price of a stock as the discounted value of expected dividends at future dates, namely the cash flow received from the investment:

$$P_0 = \sum_{t=1}^{T} \frac{E(D)_T}{(1 + r_T)^T}, \tag{1.1.6}$$

where each period's expected dividends are discounted the appropriate number of time periods T by the compound interest formula stated in (1.1.6).

Formula (1.1.6) for the stock price is more useful, since it is based on the financials of a company instead of less predictable stock prices. One must forecast dividends, and thus have a prediction of earnings of a company. This approach is more practical since the other formula (1.1.4) was based on an unknown future price. One may ask the question: How can we have two formulas for the same price?

However, both formulas (1.1.4) and (1.1.6) are identical if we assume that investors are investing for the future cash flow from holding the stock. Our price based on discounted present value of future dividends looks odd, since it appears that we would have to hold the stock indefinitely to receive the entire value. What if we sell the stock after two years for a stock paying quarterly dividends (8 quarters)?

The value of our cash flow after including the end point price would be

$$P_0 = \frac{E(D)_1}{(1+r_1)^1} + \frac{E(D)_2}{(1+r_2)^2} + \ldots + \frac{E(D)_8}{(1+r_8)^8} + \frac{P_8}{(1+r_8)^8}, \tag{1.1.7}$$

where r_t is the quarterly return for the quarter t with $t = 1, \ldots, 8$. But realizing that the buyer in quarter 8 is purchasing the subsequent cash flows until they sell the stock one year later we have

$$\frac{P_8}{(1+r_8)^8} = \frac{E(D)_9}{(1+r_9)^9} + \frac{E(D)_{10}}{(1+r_{10})^{10}} + \frac{E(D)_{11}}{(1+r_{11})^{11}} + \frac{E(D)_{12}}{(1+r_{12})^{12}} + \frac{P_{12}}{(1+r_{12})^{12}}. \tag{1.1.8}$$

And so on it goes. So that recursively substituting the future prices yields P_0 equal to all future discounted dividends. This means that even for a stock not currently paying any dividend, we can use the same discounting formula. The stock must eventually pay some return to warrant a positive price.

Using the value of dividends to price a security may be unreliable, however. The motivation for a company issuing dividends is more complex than simply paying out the profits to the owners (see Allen and Michaely, 1995, for a survey of dividend policy). First, growth companies with little excess cash flows may not pay any dividend in early years. The more distant these dividends are, the harder they are to forecast. Dividends also create a tax burden for the investor because they are taxed as current income, whereas capital gains from holding the stock are not taxed until the stock is sold. This double taxation of dividends at the corporate and individual levels leads many to question the use of dividends at all, and has led many firms to buy back shares with excess cash rather than issue dividends. Also dividends are a choice made by the firm's management. Bhattacharya (1979) shows how dividends can signal financial health of a company, so firms are seen paying out cash through dividends and then almost immediately issuing more shares of stock to raise capital.

Alternatively, since dividend amounts are chosen by the management of the firm and may be difficult to forecast, price can be modeled as the present value of future earnings, ignoring the timing of exactly when they are paid out in the form of dividends. This model assumes that earnings not paid out as dividends are reinvested in the company for T years. So that if they are not paid in the current period, they will earn a return so that each dollar of "retained earnings" pays $1 + r_T$ next period. This makes the present value of expected earnings identical to the present value of dividends. Hence a lesson for the management is that they better focus on net earnings rather than window dressing of quarterly earnings by changing the dividend payouts and the timing of cash flows. The only relevant figure for determining the stock price is the bottom line of net earnings, not how it is distributed.

1.2 THE EXPECTED RETURN

Once the expected cash flows have been identified, one needs to discount the cash flows by the appropriate return, r_T. This is another value in the formula that looks deceptively simple. In this section we discuss several areas of concern when deciding the appropriate discount rate, namely its term, taxes, inflation, and risk, as well as some historical trends in each area.

The first building block for a complete model of returns is the risk-free rate, r_T^{f}. This is the return that would be required on an investment maturing in time T with no risk whatsoever. This is the rate that is required solely to compensate the investor for the lapse of time between the investment and the payoff. The value of the risk-free rate can be seen as the equilibrium interest rate in the market for loanable funds or government (FDIC) insured return:

The Borrower

A borrower will borrow funds only if the interest rate paid is less than or equal to the return on the project being financed. The higher the interest rate, the fewer the projects that will yield a high enough return to pay the necessary return.

The Lender

A lender will invest funds only if the interest rate paid is enough to compensate the lender for the time duration. Therefore, as the interest rate increases, more investors will be willing to forgo current consumption for the higher consumption in the future.

The Market

The equilibrium interest rate is the rate at which the demand for funds by borrowers in equal to the supply of funds from lenders; it is the market clearing interest rate in the market for funds. As can be seen from the source of the demand and supply of funds, this will be the return of the marginal project being funded (the project just able to cover the return), and at the same time this will be the time discount rate of the marginal investor.

A common observation about the interest rate is that the equilibrium return tends to rise as the length of maturity increases. Plotting return against length of maturity is known as the yield curve. Because an investor will need more enticement to lend for longer maturities due to the reduced liquidity, the yield curve normally has a positive slope. A negative slope of the yield curve is seen as a sign that investors are expecting a recession (reducing projected future returns) or that they are expecting high short-term inflation.

To see how inflation affects the required return for an investor, we can augment our return to get the nominal interest rate:

$$r_T^{\mathrm{n}} = r_T^{\mathrm{f}} + \pi_T^{\mathrm{e}}, \qquad (1.2.1)$$

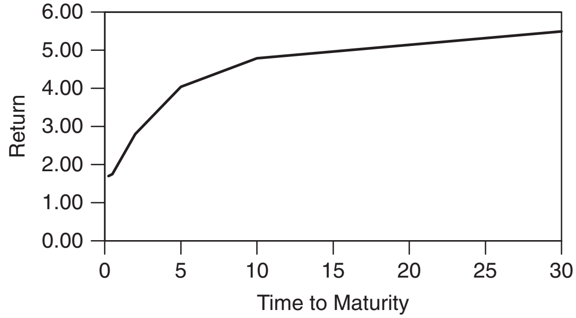

Figure 1.2.1 Yield curve for US treasuries, June 1, 2002

where π_T^e is the expected rate of inflation between time 0 and time T. For an investor to be willing to supply funds, the nominal return must not only compensate for the time the money is invested, it must also compensate for the lower value of money in the future.

For example, if \$100 is invested at 5% interest with an expected inflation rate of 3% in January 2002, payable in January 2003, the payoff of the investment after one year is \$105. But this amount cannot buy what \$105 will buy in 2002. An item that was worth \$105 in 2002 will cost \$105(1.03) = \$108.15 in 2003. To adjust for this increase in prices, to the nominal interest rate is added the cost of inflation to the return.

One may wonder about the extra 15 cents that the formula above does not include. According to the formula for nominal rate, an investor would get 5% + 3% = 8%, or \$108 at the payoff date. That is because the usual formula for nominal rate is an approximation: it only adjusts for inflation of the principal but not the interest of the loan. The precise formula will be

$$r_T^n = r_T^f + \pi_T^e + r_T^n \pi_T^e \quad \text{or} \quad r_T^n = \frac{r_T^f + \pi_T^e}{1 - \pi_T^e}. \tag{1.2.2}$$

As the additional term is the interest rate times the expected inflation rate, two numbers are usually less than one. Unless either the inflation rate or the interest rate is unusually high, the product of the two is small and the approximate formula is sufficient.

Our next adjustment comes from taxes. Not all of the nominal return is kept by the investor. When discounting expected cash flows then, the investor must ensure that the after-tax return is sufficient to cover the time discount:

$$r_T^{at} = r_T^n (1 - \tau), \tag{1.2.3}$$

where r_T^{at} is the after-tax return and τ is the tax rate for an additional dollar of investment income.

It is important to note that taxes are applied to the nominal return, not the real return (return with constant earning power). This makes an investor's forecast of inflation crucial to financial security.

Consider the following two scenarios of an investment of \$100 with a nominal return of 12.31% at a tax rate of 35% of investment income. The investor requires a risk-free real rate of interest of 5% and expects inflation to be 3%. The investment is to be repaid in one year.

Scenario 1—Correct Inflation Prediction

If inflation over the course of the investment is, indeed, 3%, then everything works correctly. The investor is paid \$112.31 after one year, \$12.31(0.35) = \$4.31 is due in taxes, so the after-tax amount is \$108.00. This covers the 5% return plus 3% to cover inflation.

Scenario 2—Underestimation of Inflation

If actual inflation over the course of the investment turns out to be 10%, the government does not consider this an expense when it comes to figuring taxable income. The investor receives \$112.31, which nominally seems to cover inflation, but then the investor must pay the same \$4.31 in taxes. The \$108.00 remaining is actually worth less than the original \$100 investment since the investor would have had to receive at least \$110 to keep the same purchasing power as the original \$100.

Scenario 2 shows how unexpectedly high inflation is a transfer from the investor, who is receiving a lower return than desired to the borrower, who pays back the investment in dollars with lower true value.

The final element in the investor's return is the risk premium, θ, so that the total return is

$$r_T^{\mathrm{n}} = \frac{r_T^{\mathrm{f}} + \pi_T^{\mathrm{e}} + \theta}{1-\tau}. \tag{1.2.4}$$

The risk premium is compensation for investing in a stock where returns are not known with certainty. The value of the risk premium is the most nebulous of the parameters in our return formula, and the task of calculating the correct risk premium and methods to lower the risk of an investment will be the subject of much of the balance of this book. At this point we will list some of the important questions in defining risk, leaving the detail for the indicated chapter.

1. How do investors feel about risk? Are they fearful of risk such that they would take a lower return to avoid risk? Or do they appreciate a bit of risk to liven up their life? Perceptions of investors to risk will be examined in Chapter 6.

2. Is risk unavoidable, or are there investment strategies that will lower risk? Certainly investors should not be compensated for taking on risk that could have been avoided. The market rarely rewards the unsophisticated investor (Chapters 2 and 3).
3. Is the unexpected return positive or negative? Most common measurements of risk (e.g., standard deviation) consider unexpected gains and losses as equally risky. An investor does not have to be enticed with a higher return to accept the "risk" of an unexpected gain. This is evidenced by the fact that individuals pay for lottery tickets, pay high prices for IPOs of unproven companies, and listen intently to rumors of the next new fad that will take the market by storm. We explain how to separate upside and downside risk in Chapter 5, and evidence of the importance of the distinction in Chapters 7, 8, and 9.

1.3 VOLATILITY

In order to develop a measure of the risk premium, we must first measure the volatility of stock returns. The term "volatility" suggests movement and change; therefore any measurement of volatility should be quantifying the extent to which stock returns deviate from the expected return, as discussed in Section 1.2. Quantifying change, however, is not a simple task. One must condense all the movements of a stock throughout the day, month, year, or even decade, into one measure. The search for a number that measures the volatility of an investment has taken numerous forms, and will be the subject of several subsequent chapters since this volatility, or movement, of stock prices, is behind our notion of risk. Without volatility, all investments are safe. With volatility, stocks yield gains and losses that deviate from the expected return.

As an example, we will use the annual return for the S&P 500 index from 1990 through 2000 shown in Table 1.3.1. The average annual return for this time period is 13.74%. Compare this return to the return of U.S. three-month Treasury bills for the same time period (Table 1.3.2) that is on average 4.94%.

Table 1.3.1 S&P 500 Index Annual Return

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
–6.56%	26.31%	4.46%	7.06%	–1.54%	34.11%	20.26%	31.01%	26.67%	19.53%	–10.14%

Table 1.3.2 US Three-Month Treasury Bills Return

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
7.50%	5.38%	3.43%	3.00%	4.25%	5.49%	5.01%	5.06%	4.78%	4.64%	5.82%

The average return of the stock index is almost three times the average return on T-bills. There must be a reason, or there should be no investor buying T-bills. Both are in dollars, so inflation is the same. Both are under the same tax system, although T-bill interest is taxed as income, and stock returns as capital gains, but that should give a higher return to T-bills. Capital gains taxes are usually lower than income taxes, and they can be delayed so they are even lower in present discounted value.

Common sense tells us the reason for the difference in returns is the volatility. While the T-bill return is consistently around 4% or 5%, the stock return has wide swings in the positive and negative range. In a free market economy, if investment in risky assets creates economic growth, new jobs, and new conveniences, these risky activities have to be rewarded. Otherwise, there will be no one taking the risks. This means the market forces must reward a higher return for investors in certain wisely chosen risky activities. Such higher return is called risk premium. Volatility is therefore very important in determining the amount of risk premium applied to a financial instrument.

To measure volatility, the simplest measure would be the range of returns, when the range is defined as the highest return less the lowest return. The S&P 500 has a range of [34.11 – (–10.14)] = 44.25. For T-bills, the range is [7.5 – 3] = 4.5. The range of returns is much larger for the S&P 500, showing the higher volatility.

The range has the benefit of ease of calculation, but the simplest measure is not always the best. The problem with the range is that it only uses two data points, and these are the two most extreme data points. This is problematic because the entire measure might be sensitive to outliers, namely to those extreme years that are atypical. For instance, a security will have the same average return and range as the S&P 500 if returns for nine years were 13.74%, the next year has a money-losing return of –8.39% and the next year has a spectacular return of 35.86%. But the volatility of this security is clearly not identical to the S&P 500, though the range is the same. This security is very consistent because only two years have extreme returns.

In order to take all years into account, one simply takes the deviations from the mean of each year's returns

$$r_t - \mu, \tag{1.3.1}$$

where μ is the average return for the respective security for each time period t. To condense these deviations into one measure, there are two common approaches. Both approaches try to put a single value on changes of the returns. Since values above or below the mean are both changes, the measure needs to treat both positive and negative values of deviations as an increase in volatility.

The mean absolute deviation (MAD) does this by taking the absolute value of the deviations, and then a simple average of the absolute values,

Table 1.3.3 S&P 500 Index Deviations from Mean

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
–20.30%	12.57%	–9.28%	–6.68%	–15.28%	20.37%	6.52%	17.27%	12.93%	5.79%	–23.88%

$$\text{MAD} = \frac{\sum_t |r_t - \mu|}{T}, \tag{1.3.2}$$

where Σ_t denotes sum from $t = 1$ to $t = T$ and where T is again the total number of years.

For the S&P 500, the mean absolute deviation is 13.72%. For the T-bill series, the mean absolute deviation is 0.84%. This shows the dramatic difference in volatility between the two securities.

The other way to transform the deviations to positive numbers is to square them. This is done with the variance, σ^2:

$$\sigma^2 = \frac{\sum_t (r_t - \mu)^2}{T}. \tag{1.3.3}$$

Note: This variance formula is often adjusted for small samples by replacing the denominator by $(T - 1)$. A discussion of sampling is in Chapter 9.

The variance formula implicitly gives larger deviations a larger impact on volatility. Therefore 10 years of a 2% deviation ($0.02^2 \times 10 = 0.004$) does not increase variance as much as one year of a 20% deviation ($0.2^2 = 0.04$).

The variance for the S&P 500 is 0.0224, for T-bills, 0.0001. It is common to present the standard deviation, which is the square root of the variance so that the measure of volatility has the same units as the average. For the S&P 500, the standard deviation is 0.1496, for T-bills, 0.0115.

The advantage in using the standard deviation is that all available data can be utilized. Also some works have shown that alternate definitions of a deviation can be used. Rather than strictly as deviations from the mean, risk can be defined as deviations from the risk-free rate (CAPM, ch. 2). Tracking error (Vardharaj, Jones, and Fabozzi, 2002) can be calculated as differences from the target return for the portfolio. When an outside benchmark is used as the target, the tracking error is more robust to prolonged downturns, which otherwise would cause the mean to be low in standard deviation units. Although consistent loss will show a low standard deviation, which is the worst form of risk for a portfolio, it will show up correctly if we use tracking error to measure volatility.

Other methods have evolved for refining the risk calculation. The intraday volatility method involves calculating several standard deviations throughout the day, and averaging them. Some researchers are developing methods of

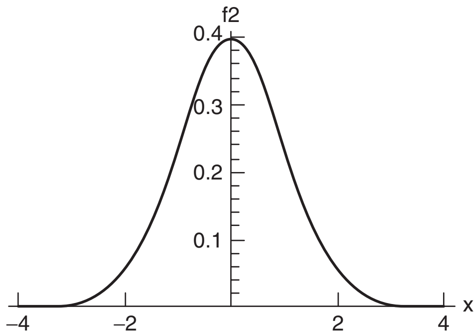

Figure 1.3.1 Probability density function for the standard normal N(0, 1) distribution

using the intraday range of prices in the calculation of standard deviations over several days. By taking the high and low price instead of the opening and closing prices, one does not run the risk of artificially smoothing the data and ignoring the rest of the day. The high and low can come at any time during the day.

Once the expected return and volatility of returns are calculated, our next step is to understand the distribution of returns. A probability distribution assigns a likelihood, or probability, to small adjacent ranges of returns. Probability distributions on continuous numbers are represented by a probability density function (PDF), which is a function of the random variable $f(x)$. The area under the PDF is the probability of the respective small adjacent range of the variable x. One commonly used distribution is the normal distribution having mean μ and variance σ^2, $N(\mu, \sigma^2)$, written

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{((x-\mu)/\sigma)^2 / 2}, \tag{1.3.4}$$

where μ is the mean of the random variable x and σ is the standard deviation. Once we know these two parameters, we know the entire probability distribution function (pdf) of $N(\mu, \sigma^2)$.

It can be seen from the normal distribution formula (1.3.4) why the standard deviation σ is such a common measure of dispersion. If one assumes that returns follow the normal distribution, with the knowledge of only the average of returns (μ) and the standard deviation (σ), all possible probabilities can be determined from widely available tables and software sources. Therefore an infinite number of possibilities can be calculated from only two statistics. This is a powerful concept. (We will discuss the validity of using the normal distribution for stock returns in Chapter 4.)

The normal distribution is a common distribution because it seems to possess several characteristics that occur in nature. The normal distribution has most of the probability around the average. It is symmetrical, meaning the

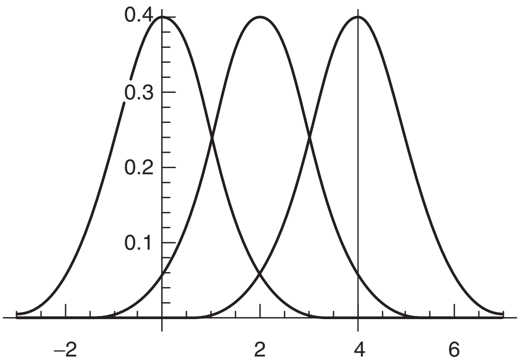

Figure 1.3.2 Normal distribution with change in the average $\mu(= 0, 2, 4)$ with $\sigma = 1$

probability density function above and below the average are mirror images. The probability of getting outcomes an extreme distance above or below the average are progressively unlikely, although the density function never goes to zero, so all outcomes are possible. Children's growth charts, IQ tests, and bell curves are examples of scales that follow the normal distribution.

Since one need only know the average and standard deviation to draw a specific normal distribution, it is a useful tool for understanding the intuition of expected value and volatility. Probability statements can be made in terms of a certain number of standard deviations from the mean. There is a 68.3% probability of x falling within one standard deviation of the mean, 95.5% probability two standard deviations of the mean, 99.74% probability three standard deviation from the mean, and so on.

The normal probability can change dramatically with changes of the parameters. Increases in the average will shift the location of the normal distribution. Increases in the standard deviation will widen the normal distribution. Decreases in the standard deviation will narrow the distribution.

Because the normal distribution changes with a change in the average or standard deviation, a useful tool is standardization. This way the random variable can be measured in units of the number of standard deviations measured from the mean:

$$z = \frac{x - \mu}{\sigma}. \tag{1.3.5}$$

If x is normally distributed with mean μ and standard deviation σ, then the standardized value z will be standard normally distributed with mean of zero, and standard deviation equal to one. In statistical literature this relation is often stated by using the compact notation: $x \sim N(\mu, \sigma^2)$ and $z \sim N(0, 1)$. It can be verified by some simple rules on the expectations (averages) of random numbers stated below. Given a and b as some constant real numbers, we have:

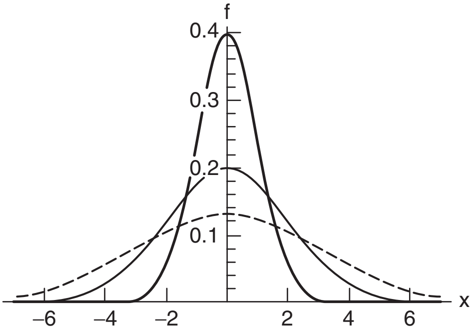

Figure 1.3.3 Normal distribution with change in σ, the standard deviation

1. If the average of $x = \mu$, then the average of $a(x) = a(\mu)$.
2. If the average of $x = \mu$, then the average of $(x + b) = \mu + b$. Therefore the average of $x - \mu = \mu - \mu = 0$.
3. If the standard deviation of $x = \sigma$, then the standard deviation of $a(x) = a(\sigma)$. (Note: The variance of $a(x) = a^2\sigma^2$.)
4. If the standard deviation of $x = \sigma$, then the standard deviation of $(x + b) = \sigma$. Therefore the standard deviation of $(x - \mu)/\sigma = (1/\sigma)\sigma = 1$.

Through standardization, tables of the area under the standard normal distribution can be used for normal distributions with any average and standard deviation. To use the tables, one converts the x value under the normal distribution to the standardized z statistic under the standard normal and looks up the z value in the table. The probability relates back to the original x value, which is then the number of standard deviations from the mean. With the wide availability of Excel software workbooks, nowadays it is possible to avoid the normal distribution tables and get the results directly for $x \sim N(\mu, \sigma^2)$ or for $z \sim N(0, 1)$.

Therefore, for the normal distribution, mean and standard deviation are the end of the story. Since symmetry is assumed, the distinction of downside risk is moot. However, for this reason the normal distribution is not always a practical assumption, but it provides a valuable baseline against which to measure downside adjustments. The next section provides a dynamic framework of modeling stock returns following the normal distribution.

1.4 MODELING OF STOCK PRICE DIFFUSION

A probability distribution gives the likelihood of ranges of returns. If one assumes the normal distribution, then the distribution is completely defined by its average and standard deviation. Knowing this, one can model the discrete movement of a stock price over time through the diffusion equation, which combines the average return μ and the volatility measured by the standard deviation σ.

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma z \sqrt{\Delta t}, \tag{1.4.1}$$

where Δ is the difference operator ($\Delta S = \Delta S_t = S_t - S_{t-\Delta t}$), S is the stock price, Δt is the time duration, and z is N(0, 1) variable. Note that $\Delta S/S$ is the relative change in the stock price, and the relative changes times 100 is the percentage change. Equation (1.4.1) seeks to explain how relative changes are diffused as the time passes around their average, subject to random variation.

The diffusion equation (1.4.1) has two parts: the first part of the percentage change is the average return μ per time period (or drift), multiplied by the number of time periods that have elapsed; the second part is the random component that measures the extent to which the return can deviate from the average. Also we see that the standard deviation is increased by the square root of the time change. The root term arises because it can be shown that (1.4.1) follows what is known as a random walk (also known as Brownian motion, or Weiner process). If S_t follows a random walk, it can be written as $S_t = S_{t-1} + \delta + \varepsilon$, where the value of the stock price at any point in time is the previous price, plus the drift (= δ), plus some random shock (= ε). The diffusion process is obviously more general than a simple random walk with drift. The more time periods out you go, the more random shocks are incorporated into the price. Since each one of these shocks has its own variance, the total variance for a length of time of Δt will be $\sigma^2 \Delta t$. Thus the standard deviation will be the square root of the variance.

The cumulative effect of these shocks from (1.4.1) can be seen by performing a simple simulation starting at a stock price of 100 for a stock with an average return of 12% per year and a standard deviation of 5% and the following random values for z. For a complete discussion of simulations, see Chapter 9.

At any time t, the price the next day ($\Delta t = 1/365 = 0.0027$) will be

$$S_{t+1} = S_t + \Delta S_t = S_t + \mu S_t \Delta t + \sigma S_t z_t \sqrt{\Delta t}. \tag{1.4.2}$$

Simulating random numbers for 30 days yields the stock prices in Table 1.4.1.

Looking at the stock prices in a graph, we can see that simulation using a random walk with drift gives a plausible series of stock prices. A few other insights can be gained from the graph. We can see that a random walk, as the name implies, is a movement from each successive stock price, not reverting back to an average stock price (this is the basis for another term associated with random walk: nonstationary). Also, as the stock price gets higher, the movements get larger since the price is the percentage of a larger base.

Table 1.4.1 Simulated Stock Price Path

t	z_t	ΔS_t	S_t
0			100.00
1	1.24	0.35	100.35
2	0.23	0.09	100.45
3	−1.08	−0.25	100.20
4	0.36	0.13	100.33
5	−0.21	−0.02	100.30
6	0.07	0.05	100.35
7	−0.37	−0.06	100.29
8	−0.30	−0.05	100.25
9	0.37	0.13	100.38
10	−0.03	0.02	100.40
11	−0.16	−0.01	100.39
12	0.61	0.19	100.58
13	0.03	0.04	100.62
14	−0.44	−0.08	100.54
15	0.02	0.04	100.58
16	−0.67	−0.14	100.44
17	0.19	0.08	100.52
18	1.04	0.30	100.82
19	1.42	0.40	101.23
20	1.49	0.42	101.65
21	0.48	0.16	101.81
22	−0.52	−0.10	101.71
23	0.53	0.17	101.88
24	1.41	0.41	102.29
25	2.05	0.58	102.86
26	0.45	0.15	103.02
27	−0.79	−0.18	102.84
28	0.78	0.24	103.08
29	2.05	0.58	103.67
30	−0.58	−0.12	103.54

This general diffusion model of stock prices has gone through many alterations for specific stock pricing situations. The descriptions that follow cover only a few of these adaptations.

1.4.1 Continuous Time

For empirical use, or for producing simulations, we can only work with discrete time changes, but theoretically a continuous time approach (as $\Delta t \to 0$) can model the path of stock prices at each moment in time. This often simplifies calculations and yields more elegant results. The continuous time diffusion equation is

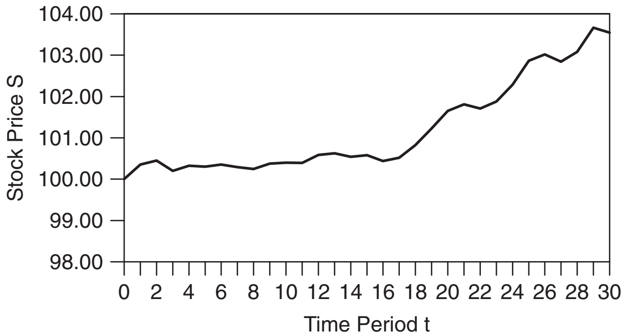

Figure 1.4.1 Simulated stock prices over 30 time periods

$$\frac{dS}{S} = \mu dt + \sigma dz, \tag{1.4.3}$$

where d denotes an instantaneous change. The dz in (1.4.3) represents a standard Wiener process or Brownian motion (Campbell et al., 1997, p. 344) that is a continuous time analogue of the random walk mentioned above.

1.4.2 Jump Diffusion

The jump diffusion process recognizes the fact that not all stock movements follow a continuous smooth process. Natural disasters, revelation of new information, and other shock can cause a massive, instantaneous revaluation of stock prices. To account for these large shocks, the normal diffusion is augmented with a third term representing these jumps:

$$\frac{dS}{S} = (\mu - \lambda k)dt + \sigma dz + dq, \tag{1.4.4}$$

where λ is the average number of jumps per unit of time, k is the average proportionate change of the jump (the variance of the jump is δ^2 to be used later), and dq is a Poisson process. The adjustment to the drift term ensures that the total average return is still μ: $(\mu - \lambda k)$ from the usual random walk drift, plus λk from the jump process leading to a cancellation of λk.

In the Poisson process, the probability of j number of jumps in T time periods is determined by the Poisson discrete probability function

$$P(j) = \frac{e^{-\lambda T}(\lambda T)^j}{j!}. \tag{1.4.5}$$

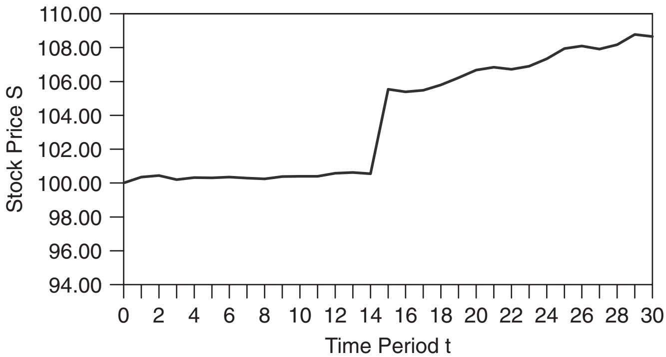

Figure 1.4.2 Simulated stock prices for jump diffusion process

A graph of the same stock diffusion in Table 1.4.1 with a jump of \$5 occurring on the 15th day is given in Figure 1.4.2. As can be seen from the graph, a jump will increase the volatility of the stock returns dramatically, depending on the volatility of the jump and the average number of jumps that occur. The total variance of the process is then $\sigma^2 + \lambda\delta^2$ per unit of time. This can also be used as a method to model unexpected downside shocks through a negative jump.

1.4.3 Mean Reversion in the Diffusion Context

For stock prices that should be gaining a return each period, a random walk with drift seems a reasonable model for stock prices. For some investments, however, it does not seem reasonable that their price should constantly be wandering upward. Interest rates or the real price for commodities, such as oil or gold, are two such examples in finance of values that are not based on future returns, and thus have an intrinsic value which should not vary over time. Prices that revert back to a long-term average are known as mean reverting (or stationary).

Mean reversion can be modeled directly in the diffusion model

$$dS = \eta(\bar{S} - S)dt + \sigma dz, \tag{1.4.6}$$

where $\bar{S}$ is the average value of the financial asset and $0 < \eta < 1$ is the speed at which the asset reverts to its mean value. Since a mean reverting process is centered around $\bar{S}$ and always has the same order of magnitude, the diffusion need not be specified in terms of percentage changes. Graphing the diffusion using the random numbers as above and an average price of \$100 gives the path for two different speeds $\eta = 0.8$ and $\eta = 0.2$ of mean reversion.

From Figure 1.4.3 both processes stay near 100. The solid line path with the higher reversion speed ($\eta = 0.8$) snaps back to 100 quicker, even after large shocks to the average price level. For the stock with the lower reversion speed,

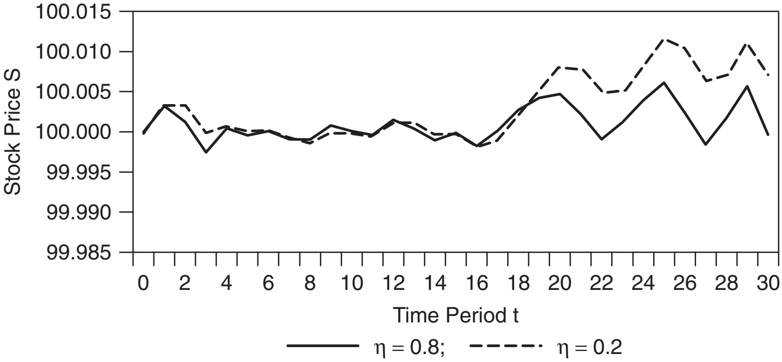

Figure 1.4.3 Simulated stock prices for two mean-reverting processes (——— $\eta = 0.8$; ------ $\eta = 0.2$)

large increases or decreases linger because the stock takes smaller steps back to its average price similar to the dashed line. We discuss mean reversion in a general context in Section 4.1.2.

1.4.4 Higher Order Lag Correlations

In the mean reverting process, the change in a stock price S_t is affected by how far the previous period is from the mean $(S_{t-1} - \mu)$. But with the high frequency at which stock data is available (e.g., hourly), it is realistic that correlations could last longer than one period. Would one expect that a boost in sales at the end of February would immediately be gone in the beginning of March? Would a news report at 10:00 am on Tuesday morning be completely reflected in the stock price by 10:01 am? No.

A method to account for these holdovers from past periods $(S_{t-2}, S_{t-3}, \ldots)$ is the ARIMA model. The AR stands for autoregressive, or the previous period returns that are directly affecting the current price. The MA stands for moving average, or the previous period shocks that are directly affecting the current price. The "I" in the middle of ARIMA stands for integrated, which is the number of times the data must be transformed by taking first differences $(S_t - S_{t-1})$ over time. If L denotes the lag operator $LS_t = S_{t-1}$, $(S_t - S_{t-1})$ becomes $(1 - L)S_t = \Delta S_t$. If $(1 - L) = 0$ is a polynomial in the lag operator, its root is obviously $L = 1$, which is called the unit root. Most stocks have unit root and are said to be *integrated of order 1*, I(1), meaning that one time difference is necessary to have a stationary process. Since taking returns accomplishes this, returns would be *stationary* or integrated of order zero, I(0). So we can work with returns directly.

When the stock price increases, ΔS_t is positive. Let us ignore the dividends temporarily, and let $r_t = \Delta S_t/S_t$ denote the stock return for time t. The autoregressive model of order p, AR(p), with p representing the maximum lag length of correlation, would be

$$\text{AR}(p): \quad r_t = \mu(1-\rho_1-\rho_2-\ldots-\rho_p)+\rho_1 r_{t-1}+\rho_2 r_{t-2} \\ +\ldots+\rho_p r_{t-p}+z_t\sigma, \tag{1.4.7}$$

whereas the MA process of order q includes lags of only the random component

$$r_t = \mu+\phi_1 z_1\sigma+\phi_2 z_2\sigma+\ldots+\phi_q z_{t-q}\sigma+z_t\sigma, \tag{1.4.8}$$

where μ is the average return.

The AR process never completely dies since it is an iterative process. Consider an AR(p) process with $p = 1$. Now let the first period be defined at $t = 0$, and substitute in (1.4.7) to give

$$r_0 = \mu+z_0\sigma. \tag{1.4.9}$$

During the next period we have the term (ρ_1 r_0), as this value gets factored into the return by a proportion ρ_1. As a result

$$r_1 = \mu(1-\rho_1)+\rho_1 r_0+z_1\sigma = \mu+\rho_1 z_0\sigma+z_1\sigma. \tag{1.4.10}$$

Because z_0 influences r_1, it also gets passed through to the next period as

$$r_2 = \mu(1-\rho_1)+\rho_1 r_1+z_2\sigma = \mu+\rho_1(\rho_1 z_0\sigma+z_1\sigma)+z_2\sigma \tag{1.4.11}$$

so that a shock t periods ago will be reflected by a factor of $(\rho_1)^t$. In order for the process to be stationary, and eventually return to the average return, we need $|\Sigma_s\rho_s| < 1$, meaning only a fraction of the past returns are reflected in the current return. The flexible nature of this specification has made the ARIMA model important for forecasting. The estimation of the parameters and use of the ARIMA model for simulations will be discussed further in Section 4.1.2.

1.4.5 Time-Varying Variance

All of the diffusion methods used to define the change of returns can also be applied to the variance of stock prices. Stocks often go through phases of bull markets where there are rapid mostly upward price changes and high volatility, and bear markets where prices are moving mostly downward or relatively stagnant. As seen in Figure 1.4.4, the standard deviation of returns for S&P 500 has gone through several peaks and troughs over time. In the basic diffusion model (1.4.1), however, the standard deviation σ is assumed to be constant over time.

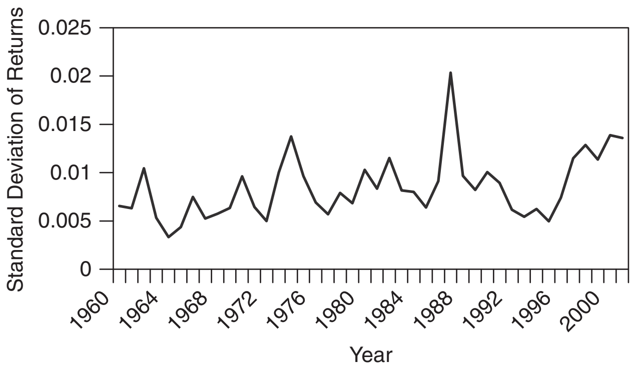

Figure 1.4.4 Standard deviation of the S&P 500 over time

A more general notation for the diffusion model would be to reflect that both the drift and the volatility are both potentially a function of the stock price and time:

$$dS = \mu(t, S)dt + \sigma(t, S)dz. \qquad (1.4.12)$$

A comprehensive parametric specification allowing changes in both drift and volatility is (see Chan et al., 1992):

$$dS = (\alpha + \beta S)dt + \sigma S^{\gamma} dz. \qquad (1.4.13)$$

This model is flexible and encompasses several common diffusion models. For example, in the drift term, if we have $\alpha = \eta\overline{S}$ and $\beta = -\eta$, there is mean reversion of (1.4.6). If $\alpha = 0$, $\beta = 1$, and $\gamma = 1$, it is a continuous version of the random walk with drift given in (1.4.1). When $\gamma > 1$, volatility is highly sensitive to the level of S.

Depending on the restrictions imposed on the parameters α, β, and γ in equation (1.4.13) one obtains several nested models. Table 1.4.2 shows the eight models (including the unrestricted) considered here and explicitly indicates parameter restrictions for each model.

The first three models impose no restrictions on either α or β. Models 4 and 5 set both α and β equal to zero, while models 6, 7, and 8 set either α or β equal to zero. Model 1, used by Brennan and Schwartz (1980), implies that the conditional volatility of changes in S is proportional to its level. Model 2 is the well-known square root model of Cox, Ingersoll, and Ross (CIR) (1985), which implies that the conditional volatility of changes in S is proportional to the square root of the level. Model 3 is the Ornstein-Uhlenbeck diffusion process first used by Vasicek (1977). The implication of this specification is that the conditional volatility of changes in S is constant. Model 4 was used by CIR (1980) and by Constantinides and Ingersoll (1984) indicates that the

Table 1.4.2 Parameter Restrictions on the Diffusion Model $dS = (\alpha + \beta S)dt + \sigma S^{\gamma}dz$

Model Name	α	β	γ	σ	Diffusion Model
1. Brennan-Schwartz			1		$dS = (\alpha + \beta S)dt + \sigma S dz$
2. CIR SR			0.5		$dS = (\alpha + \beta S)dt + \sigma S^{0.5}dz$
3. Vasicek			0		$dS = (\alpha + \beta S)dt + \sigma dz$
4. CIR V	0	0	1.5		$dS = \sigma S^{1.5}dz$
5. Dothan	0	0	1		$dS = \sigma S dz$
6. CEV	0				$dS = \beta S dt + \sigma S^{\gamma}dz$
7. GBM	0		1		$dS = \beta S dt + \sigma S dz$
8. Merton		0	0		$dS = \alpha dt + \sigma dz$

conditional volatility of changes in S is highly sensitive to the level of S. Model 5 was used by Dothan (1978), and model 6 is the constant elasticity of variance (CEV) process proposed by Cox (1975) and Cox and Ross (1976). Model 7 is the famous geometric Brownian motion (GBM) process first used by Black and Scholes (1973). Finally, model 8 is used by Merton (1973) to represent Brownian motion with drift.

This flexible parametric specification is useful since the parameters may be estimated (see Chapter 9) to determine the model that best fits a particular security. Vinod and Samanta (1997) estimate all these models to study the nature of exchange rate dynamics using daily, weekly, and monthly rates for the British pound sterling and the Deutsche mark for the 1975 to 1991 period and compare the out-of-sample forecasting performance of the models in Table 1.4.2. The models Brnnan-Schwartz, CIR-SR, and Vasicek performed poorly, whereas CIR-VR and GBM were generally the best.

Now that we have examined some of the possible ways that stock prices can move, the next section explores the way that a stock price should move in an efficient market.

1.5 EFFICIENT MARKET HYPOTHESIS

We have concentrated thus far on the similarities of investors and investments. There are many commonalities that market participants share. The mantra of Wall Street is "buy low and sell high," and all investors would prefer to do so. Looking at the big picture, this makes stock investing a tricky proposition. Any stock trade involves a buyer and a seller. They cannot both be right.

For speculators, the mantra may be valid, since they try to profit on short-term price movements. But for the average investors who are interested in receiving a fair return on their investment, a stock trade need not be a zero-sum game. While stock prices are certainly risky, it is not the same as a card game where one player's gain is another player's loss. Stock trades can be mutually beneficial, since each investor has different needs and endowments at different times in the life span. Consider a trade involving an investor who is just entering retirement and needs to liquidate some of his portfolio for living expenses selling to a new father who needs an investment for his new child's college fund. Note that both are better off if they were not able to complete the trade. Without the trade, the retiree may have to do without basic necessities if he does not have ready cash; without the college fund, the child would be worse off.

In order for investors to be confident that the stock market is better than a gamble, stocks must be priced fairly. Any major stock market in the world insists that investors have all available information about a company before purchasing shares. This ensures that the purchase is an informed decision. If a stock is trading at a price lower than its fair value, then sellers will be missing out on value due to them. Stocks that are overvalued go against the buyers. The idea that stocks are continually priced at their fair value is known as the efficient market hypothesis (EMH).

The EMH is an attractive idea for stock investors, since even investors who are not stock analysts, and do not have time to perform in depth research for every company in their portfolios, are assured of trading at a fair price by simply buying or selling at the current market price. Of course, this argument taken to its extreme has all investors relying on the market to price fairly, and no one doing the homework to figure out what this fair price should be. We can see, then, that the EMH creates a market of "smart money" investors who are well informed, as opposed to investors trading at whatever the prevailing price happens to be. It is a waste of resources to have all investors well informed if prices are fair already, but there must be some smart money to ensure the prices get to this fair value.

The EMH comes in three strengths depending on how informed we assume the market price is: weak, semi-strong, and strong.

1.5.1 Weak Form Efficiency

This form of the efficient market hypothesis assumes that all historical information is factored into the market price of a stock. Evidence of weak form efficiency can be seen by the demand for high-speed information, and the high price charged for real time financial news. Leinweber (2001) documents the speed at which earnings announcements are reflected in stock prices. He shows that in the 1980s earnings surprises could take up to two weeks to be incorporated into the stock price. In the 1990s, only a decade later, stock prices jumped within minutes of earnings announcements.

What happened to cause this change? In the 1980s, investment news was still very much a print industry. Earnings announcements were on record in the *Wall Street Journal*, which has an inevitable lag because of printing time and delivery. In the 1990s technology took over, and electronic news services such as Reuters and Bloomberg, and Internet news services were the source for late-breaking news, with the newspapers providing analysis and often ex post credibility. The speed at which transactions could take place also increased with computerized trading and discount brokers. The increased technology makes it hard to argue against weak form efficiency in a market where old news has little value.

1.5.2 Semi-strong Form Efficiency

This second level of the efficient market hypothesis encompasses weak form efficiency, and adds the additional requirement that all expectations about a firm are incorporated into the stock price. There is no reason why an investor who thought an interest rate hike was an inevitability would wait until the formal announcement to trade on the information. In previous sections we have implicitly used semi-strong form efficiency in our pricing formulas by pricing *expected* earnings or *expected* dividends.

Semi-strong form efficiency can also explain some of the counterintuitive movements in the stock market, such as the market going up after bad news, or declining on seemingly good news. If the bad news was not as bad as expected, the price can actually rebound when the uncertainty is resolved. Consider the following example: A firm has a fundamental value of \$100 per share at the current level of interest rates of 6%. Inflation starts to pick up, and the Federal Reserve Bank (the Fed) considers an interest rate hike. An interest rate hike increases the cost of funds for the company and therefore reduces its value. The firm's analysts have come up with the possible scenarios listed in Table 1.5.1.

Operating under weak form efficiency, the stock price of this firm would not necessarily move, since no announcement has been made. However, we see that all of the scenarios in Table 1.5.1 involve a rate hike and a share price decrease. Hence it stands to reason that in the last column of the table investors pay less than the original price of \$100 per share.

Under semi-strong form efficiency, the investor prices the stock based on expected price:

$$E(P) = 0.2(95) + 0.5(89) + 0.3(80) = \$87.50 \quad \text{per share.} \tag{1.5.1}$$

Since the price already reflects the consensus rate hike of a little over $\frac{1}{2}\%$, the only price movement on the day of the Fed announcement will be the

Table 1.5.1 Hypothetical Probabilities of Interest Rate Hike from the Current 6%

Probability	Interest Rate	Share Price
0.2	6.25	95
0.5	6.5	89
0.3	6.75	80

unexpected component. So, if the Fed increases the rate by $\frac{1}{4}\%$, the stock price will increase from \$87.50 to \$95.00 even though it was still a rate hike.

1.5.3 Strong Form Efficiency

This is the extreme version of the EMH. Strong form efficiency states that all information, whether historical, expected, or insider information, is already reflected in the stock price. This means that an investor will not be able to make additional profits, even with insider information. In markets with strict insider-trading regulation, it is probably an overstatement to assume that investors are always correct. In essence, there would have to be a powerful contingent of "smart money" constantly driving the price to its true level. In emerging markets, however, and in markets without restrictions on insider trading, it is not so far-fetched that this "smart money" exists. However, it may be just as far-fetched that this "smart money" is able to drive such volatile markets.

In any of the formulations of the EMH, the underlying result is that profit cannot be made from old news. At any point in time, news is reflected in the stock price, and the price change to the next period will be a function of two things: the required return and the new information or expectations that hit the market. Therefore it should be no surprise that advocates of the EMH find it convenient to model stock price movements as a random walk, as in Section 1.4, with the volatility portion representing unexpected changes in information.

Empirical evidence for or against the EMH is a tricky concept, since it is a hypothesis that states that old information has no effect on stock prices. Usually in empirical tests, researchers look for significant effects, not for the lack thereof. Therefore, disproving the EMH is much easier than proving it. In Chapter 4 we examine a number of anomalies that attempt to disprove the EMH by finding predictable patterns based on old news. These tests usually take the form of

$$r = \alpha + \delta(\text{Anomaly}) + \varepsilon, \tag{1.5.2}$$

where r is the actual stock return, α is the average return, δ represents the excess return for the anomaly in question, and ε is the random error term. If δ is statistically significant, then there is evidence of market inefficiency, since on those anomaly days the market gets a predictable excess return. For a summary of regression analysis, see the Appendix.

Another factor to be taken into account is transactions costs. If a stock price is 2 cents off of its fundamental value, but the brokerage fee is 5 cents per share to take advantage of the discrepancy, it is futile to undertake the transaction and suffer a loss of three cents. The EMH still holds in this case where no transaction takes place, since there are no net profits to be made after the transactions cost is considered.

One method that has been employed to test for evidence of efficiency rather than inefficiency has been an application of the two-sided hypothesis test (Lehmann, 1986) to the EMH by Reagle and Vinod (2003). As we stated above, an excess return less than transactions costs is not necessarily inefficient, since investors have no incentive to trade on the information. Reagle and Vinod (2003) use this feature of the EMH by setting a region around zero where $|\delta|$ is less than transactions costs, and then test that δ falls in this region at a reasonable confidence level. This is akin to a test that the anomaly is not present, and therefore it would be evidence in favor of the EMH. As with the time it takes for information to be incorporated into a stock price, many other anomalies that were common in stock prices a decade ago have gone away as transactions costs have decreased and the volume of information has increased.

If one accepts the EMH, risk becomes the central focus of the investor. All movements other than the average return are unknown. Research and data collection do not aid in predicting these movements along a random walk. So if these risks cannot be avoided, the next step is to measure and quantify the risk involved for an investment. This will be the topic of the next chapter.

APPENDIX: SIMPLE REGRESSION ANALYSIS

When there is a relationship of the form

$$y = \alpha + \beta x + \varepsilon, \tag{1.A.1}$$

where y is a dependent variable that is influenced by x, the independent variable, and the random error is ε, then regression analysis can be employed to estimate the parameters α and β.

The most common method of estimation for regression analysis is ordinary least squares (OLS), which can be used given that the following assumptions hold:

1. The regression model is specified correctly; that is, in the above case y is a linear function of x.
2. ε has a zero mean, a constant variance for all observations, and is uncorrelated between observations.
3. x is given, not correlated with the error term, and in the case where there is more than one independent variable, the independent variables are not highly correlated with each other.
4. There are more observations than the number of parameters being estimated.

These assumptions basically state that the parameters can be estimated from data, and that all the available information is used. These assumptions may be tested, and in most cases the regression model may be modified if one or more assumptions fail. In this appendix we cover OLS estimation and interpretation. For a complete reference on extensions to OLS see Greene (2000) or Mittlehammer, Judge, and Miller (2000).

OLS involves finding the estimates for the parameters that give the smallest squared prediction errors. Since ε is zero mean, our prediction of the dependent variable is

$$\hat{y} = a + bx, \tag{1.A.2}$$

where a and b are the estimates of α and β, respectively.

Prediction error is then the difference between the actual value of the dependent variable y, and the predicted value $\hat{y}$:

$$e = y - \hat{y}. \tag{1.A.3}$$

The OLS solution, then, is the value of a and b that solve the optimization problem:

$$\min_{a,b} \sum_n e_i^2, \tag{1.A.4}$$

where the minimization is with respect to the parameters a and b and where n is the number of observed data points and Σ_n denotes summation from $i = 1$ to $i = n$.

The values of a and b that solve the OLS minimization are

$$b = \frac{\sum_n (y_i - \bar{y})(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2} \quad \text{and} \quad a = \bar{y} - b\bar{x}, \tag{1.A.5}$$

where $\bar{y}$ and $\bar{x}$ denote the average of y and x, respectively. The numerator of b is also known as the covariance between x and y (multiplied by n), and the denominator is identical to the variance of x (multiplied by n). The presence of the covariance is intuitive since b estimates the amount of change in y for a change in x. Dividing by the variance of x discounts this movement of x by its volatility, since high volatility independent variables may move a large distance for small changes in y.

These are the estimated values of the OLS parameters that give the lowest sum of squared prediction errors. They do not give a precise value, however. One can think of the estimates as being in a range around the true value. Given a large enough sample size (over 30 observations) this range can be determined by the normal distribution.

Given that the OLS assumptions hold, the estimates a and b in (1.A.5) are unbiased, meaning that on average they fall around the true parameter value, and they are the "best" estimates in that they have the lowest variance around the true parameter of any other unbiased estimator of a linear relationship. These properties are referred to as BLUE, or best linear unbiased estimator.

Since the estimators will have some error compared to the true parameters, this error can be quantified by the standard deviation of the normal distribution around the true parameter, also known as the estimate's standard error:

$$\sigma_b = \sqrt{\frac{\sum_n e_i^2}{(n-z)\sum_n (x_i - \bar{x})^2}} \tag{1.A.6}$$

From the properties of the normal distribution, 95% of the probability falls within 1.96 standard deviations of the mean. This allows us to construct a 95% confidence interval for the true parameter based on the estimated value

$$b \pm 1.96\sigma_b. \tag{1.A.7}$$

Roughly speaking, the true unknown parameter β will only fall outside of this range 5% of the time. The 5% error is known as the significance level.

A further method of statistical inference using the OLS estimate is hypothesis testing. Hypothesis testing sets up two competing hypothesis about the parameter, and then uses the estimates from the data to choose between them. Usually the accepted theory is used as the null hypothesis H_0, and the null is assumed to be valid unless rejected by the data, in which case the default hypothesis is the alternative hypothesis H_A.

To test the significance of b, the usual null hypothesis is $H_0: \beta = 0$. If we assume the null is true, the observed b should fall within 1.96 standard deviations 95% of the time. Therefore $[b - 1.96\sigma_b, b + 1.96\sigma_b]$ is our acceptance region where the null hypothesis is reasonable (alternatively, b could be divided by the standard error to obtain a z value—or t value in small samples—and compared to $-1.96 < z < 1.96$). If the estimated statistic falls outside this region, the null is not reasonable since this would be a rare event if the true parameter were, zero. Then the null hypothesis would be rejected, and the alternative $H_A: \beta \neq 0$ would be accepted (strictly speaking, not rejected). In the case of rejection of the null, we say that b is statistically significant, or statistically significantly different from zero.

The regression model is often extended to allow for several independent variables in a multiple regression. The interpretation of regression coefficients, their estimators, standard errors, confidence intervals, are analogous to the simple regression above, although a compact solution requires matrix algebra, as we will see in later chapters.