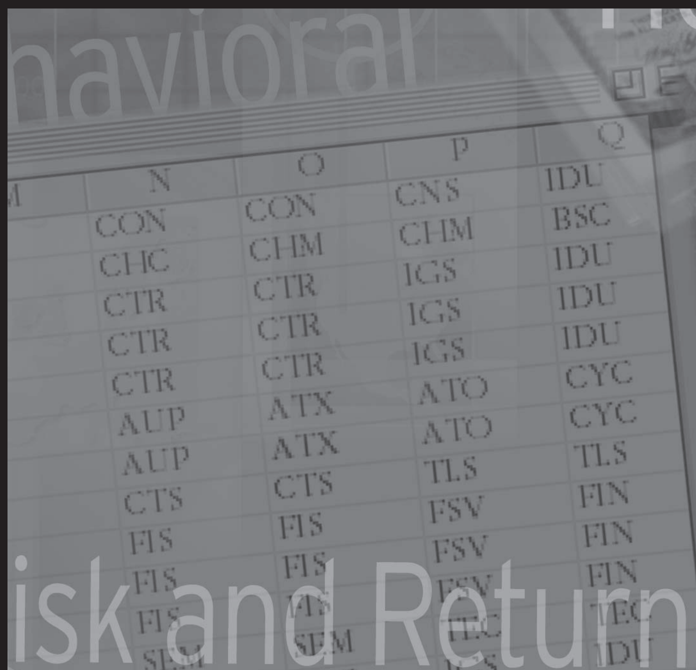
APPENDICES

APPENDIX A

QUANTITATIVE REVIEW

Students in management and investment courses typically come from a variety of backgrounds. Some, who have had strong quantitative training, may feel perfectly comfortable with formal mathematical presentation of material. Others, who have had less technical training, may easily be overwhelmed by mathematical formalism. Most students, however, will benefit from some coaching to make the study of investment easier and more efficient. If you had a good introductory quantitative methods course, and like the text that was used, you may want to refer to it whenever you feel in need of a refresher. If you feel uncomfortable with standard quantitative texts, this reference is for you. Our aim is to present the essential quantitative concepts and methods in a self-contained, nontechnical, and intuitive way. Our approach is structured in line with requirements for the CFA program. The material included is relevant to investment management by the ICFA, the Institute of Chartered Financial Analysts. We hope you find this appendix helpful. Use it to make your venture into investments more enjoyable.



A.1 PROBABILITY DISTRIBUTIONS

Statisticians talk about "experiments," or "trials," and refer to possible outcomes as "events." In a roll of a die, for example, the "elementary events" are the numbers 1 through 6. Turning up one side represents the most disaggregate *mutually exclusive* outcome. Other events are *compound,* that is, they consist of more than one elementary event, such as the result "odd number" or "less than 4." In this case "odd" and "less than 4" are not mutually exclusive. Compound events can be mutually exclusive outcomes, however, such as "less than 4" and "equal to or greater than 4."

In decision making, "experiments" are circumstances in which you contemplate a decision that will affect the set of possible events (outcomes) and their likelihood (probabilities). Decision theory calls for you to identify optimal decisions under various sets of circumstances (experiments), which you may do by determining losses from departures from optimal decisions.

When the outcomes of a decision (experiment) can be quantified, that is, when a numerical value can be assigned to each elementary event, the decision outcome is called a *random variable.* In the context of investment decision making, the random variable (the payoff to the investment decision) is denominated either in dollars or as a percentage rate of return.

The set or list of all possible values of a random variable, *with* their associated probabilities, is called the *probability distribution* of the random variable. Values that are impossible for the random variable to take on are sometimes listed with probabilities of zero. All possible elementary events are assigned values and probabilities, and thus the probabilities have to sum to 1.0.

Sometimes the values of a random variable are *uncountable,* meaning that you cannot make a list of all possible values. For example, suppose you roll a ball on a line and report the distance it rolls before it comes to rest. Any distance is possible, and the precision of the report will depend on the need of the roller and/or the quality of the measuring device. Another uncountable random variable is one that describes the weight of a newborn baby. Any positive weight (with some upper bound) is possible.

We call uncountable probability distributions *continuous,* for the obvious reason that, at least within a range, the possible outcomes (those with positive probabilities) lie anywhere on a continuum of values. Because there is an infinite number of possible values for the random variable in any continuous distribution, such a probability distribution has to be described by a formula that relates the values of the random variable and their associated probabilities, instead of by a simple list of outcomes and probabilities. We discuss continuous distributions later in this section.

Even countable probability distributions can be complicated. For example, on the New York Stock Exchange stock prices are quoted in eighths. This means the price of a stock at some future date is a *countable* random variable. Probability distributions of countable random variables are called *discrete distributions.* Although a stock price cannot dip below zero, it has no upper bound. Therefore a stock price is a random variable that can take on infinitely many values, even though they are countable, and its discrete probability distribution will have to be given by a formula just like a continuous distribution.

There are random variables that are both discrete and finite. When the probability distribution of the relevant random variable is countable and finite, decision making is tractable and relatively easy to analyze. One example is the decision to call a coin toss "heads" or "tails," with a payoff of zero for guessing wrong and 1 for guessing right. The random variable of the decision to guess "heads" has a discrete, finite probability distribution. It can be written as

Event	Value	Probability
Heads	1	.5
Tails	0	.5

This type of analysis usually is referred to as *scenario analysis*. Because scenario analysis is relatively simple, it is used sometimes even when the actual random variable is infinite and uncountable. You can do this by specifying values and probabilities for a set of compound, yet exhaustive and mutually exclusive, events. Because it is simple and has important uses, we handle this case first.

Here is a problem from the 1988 CFA examination.

> Mr. Arnold, an Investment Committee member, has confidence in the forecasting ability of the analysts in the firm's research department. However, he is concerned that analysts may not appreciate risk as an important investment consideration. This is especially true in an increasingly volatile investment environment. In addition, he is conservative and risk averse. He asks for your risk analysis for Anheuser-Busch stock.
>
> 1. Using Table A.1, calculate the following measures of dispersion of returns for Anheuser-Busch stock under each of the three outcomes displayed. Show calculations.
>   *a.* Range.
>   *b.* Variance: $\Sigma \Pr(i)[r_i - E(r)]^2$.
>   *c.* Standard deviation.
>   *d.* Coefficient of variation: $CV = \sigma/E(r)$.
> 2. Discuss the usefulness of each of the four measures listed in quantifying risk.

The examination questions require very specific answers. We use the questions as framework for exposition of scenario analysis.

Table A.1 specifies a three-scenario decision problem. The random variable is the rate of return on investing in Anheuser-Busch stock. However, the third column, which specifies the value of the random variable, does not say simply "Return"—it says "Expected Return." This tells us that the scenario description is a compound event consisting of many elementary events, as is almost always the case. We streamline or simplify reality in order to gain tractability.

Analysts who prepare input lists must decide on the number of scenarios with which to describe the entire probability distribution, as well as the rates of return to allocate to each one. This process calls for determining the probability of occurrence of each scenario, *and* the expected rate of return *within* (conditional on) each scenario, which governs the outcome of each scenario. Once you become familiar with scenario analysis, you will be able to build a simple scenario description from any probability distribution.

Expected Returns

The expected value of a random variable is the answer to the question, What would be the average value of the variable if the "experiment" (the circumstances and the decision) were repeated infinitely? In the case of an investment decision, your answer is meant to describe the reward from making the decision.

Table A.1 Anheuser-Busch Companies, Inc., Dispersion of Potential Returns

Outcome	Probability	Expected Return*
Number 1	.20	20%
Number 2	.50	30
Number 3	.30	50

*Assume for the moment that the expected return in each scenario will be realized with certainty. This is the way returns were expressed in the original question.

Note that the question is hypothetical and abstract. It is hypothetical because, practically, the exact circumstances of a decision (the "experiment") often cannot be repeated even once, much less infinitely. It is abstract because, even if the experiment were to be repeated many times (short of infinitely), the *average* rate of return may not be one of the possible outcomes. To demonstrate, suppose that the probability distribution of the rate of return on a proposed investment project is +20% or −20%, with equal probabilities of .5. Intuition indicates that repeating this investment decision will get us ever closer to an average rate of return of zero. But a one-time investment cannot produce a rate of return of zero. Is the "expected" return still a useful concept when the proposed investment represents a one-time decision?

One argument for using expected return to measure the reward from making investment decisions is that, although a specific investment decision may be made only once, the decision maker will be making many (although different) investment decisions over time. Over time, then, the average rate of return will come close to the average of the expected values of all the individual decisions. Another reason for using the expected value is that admittedly we lack a better measure.[1]

The probabilities of the scenarios in Table A.1 predict the relative frequencies of the outcomes. If the current investment in Anheuser-Busch could be replicated many times, a 20% return would occur 20% of the time, a 30% return would occur 50% of the time, and 50% return would occur the remaining 30% of the time. This notion of probabilities and the definition of the expected return tells us how to calculate the expected return.[2]

$$E(r) = .20 \times .20 + .50 \times .30 + .30 \times .50 = .34 \text{ (or 34\%)}$$

Labeling each scenario $i = 1, 2, 3$, and using the summation sign, Σ, we can write the formula for the expected return:

$$\begin{aligned} \mathrm{E}(r) &= \Pr(1)r_1 + \Pr(2)r_2 + \Pr(3)r_3 \\ &= \sum_{i=1}^{3} \Pr(i)r_i \end{aligned} \tag{A.1}$$

The definition of the expectation in equation A.1 reveals two important properties of random variables. First, if you add a constant to a random variable, its expectation is also increased by the same constant. If, for example, the return in each scenario in Table A.1 were increased by 5%, the expectation would increase to 39%. Try this, using equation A.1. If a random variable is multiplied by a constant, its expectation will change by that same proportion. If you multiply the return in each scenario by 1.5, $E(r)$ would change to $1.5 \times .34 = .51$ (or 51%).

Second, the deviation of a random variable from its expected value is itself a random variable. Take any rate of return r_1 in Table A.1 and define its deviation from the expected value by

$$d_i = r_i - E(r)$$

What is the expected value of d? $E(d)$ is the expected deviation from the expected value, and by equation A.1 it is necessarily zero because

$$\begin{aligned} E(d) &= \Sigma \Pr(i)d_i = \Sigma \Pr(i)[r_i - E(r)] \\ &= \Sigma \Pr(i)r_i - E(r)\,\Sigma \Pr(i) \\ &= E(r) - E(r) = 0 \end{aligned}$$

[1] Another case where we use a less-than-ideal measure is the case of yield to maturity on a bond. The YTM measures the rate of return from investing in a bond *if* it is held to maturity and *if* the coupons can be reinvested at the same yield to maturity over the life of the bond.

[2] We will consistently perform calculations in decimal fractions to avoid confusion.

Measures of Dispersion

The Range. Assume for a moment that the expected return for each scenario in Table A.1 will be realized with certainty in the event that scenario occurs. Then the set of possible return outcomes is unambiguously 20%, 30%, and 50%. The *range* is the difference between the maximum and the minimum values of the random variable, 50% − 20% = 30% in this case. Range is clearly a crude measure of dispersion. Here it is particularly inappropriate because the scenario returns themselves are given as expected values, and therefore the true range is unknown. There is a variant of the range, the *interquartile range,* that we explain in the discussion of descriptive statistics.

The Variance. One interpretation of variance is that it measures the "expected surprises." Although that may sound like a contradiction in terms, it really is not. First, think of a surprise as a deviation from expectation. The surprise is not in the *fact* that expectation has not been realized, but rather in the *direction* and *magnitude* of the deviation.

The example in Table A.1 leads us to *expect* a rate of return of 34% from investing in Anheuser-Busch stock. A second look at the scenario returns, however, tells us that we should stand ready to be surprised because the probability of earning exactly 34% is zero. Being sure that our expectation will not be realized does not mean that we can be sure what the realization is going to be. The element of surprise lies in the direction and magnitude of the deviation of the actual return from expectation, and that is the relevant random variable for the measurement of uncertainty. Its probability distribution adds to our understanding of the nature of the uncertainty that we are facing.

We measure the reward by the expected return. Intuition suggests that we measure uncertainty by the expected *deviation* of the rate of return from expectation. We showed in the previous section, however, that the expected deviation from expectation must be zero. Positive deviations, when weighted by probabilities, are exactly offset by negative deviations. To get around this problem, we replace the random variable "deviation from expectations" (denoted earlier by d) with its square, which must be positive even if d itself is negative.

We define the *variance,* our measure of surprise or dispersion, by the *expected squared deviation of the rate of return from its expectation.* With the Greek letter sigma square denoting variance, the formal definition is

$$\sigma^2(r) = E(d^2) = E[r_i - E(r)]^2 = \Sigma \Pr(i)[r_i - E(r)]^2 \quad \text{(A.2)}$$

Squaring each deviation eliminates the sign, which eliminates the offsetting effects of positive and negative deviations.

In the case of Anheuser-Busch, the variance of the rate of return on the stock is

$$\sigma^2(r) = .2(.20 - .34)^2 + .5(.30 - .34)^2 + .3(.50 - .34)^2 = .0124$$

Remember that if you add a constant to a random variable, the variance does not change at all. This is because the expectation also changes by the same constant, and hence deviations from expectation remain unchanged. You can test this by using the data from Table A.1.

Multiplying the random variable by a constant, however, *will* change the variance. Suppose that each return is multiplied by the factor k. The new random variable, kr, has expectation of $E(kr) = kE(r)$. Therefore, the deviation of kr from its expectation is

$$d(kr) = kr - E(kr) = kr - kE(r) = k[r - E(r)] = kd(r)$$

If each deviation is multiplied by k, the squared deviations are multiplied by the square of k:

$$\sigma^2(kr) = k^2\sigma^2(r)$$

To summarize, adding a constant to a random variable does not affect the variance. Multiplying a random variable by a constant, though, will cause the variance to be multiplied by the square of that constant.

The Standard Deviation. A closer look at the variance will reveal that its dimension is different from that of the expected return. Recall that we squared deviations from the expected return in order to make all values positive. This alters the *dimension* (units of measure) of the variance to "square percents." To transform the variance into terms of percentage return, we simply take the square root of the variance. This measure is the *standard deviation.* In the case of Anheuser-Busch's stock return, the standard deviation is

$$\sigma = (\sigma^2)^{1/2} = \sqrt{.0124} = .1114 \text{ (or 11.14\%)} \tag{A.3}$$

Note that you always need to calculate the variance first before you can get the standard deviation. The standard deviation conveys the same information as the variance but in a different form.

We know already that adding a constant to r will not affect its variance, and it will not affect the standard deviation either. We also know that multiplying a random variable by a constant multiplies the variance by the square of that constant. From the definition of the standard deviation in equation A.3, it should be clear that multiplying a random variable by a constant will multiply the standard deviation by the (absolute value of this) constant. The absolute value is needed because the sign of the constant is lost through squaring the deviations in the computation of the variance. Formally,

$$\sigma(kr) = \text{Abs}(k)\sigma(r)$$

Try a transformation of your choice using the data in Table A.1.

The Coefficient of Variation. To evaluate the magnitude of dispersion of a random variable, it is useful to compare it to the expected value. The ratio of the standard deviation to the expectation is called the *coefficient of variation.* In the case of returns on Anheuser-Busch stock, it is

$$CV = \frac{\sigma}{E(r)} = \frac{.1114}{.3400} = .3285 \tag{A.4}$$

The standard deviation of the Anheuser-Busch return is about one-third of the expected return (reward). Whether this value for the coefficient of variation represents a big risk depends on what can be obtained with alternative investments.

The coefficient of variation is far from an ideal measure of dispersion. Suppose that a plausible expected value for a random variable is zero. In this case, regardless of the magnitude of the standard deviation, the coefficient of variation will be infinite. Clearly, this measure is not applicable in all cases. Generally, the analyst must choose a measure of dispersion that fits the particular decision at hand. In finance, the standard deviation is the measure of choice in most cases where overall risk is concerned. (For individual assets, the measure β, explained in the text, is the measure used.)

Skewness

So far, we have described the measures of dispersion as indicating the size of the average surprise, loosely speaking. The standard deviation is not exactly equal to the average surprise though, because squaring deviations and then taking the square root of the average square deviation results in greater weight (emphasis) placed on larger deviations. Other than that, it is simply a measure that tells us how big a deviation from expectation can be expected.

Most decision makers agree that the expected value and standard deviation of a random variable are the most important statistics. However, once we calculate them another question about risk (the nature of the random variable describing deviations from expectations) is pertinent: Are the larger deviations (surprises) more likely to be positive? Risk-averse decision makers worry about bad surprises, and the standard deviation does not distinguish good from bad ones. Most risk avoiders are believed to prefer random variables with likely *small negative surprises* and *less* likely *large positive surprises,* to the reverse, likely *small good surprises* and *less* likely *large bad surprises.* More than anything, risk is really defined by the possibility of disaster (large bad surprises).

One measure that distinguishes between the likelihood of large good-vs.-bad surprises is the "third moment." It builds on the behavior of deviations from the expectation, the random variable we have denoted by d. Denoting the *third moment* by M_3, we define it:

$$M_3 = E(d^3) = E[r_i - E(r)]^3 = \Sigma \Pr(i)[r_i - E(r)]^3 \quad \text{(A.5)}$$

Cubing each value of d (taking it to the third power) magnifies larger deviations more than smaller ones. Raising values to an odd power causes them to retain their sign. Recall that the sum of all deviations multiplied by their probabilities is zero because positive deviations weighted by their probabilities exactly offset the negative. When *cubed* deviations are multiplied by their probabilities and then added up, however, large deviations will dominate. The sign will tell us in this case whether *large positive* deviations dominate (positive M_3) or whether *large negative* deviations dominate (negative M_3).

Incidentally, it is obvious why this measure of skewness is call the third moment; it refers to cubing. Similarly, the variance is often referred to as the second moment because it requires squaring.

Returning to the investment decision described in Table A.1, with the expected value of 34%, the third moment is

$$M_3 = .2(.20 - .34)^3 + .5(.30 - .34)^3 + .3(.50 - .34)^3 = .000648$$

The sign of the third moment tells us that larger *positive* surprises dominate in this case. You might have guessed this by looking at the deviations from expectation and their probabilities; that is, the most likely event is a return of 30%, which makes for a small negative surprise. The other negative surprise (20% − 34% = −14%) is smaller in magnitude than the positive surprise (50% − 34% = 16%) *and* is also *less* likely (probability .20) relative to the positive surprise, 30% (probability .30). The difference appears small, however, and we do not know whether the third moment may be an important issue for the decision to invest in Anheuser-Busch.

It is difficult to judge the importance of the third moment, here .000648, without a benchmark. Following the same reasoning we applied to the standard deviation, we can take the *third root* of M_3 (which we denote m_3) and compare it to the standard deviation. This yields $m_3 = .0865 = 8.65\%$, which is not trivial compared with the standard deviation (11.14%).

Another Example: Options on Anheuser-Busch Stock

Suppose that the current price of Anheuser-Busch stock is $30. A call option on the stock is selling for 60 cents, and a put is selling for $4. Both have an exercise price of $42 and maturity date to match the scenarios in Table A.1.

The call option allows you to buy the stock at the exercise price. You will choose to do so if the call ends up "in the money," that is, the stock price is above the exercise price. The profit in this case is the difference between the stock price and the exercise price, less cost of the call. Even if you exercise the call, your profit may still be negative if the cash flow from the exercise of the call does not cover the initial cost of the call. If the call ends up "out of the money," that is, the stock price is below the exercise price, you will let the call expire worthless and suffer a loss equal to the cost of the call.

The put option allows you to sell the stock at the exercise price. You will choose to do so if the put ends up "in the money," that is, the stock price is below the exercise price.

Your profit is then the difference between the exercise price and the stock price, less the initial cost of the put. Here again, if the cash flow is not sufficient to cover the cost of the put, the investment will show a loss. If the put ends up "out of the money," you again let it expire worthless, taking a loss equal to the initial cost of the put.

The scenario analysis of these alternative investments is described in Table A.2.

The expected rates of return on the call and put are

$$E(r_{\text{call}}) = .2(-1) + .5(-1) + .3(4) = .5 \text{ (or 50\%)}$$
$$E(r_{\text{put}}) = .2(.5) + .5(-.25) + .3(-1) = -.325 \text{ (or } -32.5\%)$$

The negative expected return on the put may be justified by the fact that it is a hedge asset, in this case an insurance policy against losses from holding Anheuser-Busch stock. The variance and standard deviation of the two investments are

$$\sigma^2_{\text{call}} = .2(-1 - .5)^2 + .5(-1 - .5)^2 + .3(4 - .5)^2 = 5.25$$
$$\sigma^2_{\text{put}} = .2[.5 - (-.325)]^2 + .5[-.25 - (-.325)]^2 + .3[-1 - (-.325)]^2 = .2756$$
$$\sigma_{\text{call}} = \sqrt{5.25} = 2.2913 \text{ (or 229.13\%)}$$
$$\sigma_{\text{put}} = \sqrt{.2756} = .525 \text{ (or 52.5\%)}$$

These are very large standard deviations. Comparing the standard deviation of the call's return to its expected value, we get the coefficient of variation:

$$\text{CV}_{\text{call}} = \frac{2.2913}{.5} = 4.5826$$

Refer back to the coefficient of variation for the stock itself, .3275, and it is clear that these instruments have high standard deviations. This is quite common for stock options. The negative expected return of the put illustrates again the problem in interpreting the magnitude of the "surprise" indicated by the coefficient of variation.

Moving to the third moments of the two probability distributions:

$$M_3(\text{call}) = .2(-1 - .5)^3 + .5(-1 - .5)^3 + .3(4 - .5)^3 = 10.5$$
$$M_3(\text{put}) = .2[.5 - (-.325)]^3 + .5[-.25 - (-.325)]^3 + .3[-1 - (-.325)]^3$$
$$= .02025$$

Table A.2 Scenario Analysis for Investment in Options on Anheuser-Busch Stock

	Scenario 1	Scenario 2	Scenario 3
Probability	.20	.50	.30
Event			
1. Return on stock	20%	30%	50%
Stock price (initial price = $30)	$36.00	$39.00	$45.00
2. Cash flow from call (exercise price = $42)	0	0	$3.00
Call profit (initial price = $.60)	−$.60	−$.60	$2.40
Call rate of return	−100%	−100%	400%
3. Cash flow from put (exercise price = $42)	$6.00	$3.00	0
Put profit (initial price = $4)	$2.00	−$1.00	−$4.00
Put rate of return	50%	−25%	−100%

Both instruments are positively skewed, which is typical of options and one part of their attractiveness. In this particular circumstance the call is more skewed than the put. To establish this fact, note the third root of the third moment:

$$m_3(\text{call}) = M_3(\text{call})^{1/3} = 2.1898 \text{ (or } 218.98\%)$$
$$m_3(\text{put}) = .02^{1/3} = .2725 \text{ (or } 27.25\%)$$

Compare these figures to the standard deviations, 229.13% for the call and 52.5% for the put, and you can see that a large part of the standard deviation of the option is driven by the possibility of large good surprises instead of by the more likely, yet smaller, bad surprises.[3]

So far we have described discrete probability distributions using scenario analysis. We shall come back to decision making in a scenario analysis framework in Section A.3 on multivariate statistics.

Continuous Distributions: Normal and Lognormal Distributions

When a compact scenario analysis is possible and acceptable, decisions may be quite simple. Often, however, so many relevant scenarios must be specified that scenario analysis is impossible for practical reasons. Even in the case of Anheuser-Busch, as we were careful to specify, the individual scenarios considered actually represented compound events.

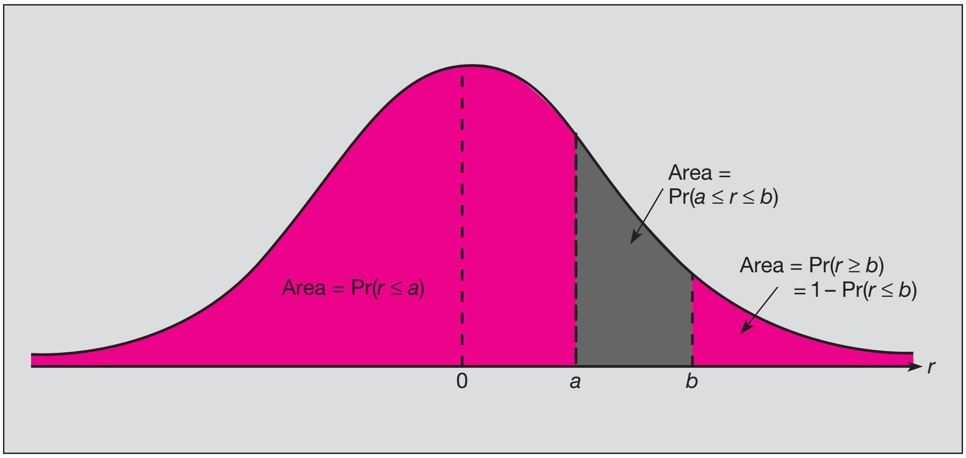
When many possible values of rate of return have to be considered, we must use a formula that describes the probability distribution (relates values to probabilities). As we noted earlier, there are two types of probability distributions: discrete and continuous. Scenario analysis involves a discrete distribution. However, the two most useful distributions in investments, the normal and lognormal, are continuous. At the same time they are often used to approximate variables with distributions that are known to be discrete, such as stock prices. The probability distribution of future prices and returns is discrete—prices are quoted in eighths. Yet the industry norm is to approximate these distributions by the normal or lognormal distribution.

Standard Normal Distribution. The normal distribution, also known as Gaussian (after the mathematician Gauss) or bell-shaped, describes random variables with the following properties and is shown in Figure A.1:

- The expected value is the mode (the most frequent elementary event) and also the median (the middle value in the sense that half the elementary events are greater and half smaller). Note that the expected value, unlike the median or mode, requires weighting by probabilities to produce the concept of central value.
- The normal probability distribution is symmetric around the expected value. In other words, the likelihood of equal absolute-positive and negative deviations from expectation is equal. Larger deviations from the expected value are less likely than are smaller deviations. In fact, the essence of the normal distribution is that the probability of deviations decreases exponentially with the magnitude of the deviation (positive and negative alike).
- A normal distribution is identified completely by two parameters, the expected value and the standard deviation. The property of the normal distribution that

[3] Note that the expected return of the put is −32.5%; hence the worst surprise is −67.5%, and the best is 82.5%. The middle scenario is also a positive deviation of 7.5% (with a high probability of .50). These two elements explain the positive skewness of the put.

Figure A.1 **Probabilities under the normal density**



makes it most convenient for portfolio analysis is that any weighted sum or normally distributed random variables produce a random variable that also is normally distributed. This property is called *stability.* It is also true that if you add a constant to a "normal" random variable (meaning a random variable with a normal probability distribution) or multiply it by a constant, then the transformed random variable also will be normally distributed.

Suppose that n is any random variable (not necessarily normal) with expectation μ and standard deviation σ. As we showed earlier, if you add a constant c to n, the standard deviation is not affected at all, but the mean will change to $\mu + c$. If you multiply n by a constant b, its mean and standard deviation will change by the same proportion to $b\mu$ and $b\sigma$. If n is normal, the transformed variable also will be normal.

Stability, together with the property that a normal variable is completely characterized by its expectation and standard deviation, implies that if we know one normal probability distribution with a given expectation and standard deviation, we know them all.

Subtracting the expected value from each observation and then dividing by the standard deviation we obtain the *standard normal distribution,* which has an expectation of zero and both variance and standard deviation equal to 1.0. Formally, the relationship between the value of the standard normal random variable, z, and its probability, f, is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \quad \text{(A.6)}$$

where "exp" is the quantity e to the power of the expression in the parentheses. The quantity e is an important number just like the well-known π, which also appears in the function. It is important enough to earn a place on the keyboard of your financial calculator, mostly because it is used also in continuous compounding.

Probability functions of continuous distributions are called *densities* and denoted by f, rather than by the "Pr" of scenario analysis. The reason is that the probability of any of the infinitely many possible values of z is infinitesimally small. Density is a function that allows us to obtain the probability of a *range of values* by integrating it over a desired range. In other words, whenever we want the probability that a standard normal variate (a random variable) will fall in the range from $z = a$ to $z = b$, we have to add up the density values, $f(z)$, for all zs from a to b. There are infinitely many zs in that range, regardless how close a is to b. *Integration* is the mathematical operation that achieves this task.

Consider first the probability that a standard normal variate will take on a value less than or equal to a, that is, z is in the range $[-\infty, a]$. We have to integrate the density from ∞ to a. The result is called the *cumulative (normal) distribution,* denoted by $N(a)$. When a approaches infinity, any value is allowed for z; hence the probability that z will end up in that range approaches 1.0. It is a property of any density that when it is integrated over the entire range of the random variable, the cumulative distribution is 1.0.

In the same way, the probability that a standard normal variate will take on a value less than or equal to b is $N(b)$. The probability that a standard normal variate will take on a value in the range $[a, b]$ is just the difference between $N(b)$ and $N(a)$. Formally,

$$\Pr(a \le z \le b) = N(b) - N(a)$$

These concepts are illustrated in Figure A.1. The graph shows the normal density. It demonstrates the symmetry of the normal density around the expected value (zero for the standard normal variate, which is also the mode and the median), and the smaller likelihood of larger deviations from expectation. As is true for any density, the entire area under the density graph adds up to 1.0. The values a and b are chosen to be positive, so they are to the right of the expected value. The leftmost colored shaded area is the proportion of the area under the density for which the value of z is less than or equal to a. Thus this area yields the cumulative distribution for a, the probability that z will be smaller than or equal to a. The gray shaded area is the area under the density graph between a and b. If we add that area to the cumulative distribution of a, we get the entire area up to b, that is, the probability that z will be anywhere to the left of b. Thus the area between a and b has to be the probability that z will fall between a and b.

Applying the same logic, we find the probability that z will take on a value greater than b. We know already that the probability that z will be smaller than or equal to b is $N(b)$. The compound events "smaller than or equal to b" and "greater than b" are mutually exclusive *and* "exhaustive," meaning that they include all possible outcomes. Thus their probabilities sum to 1.0, and the probability that z is greater than b is simply equal to one minus the probability that z is less than or equal to b. Formally, $\Pr(z > b) = 1 - N(b)$.

Look again at Figure A.1. The area under the density graph between b and infinity is just the difference between the entire area under the graph (equal to 1.0) and the area between minus infinity and b, that is, $N(b)$.

The normal density is sufficiently complex that its cumulative distribution, its integral, does not have an exact formulaic closed-form solution. It must be obtained by numerical (approximation) methods. These values are produced in tables that give the value $N(z)$ for any z, such as Table 21.2 of this text.

To illustrate, let us find the following probabilities for a standard normal variate:

$$\Pr(z \le -.36) = N(-.36) = \text{Probability that } z \text{ is less than or equal to } -.36$$
$$\Pr(z \le .94) = N(.94) = \text{Probability that } z \text{ is less than or equal to } .94$$
$$\Pr(-.36 \le z \le .94) = N(.94) - N(-.36) = \text{Probability that } z \text{ will be in the range } [-.36, .94]$$
$$\Pr(z > .94) = 1 - N(.94) = \text{Probability that } z \text{ is greater than } .94$$

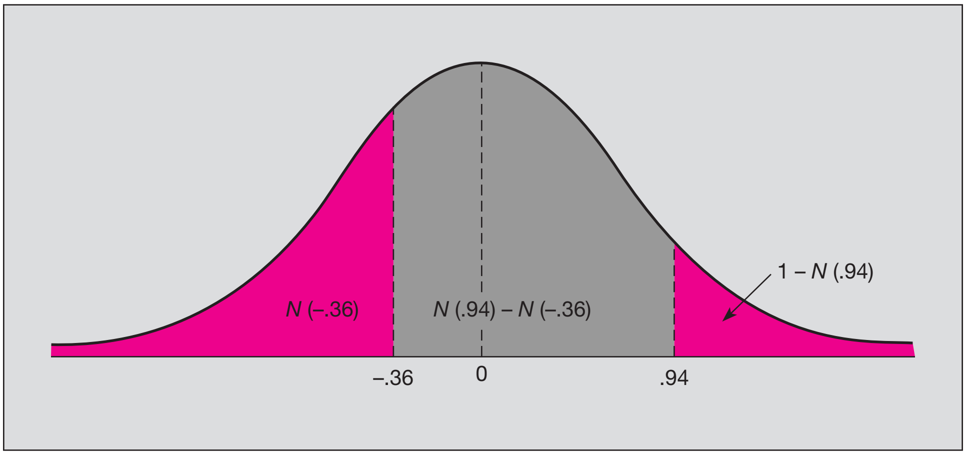
Use Table 21.2 of the cumulative standard normal (sometimes called the area under the normal density) and Figure A.2. The table shows that

$$N(-.36) = .3594$$
$$N(.94) = .8264$$

In Figure A.2 the area under the graph between $-.36$ and $.94$ is the probability that z will fall between $-.36$ and $.94$. Hence

$$\Pr(-.36 \le z \le .94) = N(.94) - N(-.36) = .8264 - .3594 = .4670$$

Figure A.2 Probabilities and the cumulative normal distribution



The probability that z is greater than .94 is the area under the graph in Figure A.2, between .94 and infinity. Thus it is equal to the entire area (1.0) less the area from minus infinity to .94. Hence

$$\Pr(z > .94) = 1 - N(.94) = 1 - .8264 = .1736$$

Finally, one can ask, What is the value a for which z will be smaller than or equal to a with probability P? The notation for the function that yields the desired value of a is $\Phi(P)$, so that

$$\text{If } \Phi(P) = a\text{, then } P = N(a) \qquad \text{(A.7)}$$

For instance, suppose the question is, Which value has a cumulative density of .50? A glance at Figure A.2 reminds us that the area between minus infinity and zero (the expected value) is .5. Thus we can write

$$\Phi(.5) = 0\text{, because } N(0) = .5$$

Similarly,

$$\Phi(.8264) = .94\text{, because } N(.94) = .8264$$

and

$$\Phi(.3594) = -.36$$

For practice, confirm with Table 21.2 that $\Phi(.6554) = .40$, meaning that the value of z with a cumulative distribution of .6554 is $z = .40$.

Nonstandard Normal Distributions. Suppose that the monthly rate of return on a stock is closely approximated by a normal distribution with a mean of .015 (1.5% per month), and standard deviation of .127 (12.7% per month). What is the probability that the rate of return will fall below zero in a given month? Recall that because the rate is a normal variate, its cumulative density has to be computed by numerical methods. The standard normal table can be used for any normal variate.

Any random variable, x, may be transformed into a new standardized variable, x^*, by the following rule:

$$x^* = \frac{x - E(x)}{\sigma(x)} \qquad \text{(A.8)}$$

Note that all we have done to x was (1) *subtract* its expectation and (2) *multiply* by one over its standard deviation, $1/[\sigma(x)]$. According to our earlier discussion, the effect of transforming a random variable by adding and multiplying by a constant is such that the expectation and standard deviation of the transformed variable are

$$E(x^*) = \frac{E(x) - E(x)}{\sigma(x)} = 0; \qquad \sigma(x^*) = \frac{\sigma(x)}{\sigma(x)} = 1 \tag{A.9}$$

From the stability property of the normal distribution we also know that if x is normal, so is x^*. A normal variate is characterized completely by two parameters: its expectation and standard deviation. For x^*, these are zero and 1.0, respectively. When we subtract the expectation and then divide a normal variate by its standard deviation, we standardize it; that is, we transform it to a standard normal variate. This trick is used extensively in working with normal (and approximately normal) random variables.

Returning to our stock, we have learned that if we subtract .015 and then divide the monthly returns by .127, the resultant random variable will be standard normal. We can now determine the probability that the rate of return will be zero or less in a given month. We know that

$$z = \frac{r - .015}{.127}$$

where z is standard normal and r the return on our stock. Thus if r is zero, z has to be

$$z(r = 0) = \frac{0 - .015}{.127} = -.1181$$

For r to be zero, the corresponding standard normal has to be -11.81%, a negative number. The event "r will be zero or less" is identical to the event "z will be $-.1181$ or less." Calculating the probability of the latter will solve our problem. That probability is simply $N(-.1181)$. Visit the standard normal table and find that

$$\Pr(r \leq 0) = N(-.1181) = .5 - .047 = .453$$

The answer makes sense. Recall that the expectation of r is 1.5%. Thus, whereas the probability that r will be 1.5% or less is .5, the probability that it will be *zero* or less has to be close, but somewhat lower.

Confidence Intervals. Given the large standard deviation of our stock, it is logical to be concerned about the likelihood of extreme values for the monthly rate of return. One way to quantify this concern is to ask: What is the interval (range) within which the stock return will fall in a given month, with a probability of .95? Such an interval is called the *95% confidence interval.*

Logic dictates that this interval be centered on the expected value, .015, because r is a normal variate (has a normal distribution) which is symmetric around the expectation. Denote the desired interval by

$$[E(r) - a, E(r) + a] = [.015 - a, .015 + a]$$

which has a length of $2a$. The probability that r will fall within this interval is described by the following expression:

$$\Pr(.015 - a \leq r \leq .015 + a) = .95$$

To find this probability, we start with a simpler problem, involving the *standard normal variate,* that is, a normal with expectation of zero and standard deviation of 1.0.

What is the 95% confidence interval for the standard normal variate, z? The variable will be centered on zero, so the expression is

$$\Pr(-a^* \leq z \leq a^*) = N(a^*) - N(-a^*) = .95$$

You might best understand the substitution of the difference of the appropriate cumulative distributions for the probability with the aid of Figure A.3. The probability of falling outside of the interval is $1 - .95 = .05$. By the symmetry of the normal distribution, z will be equal to or less than $-a^*$ with probability of .025, and with probability .025, z will be greater than a^*. Thus we solve for a^* using

$$-a^* = \Phi(.025), \text{ which is equivalent to } N(-a^*) = .025$$

We can summarize the chain that we have pursued so far as follows. If we seek a $P = .95$ level confidence interval, we define α as the probability that r will fall outside the confidence interval. Because of the symmetry, α will be split so that half of it is the probability of falling to the right of the confidence interval, while the other half is the probability of falling to the left of the confidence interval. Therefore, the relation between α and P is

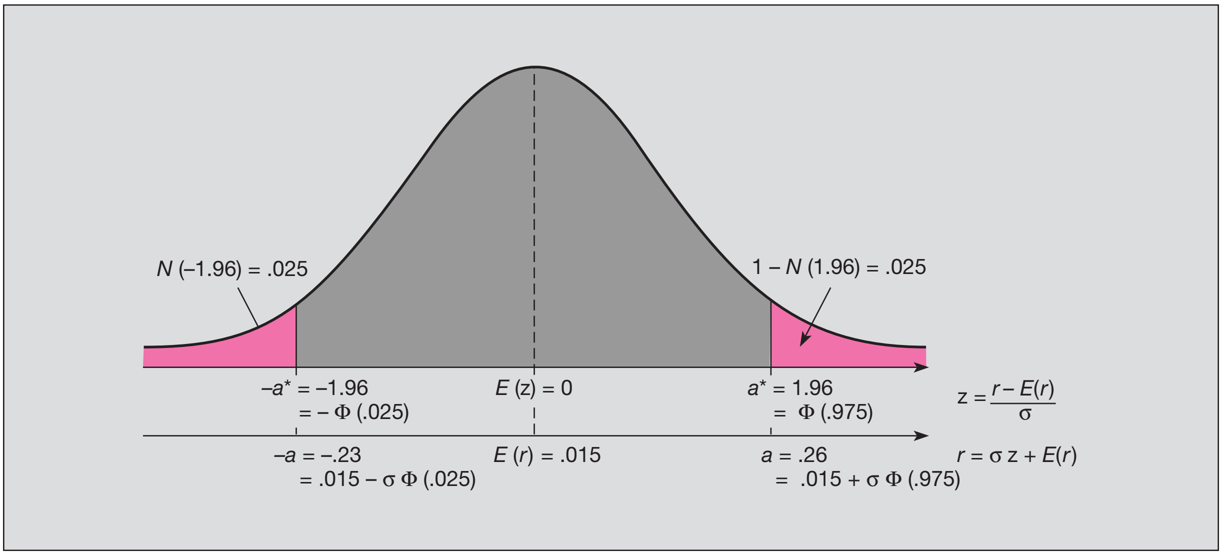
$$\alpha = 1 - P = .05; \qquad \frac{\alpha}{2} = \frac{1-P}{2} = .025$$

We use $\alpha/2$ to indicate that the area that is excluded for r is equally divided between the tails of the distributions. Each tail that is excluded for r has an area of $\alpha/2$. The value $\alpha = 1 - P$ represents the entire value that is excluded for r.

To find $z = \Phi(\alpha/2)$, which is the lower boundary of the confidence interval for the standard normal variate, we have to locate the z value for which the standard normal cumulative distribution is .025, finding $z = -1.96$. Thus we conclude that $-a^* = -1.96$ and $a^* = 1.96$. The confidence interval for z is

$$\begin{aligned}[E(z) - \Phi(\alpha/2), E(z) + \Phi(\alpha/2)] &= [-\Phi(.025), \Phi(.025)] \\ &= [-1.96, .196]\end{aligned}$$

Figure A.3 Confidence intervals and the standard normal density



To get the interval boundaries for the nonstandard normal variate r, we transform the boundaries for z by the usual relationship, $r = z\sigma(r) + E(r) = \Phi(\alpha/2)\sigma(r) + E(r)$. Note that all we are doing is setting the expectation at the center of the confidence interval and extending it by a number of standard deviations. The number of standard deviations is determined by the probability that we allow for falling outside the confidence interval (α), or, equivalently, the probability of falling in it (P). Using minus and plus 1.96 for $z = \pm\,\Phi(Ø/.25)$, the distance on each side of the expectation is $\pm 1.96 \times .127 = .249$. Thus we obtain the confidence interval

$$[E(r) - \sigma(r)\Phi(\alpha/2), E(r) + \sigma(r)\Phi(\alpha/2)] = [E(r) - .249, E(r) + .249]$$
$$= [-.234, .264]$$

so that

$$P = 1 - \alpha = \Pr[E(r) - \sigma(r)\Phi(\alpha/2) \leq r \leq E(r) + \sigma(r)\Phi(\alpha/2)]$$

which, for our stock (with expectation .015 and standard deviation .127), amounts to

$$\Pr[-.234 \leq r \leq .264] = .95$$

Note that because of the large standard deviation of the rate of return on the stock, the 95% confidence interval is 49% wide.

To reiterate with a variation on this example, suppose we seek a 90% confidence interval for the annual rate of return on a portfolio, r_p, with a monthly expected return of 1.2% and standard deviation of 5.2%.

The solution is simply

$$\Pr\left[E(r) - \sigma(r)\,\Phi\left(\frac{1-P}{2}\right) \leq r_p \leq E(r) + \sigma\,(r)\,\Phi\left(\frac{1-P}{2}\right)\right]$$
$$= \Pr[(.012 - .052 \times 1.645) \leq r_p \leq (.012 + .052 \times 1.645)]$$
$$= \Pr[-.0735 \leq r_p \leq .0975] = .90$$

Because the portfolio is of low risk this time (and we require only a 90% rather than a 95% probability of falling within the interval), the 90% confidence interval is only 2.4% wide.

The Lognormal Distribution. The normal distribution is not adequate to describe stock prices and returns for two reasons. First, whereas the normal distribution admits any value, including negative values, actual stock prices cannot be negative. Second, the normal distribution does not account for compounding. The lognormal distribution addresses these two problems.

The lognormal distribution describes a random variable that grows, *every instant,* by a rate that is a normal random variable. Thus the progression of a lognormal random variable reflects continuous compounding.

Suppose that the *annual continuously compounded* (ACC) rate of return on a stock is normally distributed with expectation $\mu = .12$ and standard deviation $\sigma = .42$. The stock price at the beginning of the year is $P_0 = \$10$. With continuous compounding (see appendix to Chapter 5), if the ACC rate of return, r_C, turns out to be .23, then the end-of-year price will be

$$P_1 = P_0\exp(r_C) = 10e^{.23} = \$12.586$$

representing an effective annual rate of return of

$$r = \frac{P_1 - P_0}{P_0} = e^{r_C} - 1 = .2586 \text{ (or 25.86\%)}$$

This is the practical meaning of r, the annual rate on the stock, being lognormally distributed. Note that however negative the ACC rate of return (r_C) is, the price, P_1, cannot become negative.

Two properties of lognormally distributed financial assets are important: their expected return and the allowance for changes in measurement period.

Expected Return of a Lognormally Distributed Asset. The expected annual rate of return of a lognormally distributed stock (as in our example) is

$$E(r) = \exp(\mu + \sigma^2/2) - 1 = \exp(.12 + .42^2/2) - 1 = e^{.2082} - 1$$
$$= .2315 \text{ (or 23.15\%)}$$

This is just a statistical property of the distribution. For this reason, a useful statistic is

$$\mu^* = \mu + \sigma^2/2 = .2082$$

When analysts refer to the expected ACC return on a lognormal asset, frequently they are really referring to μ^*. Often the asset is said to have a normal distribution of the ACC return with expectation μ^* and standard deviation σ.

Change of Frequency of Measured Returns. The lognormal distribution allows for easy change of the holding period of returns. Suppose that we want to calculate returns monthly instead of annually. We use the parameter t to indicate the fraction of the year that is desired; in the case of monthly periods, $t = 1/12$. To transform the annual distribution to a t-period (monthly) distribution, it is necessary merely to multiply the expectation and variance of the ACC return by t (in this case, 1/12).

The monthly continuously compounded return on the stock in our example has the expectation and standard deviation of

$$\mu(\text{monthly}) = .12/12 = .01 \text{ (1\% per month)}$$
$$\sigma(\text{monthly}) = .42/\sqrt{12} = .1212 \text{ (or 12.12\% per month)}$$
$$\mu^*(\text{monthly}) = .2082/12 = .01735 \text{ (or 1.735\% per month)}$$

Note that we divide variance by 12 when changing from annual to monthly frequency; the standard deviation therefore is divided by the square root of 12.

Similarly, we can convert a nonannual distribution to an annual distribution by following the same routine. For example, suppose that the weekly continuously compounded rate of return on a stock is normally distributed with $\mu^* = .003$ and $\sigma = .07$. Then the ACC return is distributed with

$$\mu^* = 52 \times .003 = .156 \text{ (or 15.6\% per year)}$$
$$\sigma = \sqrt{52} \times .07 = .5048 \text{ (or 50.48\% per year)}$$

In practice, to obtain normally distributed, continuously compounded returns, R, we take the log of 1.0 plus the raw returns:

$$R = \log(1 + r)$$

For short intervals, raw returns are small, and the continuously compounded returns, R, will be practically identical to the raw returns, r. The rule of thumb is that this conversion is not necessary for periods of 1 month or less. That is, approximating stock returns as normal will be accurate enough. For longer intervals, however, the transformation may be necessary.

A.2 DESCRIPTIVE STATISTICS

Our analysis so far has been forward looking, or, as economists like to say, ex ante. We have been concerned with probabilities, expected values, and surprises. We made our analysis more tractable by assuming that decision outcomes are distributed according to relatively simple formulas and that we know the parameters of these distributions.

Investment managers must satisfy themselves that these assumptions are reasonable, which they do by constantly analyzing observations from relevant random variables that accumulate over time. Distribution of past rates of return on a stock is one element they need to know in order to make optimal decisions. True, the distribution of the rate of return itself changes over time. However, a sample that is not too old does yield information relevant to the next-period probability distribution and its parameters. In this section we explain descriptive statistics, or the organization and analysis of such historic samples.

Histograms, Boxplots, and Time Series Plots

Table A.3 shows the annual excess returns (over the T-bill rate) for two major classes of assets, the S&P 500 index and a portfolio of long-term government bonds, for the period 1926 to 1993.

One way to understand the data is to present it graphically, commonly in a *histogram* or frequency distribution. Histograms of the 68 observations in Table A.3 are shown in Figure A.4. We construct a histogram according to the following principles:

- The range (of values) of the random variable is divided into a relatively small number of equal-sized intervals. The number of intervals that makes sense depends on the number of available observations. The data in Table A.3 provide 68 observations, and thus deciles (10 intervals) seem adequate.
- A rectangle is drawn over each interval. The height of the rectangle represents the frequency of observations for each interval.
- If the observations are concentrated in one part of the range, the range may be divided to unequal intervals. In that case the rectangles are scaled so that their *area* represents the frequency of the observations for each interval. (This is not the case in our samples, however.)
- If the sample is representative, the shape of the histogram will reveal the probability distribution of the random variable. Our total of 68 observations is not a large sample, but a look at the histogram does suggest that the returns may be reasonably approximated by a normal or lognormal distribution.

Another way to represent sample information graphically is by *boxplots*. Figure A.5 is an example that uses the same data as in Table A.3. Boxplots are most useful to show the dispersion of the sample distribution. A commonly used measure of dispersion is the *interquartile range*. Recall that the range, a crude measure of dispersion, is defined as the distance between the largest and smallest observations. By its nature, this measure is unreliable because it will be determined by the two most extreme outliers of the sample.

The interquartile range, a more satisfactory variant of the simple range, is defined as the difference between the lower and upper quartiles. Below the *lower* quartile lies 25% of the sample; similarly, above the *upper* quartile lies 25% of the sample. The interquartile range therefore is confined to the central 50% of the sample. The greater the dispersion of a sample, the greater the distance between these two values.

In the boxplot the horizontal broken line represents the median, the box the interquartile range, and the vertical lines extending from the box the range. The vertical lines representing the range often are restricted (if necessary) to extend only to 1.5 times the interquartile

Table A.3 Excess Return (Risk Premiums) on Stocks and Long-Term Treasury Bonds (Maturity Premiums)

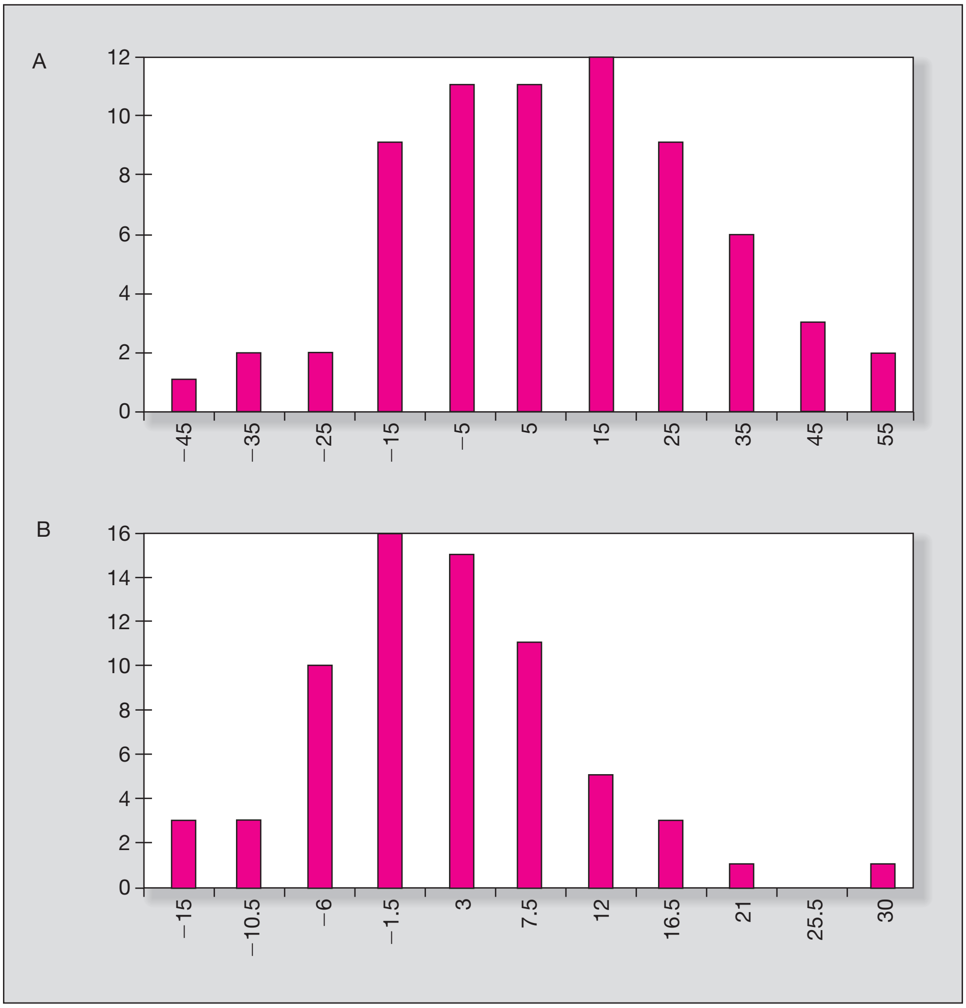
Year	Equity Risk Premium	Maturity Premium	Year	Equity Risk Premium	Maturity Premium
1926	8.35	4.50	1963	19.68	−1.91
1927	34.37	5.81	1964	12.94	−0.03
1928	40.37	−3.14	1965	8.52	−3.22
1929	−13.17	−1.33	1966	−14.82	−1.11
1930	−27.31	2.25	1967	19.77	−13.40
1931	−44.41	−6.38	1968	5.85	−5.47
1932	−9.15	15.88	1969	−15.08	−11.66
1933	53.69	−0.38	1970	−2.52	5.57
1934	−1.60	9.86	1971	9.92	8.84
1935	47.50	4.81	1972	15.14	1.84
1936	33.74	7.33	1973	−21.59	−8.04
1937	−35.34	−0.08	1974	−34.47	−3.65
1938	31.14	5.55	1975	31.40	3.39
1939	−0.43	5.92	1976	18.76	11.67
1940	−9.78	6.09	1977	−12.30	−5.79
1941	−11.65	0.87	1978	−0.62	−8.34
1942	20.07	2.95	1979	8.06	−11.60
1943	25.55	1.73	1980	21.18	−15.19
1944	19.42	2.48	1981	−19.62	−12.86
1945	36.11	10.40	1982	10.87	29.81
1946	−8.42	−0.45	1983	13.71	−8.12
1947	5.21	−3.13	1984	−3.58	5.58
1948	4.69	2.59	1985	24.44	23.25
1949	17.69	5.35	1986	12.31	18.28
1950	30.51	−1.14	1987	−0.24	−8.16
1951	22.53	−5.43	1988	10.46	3.32
1952	16.71	−0.50	1989	23.12	9.74
1953	−2.81	1.81	1990	−10.98	−1.63
1954	51.76	6.33	1991	24.95	13.70
1955	29.99	−2.87	1992	4.16	4.54
1956	4.10	−8.05	1993	7.09	15.34
1957	−13.92	4.31			
1958	41.82	−7.64	Average	8.57	1.62
1959	9.01	−5.21	Standard deviation	20.90	8.50
1960	−3.13	11.12	Minimum	−44.41	−15.19
1961	24.76	−1.16	Maximum	53.69	29.81
1962	−11.46	4.16			

Source: Data from the Center for Research of Security Prices, University of Chicago.

range, so that the more extreme observations can be shown separately (by points) as outliers.

As a concept check, verify from Table A.3 that the points on the boxplot of Figure A.5 correspond to the following list on page 1025:

Figure A.4
A. Histogram of the equity risk premium; B. histogram of the bond maturity premium.



Source: *The Wall Street Journal,* October 15, 1997. Reprinted by permission of *The Wall Street Journal,*

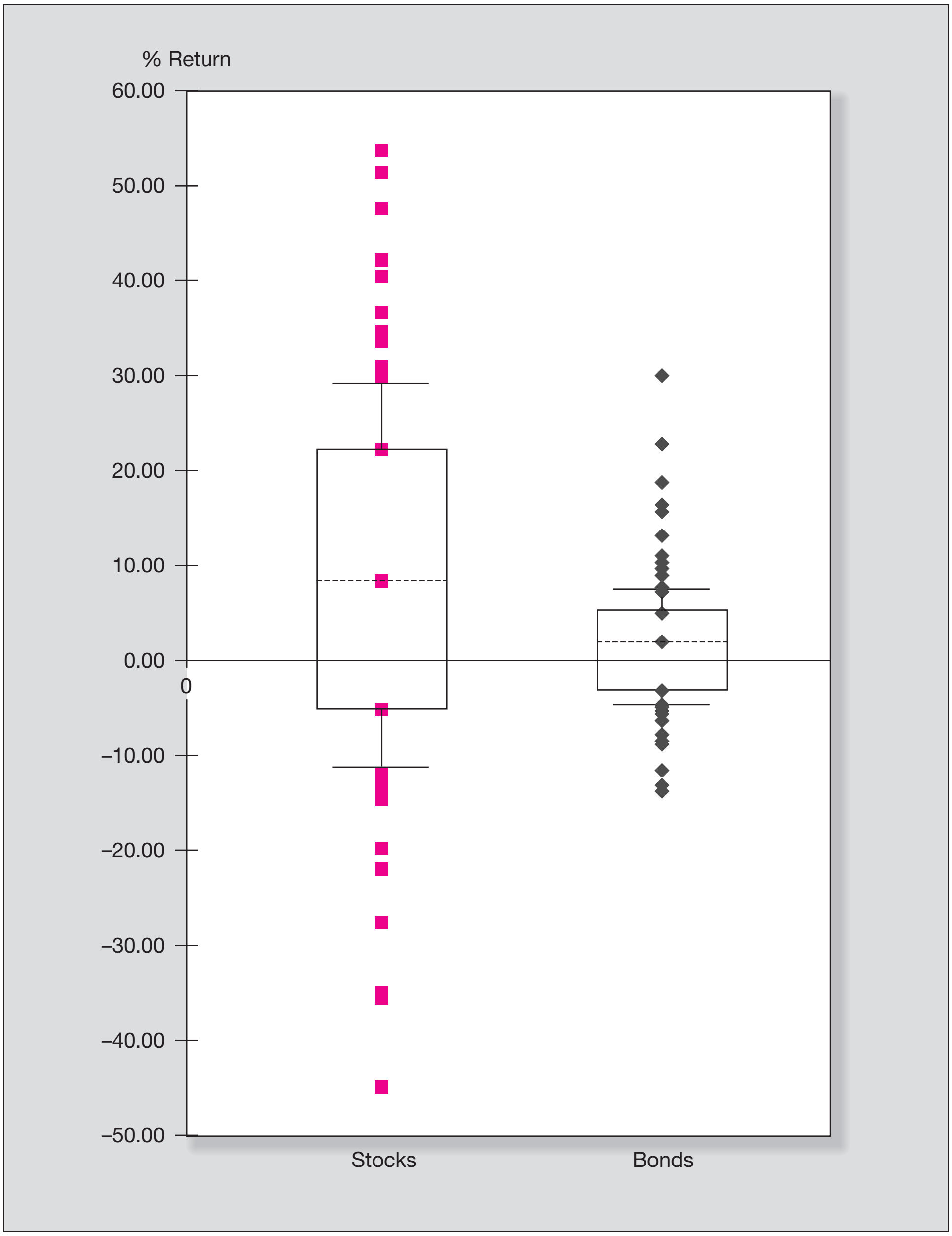
	Equity Risk Premium	Bond Maturity Premium
Lowest extreme points	−44.41	−15.19
	−35.34	−13.40
	−34.47	−12.86
	−27.31	−11.66
	−21.59	−11.60
	−19.62	−8.34
	−15.08	−8.16
	−14.82	−8.12
	−13.92	−8.05
	−13.17	−8.04
	−12.30	−7.64
		−6.38
		−5.79
		−5.47
		−5.43
		−5.21
Lowest quartile	−4.79	−3.33
Median	8.77	1.77
Highest quartile	22.68	5.64
Highest extreme points	29.99	8.84
	30.51	9.74
	31.14	9.86
	31.40	10.40
	33.74	11.12
	34.37	11.67
	36.11	13.70
	40.37	15.34
	41.82	15.88
	47.50	18.28
	51.76	23.25
	53.69	29.81
Interquartile range	27.47	8.97
1.5 times the interquartile range	41.20	13.45
From:	−11.84	−4.95
To:	29.37	8.49

Finally, a third form of graphing is *time series plots,* which are used to convey the behavior of economic variables over time. Figure A.6 shows a time series plot of the excess returns on stocks and bonds from Table A.3. Even though the human eye is apt to see patterns in randomly generated time series, examining time series' evolution over a long period does yield some information. Sometimes, such examination can be as revealing as that provided by formal statistical analysis.

Sample Statistics

Suppose we can assume that the probability distribution of stock returns has not changed over the 68 years from 1926 to 1993. We wish to draw inferences about the probability distribution of stock returns from the sample of 68 observations of annual stock excess returns in Table A.3.

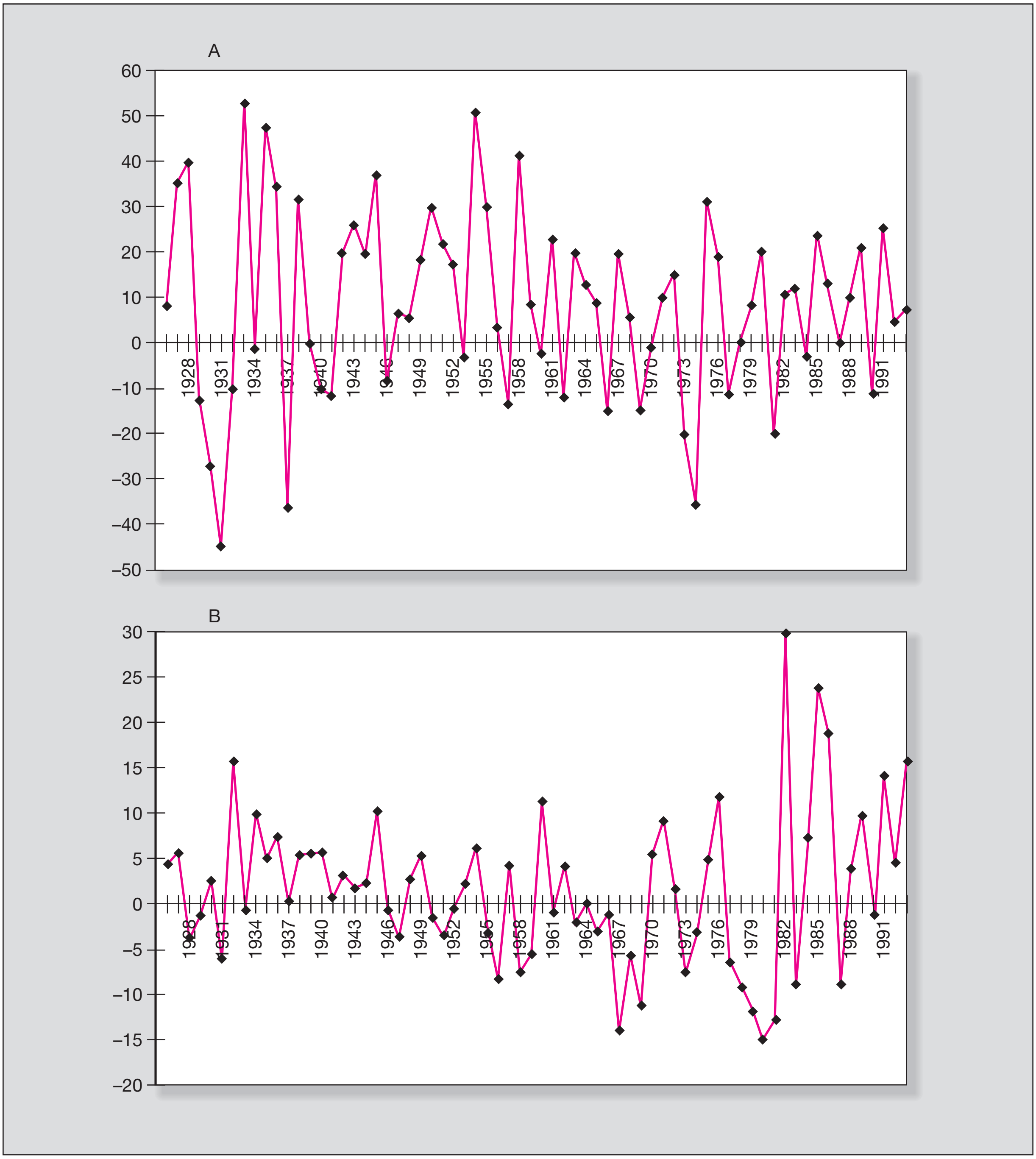
Figure A.5
Boxplots of annual equity risk premium and long-term bond (maturity) risk premium (1926–1993)



A central question is whether given observations represent independent observations from the underlying distribution. If they do, statistical analysis is quite straightforward. Our analysis assumes that this is indeed the case. Empiricism in financial markets tends to confirm this assumption in most cases.

Estimating Expected Returns from the Sample Average. The definition of expected returns suggests that the sample average be used as an estimate of the expected value. Indeed, one definition of the expected return is the average of a sample when the number of observations tends to infinity.

Figure A.6 A. Equity risk premium, 1926–1993; B. bond maturity premium, 1926–1993.



Denoting the sample returns in Table A.3 by R_t, $t = 1, \ldots, T = 68$, the estimate of the annual expected excess rate of return is

$$\bar{R} = \frac{1}{T}\Sigma R_t = 8.57\%$$

The bar over the R is a common notation for an estimate of the expectation. Intuition suggests that the larger the sample, the greater the reliability of the sample average, and the larger the standard deviation of the measured random variable, the less reliable the average. We discuss this property more fully later.

Estimating Higher Moments. The principle of estimating expected values from sample averages applies to higher moments as well. Recall that higher moments are defined as expectations of some power of the deviation from expectation. For example, the variance (second moment) is the expectation of the squared deviation from expectation. Accordingly, the sample average of the squared deviation from the average will serve as the estimate of the variance, denoted by s^2:

$$s^2 = \frac{1}{T-1}\,\Sigma(R_t - \overline{R})^2 = \frac{1}{67}\,\Sigma(R_t - .0857)^2 = .04368 \qquad (s = 20.90\%)$$

where $\overline{R}$ is the estimate of the sample average. The average of the squared deviation is taken over $T - 1 = 67$ observations for a technical reason. If we were to divide by T, the estimate of the variance would be downward-biased by the factor $(T - 1)/T$. Here too, the estimate is more reliable the larger the sample and the smaller the true standard deviation.

A.3 MULTIVARIATE STATISTICS

Building portfolios requires combining random variables. The rate of return on a portfolio is the weighted average of the individual returns. Hence understanding and quantifying the interdependence of random variables is essential to portfolio analysis. In the first part of this section we return to scenario analysis. Later we return to making inferences from samples.

The Basic Measure of Association: Covariance

Table A.4 summarizes what we have developed so far for the scenario returns on Anheuser-Busch stock and options. We know already what happens when we add a constant to one of these return variables or multiply by a constant. But what if we combine any two of them? Suppose that we add the return on the stock to the return on the call. We create a new random variable that we denote by $r(s + c) = r(s) + r(c)$, where $r(s)$ is the return on the stock and $r(c)$ is the return on the call.

Table A.4 Probability Distribution of Anheuser-Busch Stock and Options

	Scenario 1	Scenario 2	Scenario 3
Probability	.20	.50	.30
Rates of Return (%)			
Stock	20	30	50
Call option	−100	−100	400
Put option	50	−25	−100
	***E*(*r*)**	**σ**	**σ²**
Stock	.340	0.1114	0.0124
Call option	.500	2.2913	5.2500
Put option	−.325	0.5250	0.2756

From the definition, the expected value of the combination variable is

$$E[r(s + c)] = \Sigma \Pr(i) r_i(s + c) \quad \text{(A.10)}$$

Substituting the definition of $r(s + c)$ into equation A.10 we have

$$\begin{aligned} E[r(s + c)] &= \Sigma \Pr(i)[r_i(s) + r_i(c)] = \Sigma \Pr(i) r_i(s) + \Sigma \Pr(i) r_i(c) \\ &= E[r(s)] + E[r(c)] \end{aligned} \quad \text{(A.11)}$$

In words, the expectation of the sum of two random variables is just the sum of the expectations of the component random variables. Can the same be true about the variance? The answer is "no," which is, perhaps, the most important fact in portfolio theory. The reason lies in the statistical association between the combined random variables.

As a first step, we introduce the *covariance,* the basic measure of association. Although the expressions that follow may look intimidating, they are merely squares of sums; that is, $(a + b)^2 = a^2 + b^2 + 2ab$, and $(a - b)^2 = a^2 + b^2 - 2ab$, where the *a*s and *b*s might stand for random variables, their expectations or their deviations from expectations. From the definition of the variance

$$\sigma^2_{s+c} = E[r_{s+c} - E(r_{s+c})]^2 \quad \text{(A.12)}$$

To make equations A.12 through A.20 easier to read, we will identify the variables by subscripts s and c and drop the subscript i for scenarios. Substitute the definition of $r(s + c)$ and its expectation into equation A.12:

$$\sigma^2_{s+c} = E[r_s + r_c - E(r_s) - E(r_c)]^2 \quad \text{(A.13)}$$

Changing the order of variables within the brackets in equation A.13,

$$\sigma^2_{s+c} = E[r_s - E(r_s) + r_c - E(r_c)]^2$$

Within the square brackets we have the sum of the deviations from expectations of the two variables, which we denote by d. Writing this out,

$$\sigma^2_{s+c} = E[(d_s + d_c)^2] \quad \text{(A.14)}$$

Equation A.14 is the expectation of a complete square. Taking the square, we find

$$\sigma^2_{s+c} = E(d_s^2 + d_c^2 + 2d_s d_c) \quad \text{(A.15)}$$

The term in parentheses in equation A.15 is the summation of three random variables. Because the expectation of a sum is the sum of the expectations, we can write equation A.15 as

$$\sigma^2_{s+c} = E(d_s^2) + E(d_c^2) + 2E(d_s d_c) \quad \text{(A.16)}$$

In equation A.16 the first two terms on the right-hand side are the variance of the stock (the expectation of its squared deviation from expectation) plus the variance of the call. The third term is twice the expression that is the definition of the covariance discussed in equation A.17. (Note that the expectation is multiplied by 2 because expectation of twice a variable is twice the variable's expectation.)

In other words, the variance of a sum of random variables is the sum of the variances *plus* twice the covariance, which we denote by $\text{Cov}(r_s, r_c)$, or the covariance between the return on s and the return on c. Specifically,

$$\text{Cov}(r_s, r_c) = E(d_s d_c) = E\{[r_s - E(r_s)][r_c - E(r_c)]\} \quad \text{(A.17)}$$

The sequence of the variables in the expression for the covariance is of no consequence. Because the order of multiplication makes no difference, the definition of the covariance in equation A.17 shows that it will not affect the covariance either.

We use the data in Table A.4 to set up the input table for the calculation of the covariance, as shown in Table A.5.

First, we analyze the covariance between the stock and the call. In Scenarios 1 and 2, both assets show *negative* deviations from expectation. This is an indication of *positive co-movement.* When these two negative deviations are multiplied, the product, which eventually contributes to the covariance between the returns, is positive. Multiplying deviations leads to positive covariance when the variables move in the same direction and negative covariance when they move in opposite directions. In Scenario 3 both assets show *positive* deviations, reinforcing the inference that the co-movement is positive. The magnitude of the products of the deviations, weighted by the probability of each scenario, when added up, results in a covariance that shows not only the direction of the co-movement (by its sign) but also the degree of the co-movement.

The covariance is a variance-like statistic. Whereas the variance shows the degree of the movement of a random variable about its expectation, the covariance shows the degree of the co-movement of two variables about their expectations. It is important for portfolio analysis that the covariance of a variable with itself is equal to its variance. You can see this by substituting the appropriate deviations in equation A.17; the result is the expectation of the variable's squared deviation from expectation.

The first three values in the last column of Table A.5 are the familiar variances of the three assets, the stock, the call, and the put. The last three are the covariances; two of them are negative. Examine the covariance between the stock and the put, for example. In the first two scenarios the stock realizes negative deviations, while the put realizes positive deviations. When we multiply such deviations, the sign becomes negative. The same happens in the third scenario, except that the stock realizes a positive deviation and the put a negative one. Again, the product is negative, adding to the inference of negative co-movement.

With other assets and scenarios the product of the deviations can be negative in some scenarios, positive in others. The *magnitude* of the products, when *weighted* by the probabilities, determines which co-movements dominate. However, whenever the sign of the products varies from scenario to scenario, the results will offset one another, contributing to a small, close-to-zero covariance. In such cases we may conclude that the returns have either a small, or no, average co-movement.

Table A.5 Deviations, Squared Deviations, and Weighted Products of Deviations from Expectations of Anheuser-Busch Stock and Options

	Scenario 1	Scenario 2	Scenario 3	Probability-Weighted Sum
Probability	0.20	0.50	0.30	
Deviation of stock	−0.14	−0.04	0.16	
Squared deviation	0.0196	0.0016	0.0256	0.0124
Deviation of call	−1.50	−1.50	3.50	
Squared deviation	2.25	2.25	12.25	5.25
Deviation of put	0.825	0.75	−0.675	
Squared deviation	0.680625	0.005625	0.455635	0.275628
Product of deviations ($d_s d_c$)	0.21	0.06	0.56	0.24
Product of deviations ($d_s d_p$)	−0.1155	−0.003	−0.108	−0.057
Product of deviations ($d_c d_p$)	−1.2375	−0.1125	−2.3625	−1.0125

Covariance between Transformed Variables. Because the covariance is the expectation of the product of deviations from expectation of two variables, analyzing the effect of transformations on deviations from expectation will show the effect of the transformation on the covariance.

Suppose that we add a constant to one of the variables. We know already that the expectation of the variable increases by that constant, so deviations from expectation will remain unchanged. Just as adding a constant to a random variable does not affect its variance, it also will not affect its covariance with other variables.

Multiplying a random variable by a constant also multiplies its expectation, as well as its deviation from expectation. Therefore, the covariance with any other variable will also be multiplied by that constant. Using the definition of the covariance, check that this summation of the foregoing discussion is true:

$$\text{Cov}(a_1 + b_1 r_s, a_2 + b_2 r_c) = b_1 b_2 \text{Cov}(r_s, r_c) \tag{A.18}$$

The covariance allows us to calculate the variance of sums of random variables, and eventually the variance of portfolio returns.

A Pure Measure of Association: The Correlation Coefficient

If we tell you that the covariance between the rates of return of the stock and the call is .24 (see Table A.5), what have you learned? Because the sign is positive, you know that the returns generally move in the same direction. However, the number .24 adds nothing to your knowledge of the closeness of co-movement of the stock and the call.

To obtain a measure of association that conveys the degree of intensity of the co-movement, we relate the covariance to the standard deviations of the two variables. Each standard deviation is the square root of the variance. Thus the product of the standard deviations has the dimensions of the variance that is also shared by the covariance. Therefore, we can define the correlation coefficient, denoted by ρ, as

$$\rho_{sc} = \frac{\text{Cov}(r_s, r_c)}{\sigma_s \sigma_c} \tag{A.19}$$

where the subscripts on r identify the two variables involved. Because the order of the variables in the expression of the covariance is of no consequence, equation A.19 shows that the order does not affect the correlation coefficient either.

We use the covariances in Table A.5 to show the *correlation matrix* for the three variables:

	Stock	Call	Put
Stock	1.00	0.94	−0.97
Call	0.94	1.00	−0.84
Put	−0.97	−0.84	1.00

The highest (in absolute value) correlation coefficient is between the stock and the put, −.97, although the absolute value of the covariance between them is the lowest by far. The reason is attributable to the effect of the standard deviations. The following properties of the correlation coefficient are important:

- Because the correlation coefficient, just as the covariance, measures only the degree of association, it tells us nothing about causality. The direction of causality has to come from theory and be supported by specialized tests.
- The correlation coefficient is determined completely by deviations from expectations, as are the components in equation A.19. We expect, therefore, that it

is not affected by adding constants to the associated random variables. However, the correlation coefficient is invariant also to multiplying the variables by constants. You can verify this property by referring to the effect of multiplication by a constant on the covariance and standard deviation.

- The correlation coefficient can vary from −1.0, perfect negative correlation, to 1.0, perfect positive correlation. This can be seen by calculating the correlation coefficient of a variable with itself. You expect it to be 1.0. Recalling that the covariance of a variable with itself is its own variance, you can verify this using equation A.19. The more ambitious can verify that the correlation between a variable and the negative of itself is equal to −1.0. First, find from equation A.17 that the covariance between a variable and its negative equals the negative of the variance. Then check equation A.19.

Because the correlation between x and y is the same as the correlation between y and x, the *correlation matrix is symmetric about the diagonal.* The diagonal entries are all 1.0 because they represent the correlations of returns with themselves. Therefore, it is customary to present only the lower triangle of the correlation matrix.

Reexamine equation A.19. You can invert it so that the covariance is presented in terms of the correlation coefficient and the standard deviations as in equation A.20:

$$\text{Cov}(r_s r_c) = \rho_{sc}\sigma_s\sigma_c \qquad \text{(A.20)}$$

This formulation can be useful, because many think in terms of correlations rather than covariances.

Estimating Correlation Coefficients from Sample Returns. Assuming that a sample consists of independent observations, we assign equal weights to all observations and use simple averages to estimate expectations. When estimating variances and covariances, we get an average by dividing by the number of observations minus one.

Suppose that you are interested in estimating the correlation between stock and long-term default-free government bonds. Assume that the sample of 68 annual excess returns for the period 1926 to 1993 in Table A.3 is representative.

Using the definition for the correlation coefficient in equation A.19, you estimate the following statistics (using the subscripts s for stocks, b for bonds, and t for time):

$$\overline{R}_s = \frac{1}{68}\sum_{i=1}^{68} R_{s,t} = .0857; \qquad \overline{R}_b = \frac{1}{68}\sum R_{b,t} = 0.162$$

$$\sigma_s = \left[\frac{1}{67}\sum(R_{s,t} - \overline{R}_s)^2\right]^{1/2} = .2090$$

$$\sigma_b = \left[\frac{1}{67}\sum(R_{b,t} - \overline{R}_b)^2\right]^{1/2} = .0850$$

$$\text{Cov}(R_s, R_b) = \frac{1}{67}\sum[(R_{s,t} - \overline{R}_s)(R_{b,t} - \overline{R}_b)] = .00314$$

$$\rho_{sb} = \frac{\text{Cov}(R_s, R_b)}{\sigma_s\sigma_b} = .17916$$

Here is one example of how problematic estimation can be. Recall that we predicate our use of the sample on the assumption that the probability distributions have not changed over the sample period. To see the problem with this assumption, suppose that we reestimate the correlation between stocks and bonds over a more recent period—for example, beginning in 1965, about the time of onset of government debt financing of both the war in Vietnam and the Great Society programs.

Repeating the previous calculations for the period 1965 to 1987, we find

$$\overline{R}_s = .0312; \qquad \overline{R}_b = -.00317$$
$$\sigma_s = .15565; \qquad \sigma_b = .11217$$
$$\text{Cov}(R_s, R_b) = .0057; \qquad \rho_{sb} = .32647$$

A comparison of the two sets of numbers suggests that it is likely, but by no means certain, that the underlying probability distributions have changed. The variance in the rates of return and the size of the sample are why we cannot be sure. We shall return to the issue of testing the sample statistics shortly.

Regression Analysis

We will use a problem from the CFA examination (Level I, 1986) to represent the degree of understanding of regression analysis that is required for the ground level. However, first let us develop some background.

In analyzing measures of association so far, we have ignored the question of causality, identifying simply *independent* and *dependent* variables. Suppose that theory (in its most basic form) tells us that all asset excess returns are driven by the same economic force whose movements are captured by a broad-based market index, such as excess return on the S&P 500 stock index.

Suppose further that our theory predicts a simple, linear relationship between the excess return of any asset and the market index. A linear relationship, one that can be described by a straight line, takes on this form:

$$R_{j,t} = a_j + b_j R_{M,t} + e_{j,t} \tag{A.21}$$

where the subscript j represents any asset, M represents the market index (the S&P 500), and t represents variables that change over time. (In the following discussion we omit subscripts when possible.) On the left-hand side of equation A.21 is the dependent variable, the excess return on asset j. The right-hand side has two parts, the explained and unexplained (by the relationship) components of the dependent variable.

The explained component of R_j is the $a + bR_M$ part. It is plotted in Figure A.7. The quantity a, also called the intercept, gives the value of R_j when the *independent* variable is zero. This relationship assumes that it is a constant. The second term in the explained part of the return represents the driving force, R_M, times the sensitivity coefficient, b, that transmits movements in R_M to movements in R_j. The term b is also assumed to be constant. Figure A.7 shows that b is the slope of the regression line.

The unexplained component of R_j is represented by the *disturbance* term, e_j. The disturbance is assumed to be uncorrelated with the explanatory variable, R_M, and of zero expectation. Such a variable is also called a noise variable because it contributes to the variance but not to the expectation of the dependent variable, R_j.

A relationship such as that shown in equation A.21 applied to data, with coefficients estimated, is called a *regression equation.* A relationship including only one explanatory variable is called *simple regression.* The parameters a and b are called (simple) *regression coefficients.* Because every value of R_j is explained by the regression, the expectation and variance of R_j are also determined by it. Using the expectation of the expression in equation A.21, we get

$$E(R_j) = a + bE(R_M) \tag{A.22}$$

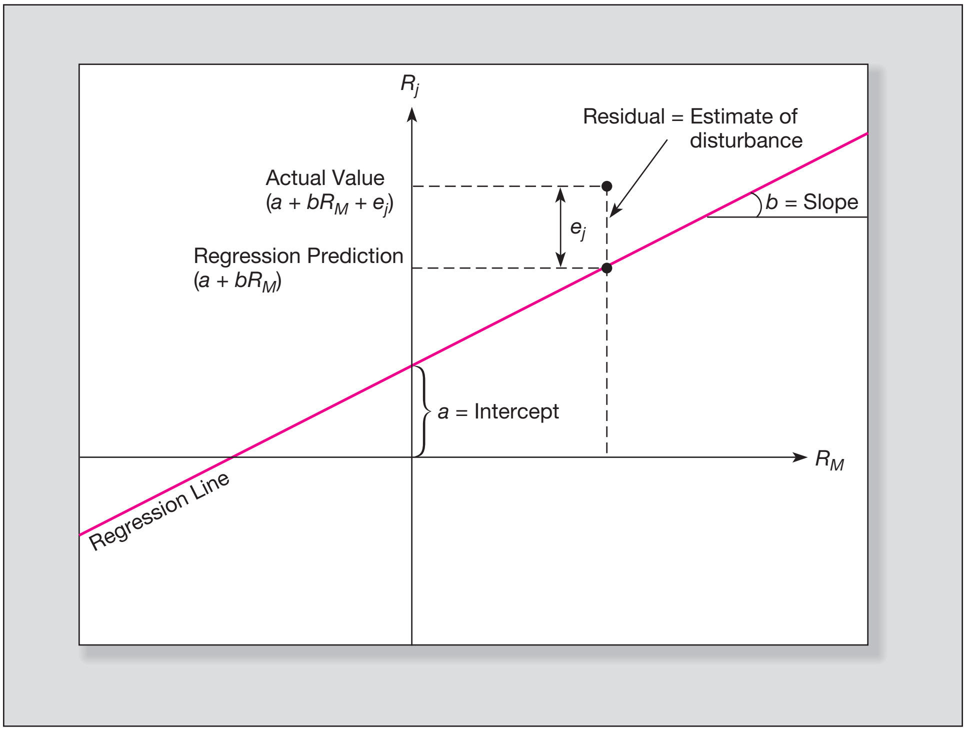


Figure A.7 **Simple regression estimates and residuals. The intercept and slope are chosen so as to minimize the sum of the squared deviations from the regression line.**

The constant a has no effect on the variance of R_j. Because the variables r_M and e_j are uncorrelated, the variance of the sum, $bR_M + e$, is the sum of the variances. Accounting for the parameter b multiplying R_M, the variance of R_j will be

$$\sigma_j^2 = b^2 \sigma_M^2 + \sigma_e^2 \tag{A.23}$$

Equation A.23 tells us that the contribution of the variance of R_M to that of R_j depends on the regression (slope) coefficient b. The term $(b\sigma_M)^2$ is called the *explained variance*. The variance of the disturbance makes up the *unexplained variance*.

The covariance between R_j and R_M is also given by the regression equation. Setting up the expression, we have

$$\begin{aligned} \text{Cov}(R_j, R_M) &= \text{Cov}(a + bR_M + e, R_M) \\ &= \text{Cov}(bR_M, R_M) = b\text{Cov}(R_M, R_M) = b\sigma_M^2 \end{aligned} \tag{A.24}$$

The intercept, a, is dropped because a constant added to a random variable does not affect the covariance with any other variable. The disturbance term, e, is dropped because it is, by assumption, uncorrelated with the market return.

Equation A.24 shows that the slope coefficient of the regression, b, is equal to

$$b = \frac{\text{Cov}(R_j, R_M)}{\sigma_M^2}$$

The slope thereby measures the co-movements of j and M as a fraction of the movement of the driving force, the explanatory variable M.

One way to measure the explanatory power of the regression is by the fraction of the variance of R_j that it explains. This fraction is called the *coefficient of determination* and denoted by ρ^2.

$$\rho_{jM}^2 = \frac{b^2\sigma_M^2}{\sigma_j^2} = \frac{b^2\sigma_M^2}{b_M^2\sigma_M^2 + \sigma_e^2} \tag{A.25}$$

Note that the unexplained variance, σ_e^2, has to make up the difference between the coefficient of determination and 1.0. Therefore, another way to represent the coefficient of determination is by

$$\rho_{jM}^2 = 1 - \frac{\sigma_e^2}{\sigma_j^2}$$

Some algebra shows that the coefficient of determination is the square of the correlation coefficient. Finally, squaring the correlation coefficient tells us what proportion of the variance of the dependent variable is explained by the independent (the explanatory) variable.

Estimation of the regression coefficients *a* and *b* is based on minimizing the sum of the square deviation of the observations from the estimated regression line (see Figure A.7). Your calculator, as well as any spreadsheet program, can compute regression estimates.

A past CFA examination included this question:

Question.

Pension plan sponsors place a great deal of emphasis on universe rankings when evaluating money managers. In fact, it appears that sponsors assume implicitly that managers who rank in the top quartile of a representative sample of peer managers are more likely to generate superior relative performance in the future than managers who rank in the bottom quartile.

The validity of this assumption can be tested by regressing percentile rankings of managers in one period on their percentile rankings from the prior period.

1. Given that the implicit assumption of plan sponsors is true to the extent that there is perfect correlation in percentile rankings from one period to the next, list the numerical values you would expect to observe for the slope of the regression, and the *R*-squared of the regression.
2. Given that there is no correlation in percentile rankings from period to period, list the numerical values you would expect to observe for the intercept of the regression, the slope of the regression, and the *R*-squared of the regression.
3. Upon performing such a regression, you observe an intercept of .51, a slope of −.05, and an *R*-squared of .01. Based on this regression, state your best estimate of a manager's percentile ranking next period if his percentile ranking this period were .15.
4. Some pension plan sponsors have agreed that a good practice is to terminate managers who are in the top quartile and to hire those who are in the bottom quartile. State what those who advocate such a practice expect implicitly about the correlation and slope from a regression of the managers' subsequent ranking on their current ranking.

Answer.

1. Intercept = 0
   Slope = 1
   R-squared = 1
2. Intercept = .50
   Slope = 0.0
   R-squared = 0.0
3. 50th percentile, derived as follows:

$$\begin{aligned} y &= a + bx \\ &= .51 - 0.05(.15) \\ &= .51 - .0075 \\ &= .5025 \end{aligned}$$

Given the very low *R*-squared, it would be difficult to estimate what the manager's rank would be.

4. Sponsors who advocate firing top-performance managers and hiring the poorest implicitly expect that both the correlation and slope would be significantly negative.

Multiple Regression Analysis

In many cases, theory suggests that a number of independent, explanatory variables drive a dependent variable. This concept becomes clear enough when demonstrated by a two-variable case. A real estate analyst offers the following regression equation to explain the return on a nationally diversified real estate portfolio:

$$RE_t = a + b_1 RE_{t-1} + b_2 NVR_t + e_t \quad (A.26)$$

The dependent variable is the period *t* real estate portfolio, RE_t. The model specifies that the explained part of that return is driven by two independent variables. The first is the previous period return, RE_{t-1}, representing persistence of momentum. The second explanatory variable is the current national vacancy rate, NVR_t.

As in the simple regression, *a* is the intercept, representing the value that RE is expected to take when the explanatory variables are zero. The (slope) regression coefficients, b_1 and b_2, represent the *marginal* effect of the explanatory variables.

The coefficient of determination is defined exactly as before. The ratio of the variance of the disturbance, *e*, to the total variance of RE is 1.0 *minus* the coefficient of determination. The regression coefficients are estimated here, too, by finding coefficients that minimize the sum of squared deviations of the observations from the prediction of the regression.

A.4 HYPOTHESIS TESTING

The central hypothesis of investment theory is that nondiversifiable (systematic) risk is rewarded by a higher *expected* return. But do the data support the theory? Consider the data on the excess return on stocks in Table A.3. The estimate of the expected excess return (the sample average) is 8.57%. This appears to be a hefty risk premium, but so is the risk—the estimate of the standard deviation for the same sample is 20.9%. Could it be that the positive average is just the luck of the draw? Hypothesis testing supplies probabilistic answers to such concerns.

The first step in hypothesis testing is to state the claim that is to be tested. This is called the *null hypothesis* (or simply the *null*), denoted by H_0. Against the null, an alternative claim (hypothesis) is stated, which is denoted by H_1. The objective of hypothesis testing is to decide whether to reject the null in favor of the alternative while identifying the probabilities of the possible errors in the determination.

A hypothesis is *specified* if it assigns a value to a variable. A claim that the risk premium on stocks is zero is one example of a specified hypothesis. Often, however, a hypothesis is general. A claim that the risk premium on stocks is not zero would be a completely general alternative against the specified hypothesis that the risk premium is zero. It amounts to "anything but the null." The alternative that the risk premium is *positive,* although not completely general, is still unspecified. Although it is sometimes desirable to test two unspecified hypotheses (e.g., the claim that the risk premium is zero or negative, against the claim that it is positive), unspecified hypotheses complicate the task of determining the probabilities of errors in judgment.

What are the possible errors? There are two, called Type I and Type II errors. Type I is the event that we will *reject* the null when it is *true.* The probability of Type I error is called the *significance level.* Type II is the event that we will *accept* the null when it is *false.*

Suppose we set a criterion for acceptance of H_0 that is so lax that we know for certain we will accept the null. In doing so we will drive the significance level to zero (which is good). If we will never reject the null, we will also never reject it when it is true. At the same time the probability of Type II error will become 1 (which is bad). If we will accept the null for certain, we must also do so when it is false.

The reverse is to set a criterion for acceptance of the null that is so stringent that we know for certain that we will reject it. This drives the probability of Type II error to zero (which is good). By never accepting the null, we avoid accepting it when it is false. Now, however, the significance level will go to 1 (which is bad). If we always reject the null, we will reject it even when it is true.

To compromise between the two evils, hypothesis testing fixes the significance level; that is, it limits the probability of Type I error. Then, subject to this present constraint, the ideal test will minimize the probability of Type II error. If we *avoid* Type II error (accepting the null when it is false) we actually *reject* the null when it is indeed *false.* The probability of doing so is one minus the probability of Type II error, which is called the *power of the test.* Minimizing the probability of Type II error maximizes the power of the test.

Testing the claim that stocks earn a risk premium, we set the hypotheses as

$H_0{:}E(R) = 0$ The expected excess return is zero

$H_1{:}E(R) > 0$ The expected excess return is positive

H_1 is an *unspecified alternative.* When a null is tested against a completely general alternative, it is called a *two-tailed test* because you may reject the null in favor of both greater or smaller values.

When both hypotheses are unspecified, the test is difficult because the calculation of the probabilities of Type I and II errors is complicated. Usually, at least one hypothesis is simple (specified) and set as the null. In that case it is relatively easy to calculate the significance level of the test. Calculating the power of the test that assumes the *unspecified* alternative is true remains complicated; often it is left unsolved.

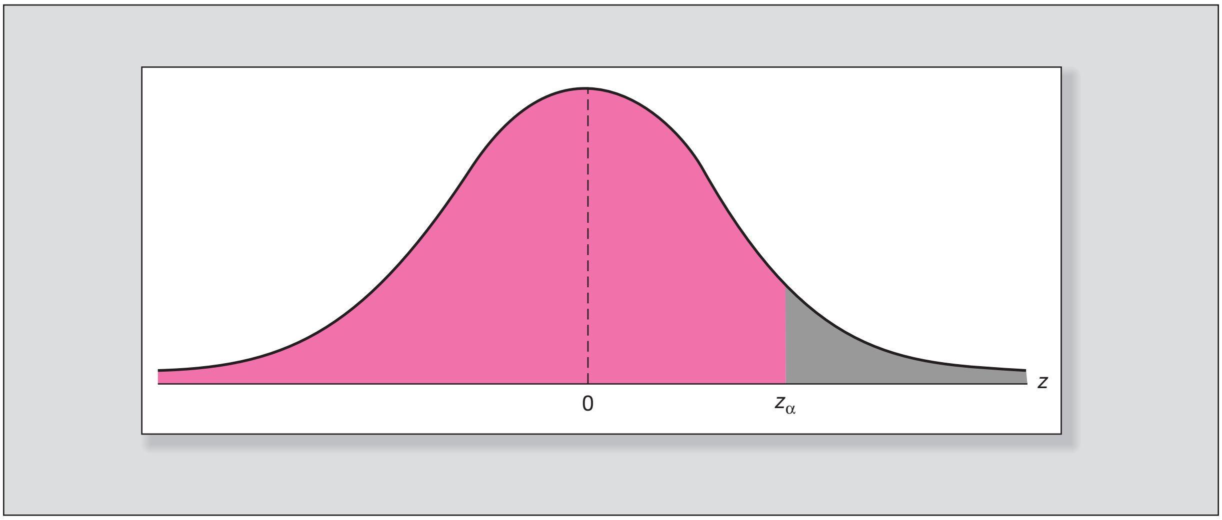
As we will show, setting the hypothesis that we wish to reject, $E(R) = 0$ as the null (the "straw man"), makes it harder to accept the alternative that we favor, our theoretical bias, which is appropriate.

In testing $E(R) = 0$, suppose we fix the significance level at 5%. This means that we will reject the null (and accept that there is a positive premium) *only* when the data suggest that the probability the null is true is 5% or less. To do so, we must find a critical value, denoted z_α (or critical values in the case of two-tailed tests), that corresponds to $\alpha = .05$, which will create two regions, an acceptance region and a rejection region. Look at Figure A.8 as an illustration.

If the sample average is to the right of the critical value (in the rejection region), the null is rejected; otherwise, it is accepted. In the latter case it is too likely (i.e., the probability is greater than 5%) that the sample average is positive simply because of sampling error. If the sample average is greater than the critical value, we will reject the null in favor of the alternative. The probability that the positive value of the sample average results from sampling error is 5% or less.

If the alternative is one-sided (one-tailed), as in our case, the acceptance region covers the entire area from minus infinity to a positive value, above which lies 5% of the distribution (see Figure A.8). The critical value is z_α in Figure A.8. When the alternative is two-tailed, the area of 5% lies at both extremes of the distribution and is equally divided

Figure A.8 Under the null hypothesis, the sample average excess return should be distributed around zero. If the actual average exceeds z_α, we conclude that the null hypothesis is false.



between them, 2.5% on each side. A two-tailed test is more stringent (it is harder to reject the null). In a one-tailed test the fact that our theory predicts the direction in which the average will deviate from the value under the null is weighted in favor of the alternative. The upshot is that for a significance level of 5%, with a one-tailed test, we use a confidence interval of $\alpha = .05$, instead of $\alpha/2 = .025$ as with a two-tailed test.

Hypothesis testing requires assessment of the probabilities of the test statistics, such as the sample average and variance. Therefore, it calls for some assumption about the probability distribution of the underlying variable. Such an assumption becomes an integral part of the null hypothesis, often an implicit one.

In this case we assume that the stock portfolio excess return is normally distributed. The distribution of the test statistic is derived from its mathematical definition and the assumption of the underlying distribution for the random variable. In our case the test statistic is the sample average.

The sample average is obtained by summing all observations ($T = 68$) and then multiplying by $1/T = 1/68$. Each observation is a random variable, drawn independently from the same underlying distribution, with an unknown expectation μ, and standard deviation σ. The expectation of the sum of all observations is the sum of the T expectations (all equal to μ) divided by T, therefore equal to the population expectation. The result is 8.57%, which is equal to the true expectation *plus* sampling errors. Under the null hypothesis, the expectation is zero, and the entire 8.57% constitutes sampling error.

To calculate the variance of the sample average, recall that we assumed that all observations were independent, or uncorrelated. Hence the variance of the sum is the sum of the variances, that is, T times the population variance. However, we also transform the sum, multiplying it by $1/T$; therefore, we have to divide the variance of the sum $T\sigma^2$ by T^2. We end up with the variance of the sample average as the population variance divided by T. The standard deviation of the sample average, which is called the *standard error,* is

$$\sigma(\text{average}) = \left(\frac{1}{T^2}\,\Sigma\sigma^2\right)^{1/2} = \left(\frac{1}{T^2}\,T\sigma^2\right)^{1/2} = \frac{\sigma}{\sqrt{T}} = \frac{.2090}{\sqrt{68}} = .0253 \qquad \text{(A.27)}$$

Our test statistic has a standard error of 2.53%. It makes sense that the larger the number of observations, the *smaller* the *standard error* of the estimate of the expectation.

However, note that it is the variance that goes down by the proportion $T = 68$. The standard error goes down by a much smaller proportion, $\sqrt{T} = 8.25$.

Now that we have the sample mean, 8.57%, its standard deviation, 2.53%, and know that the distribution under the null is normal, we are ready to perform the test. We want to determine whether 8.57% is significantly positive. We achieve this by standardizing our statistic, which means that we subtract from it its expected value under the null hypothesis and divide by its standard deviation. This standardized statistic can now be compared to z values from the standard normal tables. We ask whether

$$\frac{\overline{R} - E(R)}{\sigma} > z_\alpha$$

We would be finished except for another caveat. The assumption of normality is all right in that the test statistic is a weighted sum of normals (according to our assumption about returns). Therefore, it is also normally distributed. However, the analysis also requires that we *know* the variance. Here we are using a sample variance that is only an *estimate* of the true variance.

The solution to this problem turns out to be quite simple. The normal distribution is replaced with *Student-t* (or *t*, for short) *distribution.* Like the normal, the *t* distribution is symmetric. It depends on degrees of freedom, that is, the number of observations less one. Thus, here we replace z_α with $t_{\alpha, T-1}$.

The test is then

$$\frac{\overline{R} - E(R)}{\sigma} > t_{\alpha, T-1}$$

When we substitute in sample results, the left-hand side is a standardized statistic and the right-hand side is a *t*-value derived from *t* tables for $\alpha = .05$ and $T - 1 = 68 - 1 = 67$. We ask whether the inequality holds. If it does, we *reject* the null hypothesis with a 5% significance level; if it does not, we *cannot reject* the null hypothesis. (In this example, $t_{.05, 67} = 1.67$.) Proceeding, we find that

$$\frac{.0857 - 0}{.0253} = 3.39 > 1.67$$

In our sample the inequality holds, and we reject the null hypothesis in favor of the alternative that the risk premium is positive.

A repeat of the test of this hypothesis for the 1965 to 1987 period may make a skeptic out of you. For that period the sample average is 3.12%, the sample standard deviation is 15.57%, and there are $23 - 1 = 22$ degrees of freedom. Does that give you second thoughts?

The *t*-Test of Regression Coefficients

Suppose that we apply the simple regression model (equation A.21) to the relationship between the long-term government bond portfolio and the stock market index, using the sample in Table A.3. The estimation result (% per year) is

$$a = .9913, \qquad b = .0729, \qquad R\text{-squared} = .0321$$

We interpret these coefficients as follows. For periods when the excess return on the market index is zero, we expect the bonds to earn an excess return of 99.13 basis points. This is the role of the intercept. As for the slope, for each percentage return of the stock portfolio in any year, the bond portfolio is expected to earn, *additionally,* 7.29 basis points. With the average equity risk premium for the sample period of 8.57%, the sample average for bonds is .9913 + (.0729 × 8.57) = 1.62%. From the squared correlation coefficient you know that the variation in stocks explains 3.21% of the variation in bonds.

Can we rely on these statistics? One way to find out is to set up a hypothesis test, presented here for the regression coefficient *b.*

H_0: $b = 0$ The regression slope coefficient is zero, meaning that changes in the independent variable do not explain changes in the dependent variable

H_1: $b > 0$ The dependent variable is sensitive to changes in the independent variable (with a *positive* covariance)

Any decent regression software supplies the statistics to test this hypothesis. The regression customarily assumes that the dependent variable and the disturbance are normally distributed, with an unknown variance that is estimated from the sample. Thus the regression coefficient *b* is normally distributed. Because once again the null is that $b = 0$, all we need is an estimate of the standard error of this statistic.

The estimated standard error of the regression coefficient is computed from the estimated standard deviation of the disturbance and the standard deviation of the explanatory variable. For the regression at hand, that estimate is $s(b) = .0493$. Just as in the previous exercise, the critical value of the test is

$$s(b)t_{\alpha, T-1}$$

Compare this value to the value of the estimated coefficient *b.* We will reject the null in favor of $b > 0$ if

$$b > s(b)t_{\alpha, T-1}$$

which, because the standard deviation $s(b)$ is positive, is equivalent to the following condition:

$$\frac{b}{s(b)} > t_{\alpha, T-1}$$

The *t*-test reports the ratio of the estimated coefficient to its estimated standard deviation. Armed with this *t-ratio,* the number of observations, *T*, and a table of the *Student-t* distribution, you can perform the test at the desired significance level.

The *t*-ratio for our example is .0729/.0493 = 1.4787. The *t*-table for 68 degrees of freedom shows we cannot reject the null at a significance level of 5%, for which the critical value is 1.67.

A question from a past CFA exam calls for understanding of regression analysis and hypothesis testing.

Question.

An academic suggests to you that the returns on common stocks differ based on a company's market capitalization, its historical earnings growth, the stock's current yield, and whether or not the company's employees are unionized. You are skeptical that there are any attributes other than market exposure as measured by beta that explain differences in returns across a sample of securities.

Nonetheless, you decide to test whether or not these other attributes account for the differences in returns. You select the S&P 500 stocks as your sample, and regress their returns each month for the past five years against the company's market capitalization at the beginning of each month, the company's growth in earnings throughout the previous 12 months, the prior year's dividend divided by the stock price at the beginning of each month, and a dummy variable that has a value of 1 if employees are unionized and 0 if not.

1. The average R-squared from the regression is .15, and it varies very little from month to month. Discuss the significance of this result.
2. You note that all of the coefficients of the attributes have t-statistics greater than 2 in most of the months in which the regressions were run. Discuss the significance of these attributes in terms of explaining differences in common stock returns.
3. You observe in most of the regressions that the coefficient of the dummy variable is $-.14$ and the t-statistic is -4.74. Discuss the implication of the coefficient regarding the relationship between unionization and the return on a company's common stock.

Answer.

1. Differences in the attributes' values together explain about 15% of the differences in return among the stocks in the S&P 500 index. The remaining unexplained differences in return may be attributable to omitted attributes, industry affiliations, or stock-specific factors. This information by itself is not sufficient to form any qualitative conclusions. The fact that R-squared varied little from month to month implies that the relationship is stable and the observed results are not sample specific.
2. Given a t-statistic greater than 2 in most of the months, one would regard the attribute coefficients as statistically significant. If the attribute coefficients were not significantly different from zero, one would expect t-statistics greater than 2 in fewer than 5% of the regressions for each attribute coefficient. Because the t-statistics are greater than 2 much more frequently, one should conclude that they are definitely significant in terms of explaining differences in stock returns.
3. Because the coefficient for the dummy variable representing unionization has persistently been negative and since it persistently has been statistically significant, one would conclude that disregarding all other factors, unionization lowers a company's common stock return. That is, everything else being equal, nonunionized companies will have higher returns than companies whose employees are unionized. Of course, one would want to test the model further to see if there are omitted variables or other problems that might account for this apparent relationship.