Chapter 3

Fundamentals of Statistics

The preceding chapter was mainly concerned with the theory of probability, including distribution theory. In practice, researchers have to find methods to choose among distributions and to estimate distribution parameters from real data. The subject of sampling brings us now to the theory of statistics. Whereas probability assumes that the distributions are known, statistics attempts to make inferences from actual data.

Here we sample from the distribution of a population, say the change in the exchange rate, to make inferences about the population. The questions are, what is the best distribution for this random variable, and what are the best parameters for this distribution? Risk measurement typically deals with large numbers of random variables. So we also want to characterize the relationships between the risk factors to which the portfolio is exposed. For example, do we observe that movements in the yen/dollar rate are correlated with movements in the dollar/euro rate? Another type of problem is to develop decision rules to test some hypotheses—for instance, whether the volatility is stable over time.

These examples illustrate two important problems in statistical inference, **estimation** and **tests of hypotheses**. With estimation, we wish to estimate the value of an unknown parameter from sample data. With tests of hypotheses, we wish to verify a conjecture about the data.

This chapter reviews the fundamental tools of statistics theory for risk managers. Section 3.1 discusses the sampling of real data and the construction of returns. The problem of parameter estimation is presented in Section 3.2. Section 3.3 then turns to regression analysis, summarizing important results as well as common pitfalls in their interpretation.

3.1 Real Data

To start with an example, let us say that we observe movements in the daily yen/dollar exchange rate and wish to characterize the distribution of tomorrow's exchange rate.

The risk manager's job is to assess the range of potential gains and losses on a trader's position. He or she observes a sequence of past spot prices, $S_0, S_1, \ldots, S_t$, from which the distribution of tomorrow's price, S_{t+1}, must be inferred.

3.1.1 Measuring Returns

The truly random component in tomorrow's price is not its level, but its change relative to today's price. We measure the *relative rate of change* in the spot price:

$$r_t = (S_t - S_{t-1})/S_{t-1} \tag{3.1}$$

Alternatively, we could construct the logarithm of the price ratio:

$$R_t = \ln[S_t/S_{t-1}] \tag{3.2}$$

which is equivalent to using continuous instead of discrete compounding. This is also expressed as

$$R_t = \ln[1 + (S_t - S_{t-1})/S_{t-1}] = \ln[1 + r_t]$$

Because $\ln(1 + x)$ is close to x if x is small, R_t should be close to r_t provided the return is small. For daily data, there is typically little difference between R_t and r_t.

The return defined so far is the **capital appreciation return**, which ignores the income payment on the asset. Define the dividend or coupon as D_t. In the case of an exchange rate position, this is the interest payment in the foreign currency over the holding period. The **total return** on the asset is

$$r_t^{\text{TOT}} = (S_t + D_t - S_{t-1})/S_{t-1} \tag{3.3}$$

When the horizon is very short, the income return is typically very small compared with the capital appreciation return.

The next question is whether the sequence of variables r_t can be viewed as independent observations. If so, one could hypothesize, for instance, that the random variables are drawn from a normal distribution $N(\mu, \sigma^2)$. We could then proceed to estimate μ and σ^2 from the data and use this information to create a distribution for tomorrow's spot price change.

Independent observations have the very nice property that their joint distributions is the product of their marginal distributions, which considerably simplifies the analysis. The obvious question is whether this assumption is a workable approximation. In fact, there are good economic reasons to believe that rates of change on financial prices are nearly independent.

The hypothesis of **efficient markets** postulates that current prices convey all relevant information about the asset. If so, any change in the asset price must be dne to news or events, which are by definition impossible to forecast (otherwise, it would not be news). This implies that changes in prices are unpredictable and hence satisfy our definition of independent random variables.

This hypothesis, also known as the **random walk** theory, implies that the conditional distribution of returns depends only on current prices, and not on the previous history of prices. If so, technical analysis must be a fruitless exercise, because previous patterns in prices cannot help in forecasting price movements.

If in addition the distribution of returns is constant over time, the variables are said to be **independently and identically distributed** (i.i.d.). So we could consider that the observations r_t are independent draws from the same distribution $N(\mu, \sigma^2)$.

Later we will consider deviations from this basic model. Distributions of financial returns typically display fat tails. Also, variances are not constant and display some persistence; expected returns can also slightly vary over time.

3.1.2 Time Aggregation

It is often necessary to translate parameters over a given horizon to another horizon. For example, we may have raw data for daily returns, from which we compute a daily volatility that we want to extend to a monthly volatility.

Returns can be easily related across time when we use the log of the price ratio, because the log of a product is the sum of the logs of the indivual terms. The two-day return, for example, can be decomposed as

$$R_{02} = \ln[S_2/S_0] = \ln[(S_2/S_1) \times (S_1/S_0)] = \ln[S_1/S_0] + \ln[S_2/S_1] = R_{01} + R_{12} \tag{3.4}$$

This decomposition is only approximate if we use discrete returns, however.

The expected return and variance are then $E(R_{02}) = E(R_{01}) + E(R_{12})$ and $V(R_{02}) = V(R_{01}) + V(R_{12}) + 2\text{Cov}(R_{01}, R_{12})$. Assuming returns are uncorrelated and have identical distributions across days, we have $E(R_{02}) = 2E(R_{01})$ and $V(R_{02}) = 2V(R_{01})$.

Generalizing over T days, we can relate the moments of the T-day returns R_T to those of the one-day returns R_1:

$$E(R_T) = E(R_1)T \tag{3.5}$$

$$V(R_T) = V(R_1)T \tag{3.6}$$

Expressed in terms of volatility, this yields the **square-root-of-time rule**:

$$\text{SD}(R_T) = \text{SD}(R_1)\sqrt{T} \tag{3.7}$$

Key concept:
When successive returns are uncorrelated, the volatility increases as the horizon extends following the square root of time.

It should be emphasized that this holds only if returns have constant parameters across time and are uncorrelated. When there is nonzero correlation across days, the two-day variance is

$$V(R_2) = V(R_1) + V(R_1) + 2\rho V(R_1) = 2V(R_1)(1+\rho) \tag{3.8}$$

Because we are considering correlations in the time series of the same variable, ρ is called the **autocorrelation coefficient**. A positive value for ρ implies that a movement in one direction in one day is likely to be followed by another movement in the same direction the next day. A positive autocorrelation signals the existence of a **trend.** In this case, Equation (3.8) shows that the two-day variance is greater than that obtained by the square-root-of-time rule.

A negative value for ρ implies that a movement in one direction in one day is likely to be followed by a movement in the other direction the next day. So prices tend to revert back to a mean value. A negative autocorrelation signals **mean reversion.** In this case, the two-day variance is less than that obtained by the square-root-of-time rule.

3.1.3 Portfolio Aggregation

Let us now turn to the aggregation of returns across assets. Consider, for example, an equity portfolio consisting of investments in N shares. Define the number of each share held as q_i with unit price S_i. The portfolio value at time t is then

$$W_t = \sum_{i=1}^{N} q_i S_{i,t} \tag{3.9}$$

We can write the weight assigned to asset i as

$$w_{i,t} = \frac{q_i S_{i,t}}{W_t} \tag{3.10}$$

By construction these weights sum to unity. Using weights, however, rules out situations with zero net investment, $W_t = 0$, such as some derivatives positions. But we could have positive and negative weights if short selling is allowed, or weights greater than 1 if the portfolio can be leveraged.

In the next period, the portfolio value is

$$W_{t+1} = \sum_{i=1}^{N} q_i S_{i,t+1} \tag{3.11}$$

assuming that the unit price incorporates any income payment. The gross, or dollar, return is then

$$W_{t+1} - W_t = \sum_{i=1}^{N} q_i (S_{i,t+1} - S_{i,t}) \tag{3.12}$$

and the *rate* of return is

$$\frac{W_{t+1} - W_t}{W_t} = \sum_{i=1}^{N} \frac{q_i S_{i,t}}{W_t} \frac{(S_{i,t+1} - S_{i,t})}{S_{i,t}} = \sum_{i=1}^{N} w_{i,t} \frac{(S_{i,t+1} - S_{i,t})}{S_{i,t}} \tag{3.13}$$

So the portfolio rate of return is a linear combination of the asset returns:

$$r_{p,t+1} = \sum_{i=1}^{N} w_{i,t} r_{i,t+1} \tag{3.14}$$

The dollar return is then

$$W_{t+1} - W_t = \left[\sum_{i=1}^{N} w_{i,t} r_{i,t+1} \right] W_t \tag{3.15}$$

and has a normal distribution if the individual returns are also normally distributed.

Alternatively, we could express the individual positions in dollar terms,

$$x_{i,t} = w_{i,t} W_t = q_i S_{i,t} \tag{3.16}$$

The dollar return is also, using dollar amounts,

$$W_{t+1} - W_t = \left[\sum_{i=1}^{N} x_{i,t} r_{i,t+1} \right] \tag{3.17}$$

As we saw in the previous chapter, the variance of the portfolio dollar return is

$$V[W_{t+1} - W_t] = x' \Sigma x \tag{3.18}$$

which, along with the expected return, fully characterizes its distribution. The portfolio VAR is then

$$\text{VAR} = \alpha \sqrt{x' \Sigma x} \tag{3.19}$$

where α depends on the selected density function.

Example 3-1: FRM Exam 2002—Question 46

An investor purchases 100 shares of XYZ at the beginning of the year for \$35. The stock pays a cash dividend of \$3 per share. The price of the stock at the time of the dividend is \$30. The dividend is not reinvested. The stock is sold at the end of the year for \$28. Calculate the holding period return for this investment. It is approximately

a) −20.0%
b) −12.6%
c) −11.4%
d) −10.3%

Example 3-2: FRM Exam 1999—Question 4

A fundamental assumption of the random walk hypothesis of market returns is that returns from one time period to the next are statistically independent. This assumption implies

a) Returns from one time period to the next can never be equal.
b) Returns from one time period to the next are uncorrelated.
c) Knowledge of the returns from one time period does not help in predicting returns from the next time period.
d) Both b) and c) are true.

Example 3-3: FRM Exam 2002—Question 3

Consider a stock with daily returns that follow a random walk. The annualized volatility is 34%. Estimate the weekly volatility of this stock assuming that the year has 52 weeks.

a) 6.80%
b) 5.83%
c) 4.85%
d) 4.71%

Example 3-4: FRM Exam 1998—Question 7

Assume an asset price variance increases linearly with time. Suppose the expected asset price volatility for the next two months is 15% (annualized), and for the one month that follows, the expected volatility is 35% (annualized). What is the average expected volatility over the next three months?

a) 22%
b) 24%
c) 25%
d) 35%

Example 3-5: FRM Exam 1997—Question 15

The standard VAR calculation for extension to multiple periods assumes that returns are serially uncorrelated. If prices display trends, the true VAR will be

a) The same as the standard VAR
b) Greater than the standard VAR
c) Less than the standard VAR
d) Unable to be determined

Example 3-6: FRM Exam 2002—Question 2

Assume we calculate a one-week VAR for a natural gas position by rescaling the daily VAR using the square-root rule. Let us now assume that we determine the *true* gas price process to be mean-reverting and recalculate the VAR. Which of the following statements is true?

a) The recalculated VAR will be less than the original VAR.
b) The recalculated VAR will be equal to the original VAR.
c) The recalculated VAR will be greater than the original VAR.
d) There is no necessary relation between the recalculated VAR and the original VAR.

3.2 Parameter Estimation

Armed with our i.i.d. sample of T observations, we can start estimating the parameters of interest, the sample mean, variance, and other moments.

As in the previous chapter, define x_i as the realization of a random sample. The expected return, or mean, $\mu = E(X)$ can be estimated by the sample mean,

$$m = \widehat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i \tag{3.20}$$

Intuitively, we assign the same weight of $1/T$ to all observations because they all have the same probability. The variance, $\sigma^2 = E[(X - \mu)^2]$, can be estimated by the sample variance,

$$s^2 = \widehat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T}(x_i - \widehat{\mu})^2 \tag{3.21}$$

Note that we divide by $T - 1$ instead of T. This is because we estimate the variance around an unknown parameter, the mean. So we have fewer degrees of freedom than otherwise. As a result, we need to adjust s^2 to ensure that its expectation equals the true value. In most situations, however, T is large, so that this adjustment is minor.

It is essential to note that these estimated values depend on the particular sample and, hence, have some inherent variability. The sample mean itself is distributed as

$$m = \widehat{\mu} \sim N(\mu, \sigma^2/T) \tag{3.22}$$

If the population distribution is normal, this exactly describes the distribution of the sample mean. Otherwise, the central limit theorem states that this distribution is valid only asymptotically (i.e., for large samples).

For the distribution of the sample variance $\widehat{\sigma}^2$, one can show that when X is normal, the following ratio is distributed as a chi-square with $(T-1)$ degrees of freedom:

$$\frac{(T-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(T-1) \tag{3.23}$$

If the sample size T is large enough, the chi-square distribution converges to a normal distribution:

$$\widehat{\sigma}^2 \sim N\left(\sigma^2, \sigma^4 \frac{2}{(T-1)}\right) \tag{3.24}$$

Using the same approximation, the sample standard deviation has a normal distribution with a standard error of

$$\mathrm{se}(\widehat{\sigma}) = \sigma\sqrt{\frac{1}{2T}} \tag{3.25}$$

We can use this information for **hypothesis testing**. For instance, we would like to detect a constant trend in X. Here, the **null hypothesis** is that $\mu = 0$. To answer the question, we use the distributional assumption in Equation (3.22) and compute a standard normal variable as the ratio of the estimated mean to its standard error, or

$$z = \frac{(m - 0)}{\sigma/\sqrt{T}} \tag{3.26}$$

Because this is now a standard normal variable, we would not expect to observe values far away from zero. Typically, we would set the confidence level at 95%, which translates into a two-tailed interval for z of $[-1.96, +1.96]$. Roughly, this means that if the absolute value of z is greater than 2, we would reject the hypothesis that m came from a distribution with a mean of zero. We can have some confidence that the true μ is indeed different from zero.

In fact, we do not know the true σ and use the estimated s instead. The distribution is a Student's t with T degrees of freedom:

$$t = \frac{(m - 0)}{s/\sqrt{T}} \tag{3.27}$$

for which the cutoff values can be found from tables. For large values of T, however, this distribution is close to the normal.

At this point, we need to make an important observation. Equation (3.22) shows that when the sample size increases, the standard error of $\widehat{\mu}$ shrinks at a rate proportional to $1/\sqrt{T}$. The precision of the estimate increases as the number of observations increases.

This result will prove useful in assessing the precision of estimates generated from **numerical simulations**, which are widely used in risk management. Numerical simulations create independent random variables over a fixed number of replications T. If T is too small, the final estimates will be imprecisely measured. If T is very large, the precision will be very good. The precision of the estimates increases at a rate proportional to $1/\sqrt{T}$.

Key concept:
With independent draws, the standard deviation of most statistics is inversely related to the square root of number of observations T. Thus, more observations make for more precise estimates.

Example:

We want to characterize movements in the monthly yen/dollar exchange rate from historical data, taken over 1990 to 1999. Returns are defined in terms of continuously compounded changes, as in Equation (3.2). The sample size is $T = 120$, and estimated parameters are $m = -0.28\%$ and $s = 3.55\%$ (per month).

Using Equation (3.22), the standard error of the mean is approximately $\text{se}(m) = s/\sqrt{T} = 0.32\%$. For the null hypothesis of $\mu = 0$, this gives a t-ratio of $t = m/\text{se}(m) = -0.28\%/0.32\% = -0.87$. Because this number is less than 2 in absolute value, we cannot reject the hypothesis that the mean is zero at the 95% confidence level. This is a typical result for financial series. The mean is not precisely estimated.

Next, we turn to the precision in the sample standard deviation. By Equation (3.25), its standard error is $\text{se}(s) = \sigma\sqrt{\frac{1}{(2T)}} = 0.229\%$. For the null of $\sigma = 0$, this gives a ratio of $z = s/\text{se}(s) = 3.55\%/0.229\% = 15.5$, which is very high. So the volatility is not zero.

Therefore, there is much more precision in the measurement of s than in that of m. We can construct 95% confidence intervals around the estimated values. These are

$$[m - 1.96 \times \text{se}(m), m + 1.96 \times \text{se}(m)] = [-0.92\%, +0.35\%]$$

$$[s - 1.96 \times \text{se}(s), s + 1.96 \times \text{se}(s)] = [3.10\%, 4.00\%]$$

So we can be reasonably confident that the volatility is between 3% and 4%, but we cannot even be sure that the mean is different from zero.

3.3 Regression Analysis

Regression analysis has particular importance for risk management, because it can be used to explain and forecast financial variables.

3.3.1 Bivariate Regression

In a **linear regression**, the **dependent variable** y is projected on a set of N predetermined **independent variables**, x. In the simplest bivariate case we write

$$y_t = \alpha + \beta x_t + \epsilon_t, \qquad t = 1, \ldots, T \tag{3.28}$$

where α is called the **intercept**, or constant, β is called the **slope**, and ϵ is called the **residual**, or **error term**. This could represent a time series or a cross-section.

The **ordinary least squares** (OLS) assumptions are

- *The errors are independent of x.*
- *The errors have a normal distribution with zero mean and constant variance, conditional on x.*
- *The errors are independent across observations.*

Based on these assumptions, the usual methodology is to estimate the coefficients by minimizing the sum of squared errors. Beta is estimated by

$$\widehat{\beta} = \frac{[1/(T-1)]\sum_t (x_t - \bar{x})(y_t - \bar{y})}{[1/(T-1)]\sum_t (x_t - \bar{x})^2} \tag{3.29}$$

where $\bar{x}$ and $\bar{y}$ correspond to the means of x_t and y_t. Alpha is estimated by

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x} \tag{3.30}$$

Note that the numerator in Equation (3.29) is also the sample covariance between two series x_i and x_j, which can be written as

$$\widehat{\sigma_{ij}} = \frac{1}{(T-1)}\sum_{t=1}^{T}(x_{t,i} - \widehat{\mu_i})(x_{t,j} - \widehat{\mu_j}) \tag{3.31}$$

To interpret β, we can take the covariance between y and x, which is

$$\text{Cov}(y, x) = \text{Cov}(\alpha + \beta x + \epsilon, x) = \beta \text{Cov}(x, x) = \beta V(x)$$

because ϵ is conditionally independent of x. This shows that the population β is also

$$\beta(y, x) = \frac{\text{Cov}(y, x)}{V(x)} = \frac{\rho(y, x)\sigma(y)\sigma(x)}{\sigma^2(x)} = \rho(y, x)\frac{\sigma(y)}{\sigma(x)} \tag{3.32}$$

The **regression fit** can be assessed by examining the size of the residuals, obtained by subtracting the fitted values $\widehat{y_t}$ from y_t,

$$\widehat{\epsilon_t} = y_t - \widehat{y_t} = y_t - \widehat{\alpha} - \widehat{\beta}x_t \tag{3.33}$$

and taking the estimated variance as

$$V(\hat{\epsilon}) = \frac{1}{(T-2)} \sum_{t=1}^{T} \hat{\epsilon}_t^2 \tag{3.34}$$

We divide by $T-2$ because the estimator uses two unknown quantities, $\hat{\alpha}$ and $\hat{\beta}$. Also note that, because the regression includes an intercept, the average value of $\hat{\epsilon}$ has to be exactly zero.

The quality of the fit can be assessed using a unitless measure called the **regression R-square**. This is defined as

$$R^2 = 1 - \frac{\text{SSE}}{\text{SSY}} = 1 - \frac{\sum_t \hat{\epsilon}_t^2}{\sum_t (y_t - \bar{y})^2} \tag{3.35}$$

where SSE is the sum of squared errors, and SSY is the sum of squared deviations of y around its mean. If the regression includes a constant, we always have $0 \leq R^2 \leq 1$. In this case, R-square is also the square of the usual correlation coefficient,

$$R^2 = \rho(y, x)^2 \tag{3.36}$$

R^2 measures the degree to which the size of the errors is smaller than that of the original dependent variables y. To interpret R^2, consider two extreme cases. If the fit is excellent, the errors will all be zero, and the numerator in Equation (3.35) will be zero, which gives $R^2 = 1$. On the other hand, if the fit is poor, SSE will be as large as SSY and the ratio will be 1, giving $R^2 = 0$.

Alternatively, we can interpret R-square by decomposing the variance of $y_t = \alpha + \beta x_t + \epsilon_t$. Because ϵ and x are uncorrelated, this yields

$$V(y) = \beta^2 V(x) + V(\epsilon)$$

Dividing by $V(y)$,

$$1 = \frac{\beta^2 V(x)}{V(y)} + \frac{V(\epsilon)}{V(y)} \tag{3.37}$$

Because R-square is also $R^2 = 1 - V(\epsilon)/V(y)$, it is equal to $= \beta^2 V(x)/V(y)$, which is the contribution in the variation of y due to β and x.

Finally, we can derive the distribution of the estimated coefficients, which is normal and centered around the true values. For the slope coefficient, $\hat{\beta} \sim N(\beta, V(\hat{\beta}))$, with variance given by

$$V(\hat{\beta}) = V(\hat{\epsilon}) \frac{1}{\sum_t (x_t - \bar{x})^2} \tag{3.38}$$

This can be used to test whether the slope coefficient is significantly different from zero. The associated test statistic

$$t = \widehat{\beta}/\sigma(\widehat{\beta}) \tag{3.39}$$

has a Student's t distribution. Typically, if the absolute value of the statistic is above 2, we would reject the hypothesis that there is no relationship between y and x.

3.3.2 Autoregression

A particularly useful application is a regression of a variable on a lagged value of itself, called **autoregression**:

$$y_t = \alpha + \beta_k y_{t-k} + \epsilon_t, \qquad t = 1, \ldots, T \tag{3.40}$$

If the coefficient is significant, previous movements in the variable can be used to predict future movements. Here the coefficient β_k is known as the kth-order **autocorrelation coefficient.**

Consider, for instance, a first-order autoregression, where the daily change in the yen/dollar rate is regressed on the previous day's value. A positive coefficient $\widehat{\beta_1}$ indicates a trend. A negative coefficient indicates mean reversion. As an example, assume we find that $\widehat{\beta_1} = 0.10$, with zero intercept. One day, the yen goes up by 2%. Our best forecast for the next day is another up-move of

$$E[y_{t+1}] = \beta_1 y_t = 0.1 \times 2\% = 0.2\%$$

Autocorrelation changes normal patterns in risk across horizons. When there is no autocorrelation, risk increases with the square root of time. With positive autocorrelation, shocks have a longer-lasting effect and risk increases faster than the square root of time.

3.3.3 Multivariate Regression

More generally, the regression in Equation (3.28) can be written, with N independent variables (perhaps including a constant),

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1N} \\ \vdots & & & & \\ x_{T1} & x_{T2} & x_{T3} & \ldots & x_{TN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} \tag{3.41}$$

or in matrix notation,

$$y = X\beta + \epsilon \tag{3.42}$$

The estimated coefficients can be written in matrix notation as

$$\widehat{\beta} = (X'X)^{-1}X'y \tag{3.43}$$

and their covariance matrix as

$$V(\widehat{\beta}) = \sigma^2(\epsilon)(X'X)^{-1} \tag{3.44}$$

We can extend the t-statistic to a multivariate environment. Say we want to test whether the last m coefficients are jointly zero. Define $\widehat{\beta_m}$ as these grouped coefficients and $V_m(\widehat{\beta})$ as their covariance matrix. We set up a statistic

$$F = \frac{\widehat{\beta}'_m V_m(\widehat{\beta})^{-1}\widehat{\beta}_m/m}{\text{SSE}/(T-N)} \tag{3.45}$$

which has an F-distribution with m and $T - N$ degrees of freedom. As before, we would reject the hypothesis if the value of F is too large compared with critical values from tables.

3.3.4 Example

This section gives an example of a regression of a stock return on the market. This is useful in assessing whether movements in the stock can be hedged using stock market index futures, for instance.

We consider 10 years of data for Intel and the S&P 500, using total rates of return over a month. Figure 3-1 plots the 120 combinations of returns, or (y_t, x_t). Apparently, there is a positive relationship between the two variables, as shown by the straight line that represents the regression fit $(\widehat{y_t}, x_t)$.

Table 3-1 displays the regression results. The regression shows a positive relationship between the two variables, with $\widehat{\beta} = 1.349$. This is significantly positive, with a standard error of 0.229 and t-statistic of 5.90. The t-statistic is very high, with an associated probability value (p-value) close to zero. Thus we can be fairly confident of a positive association between the two variables.

FIGURE 3-1 Intel Return vs. S&P Return

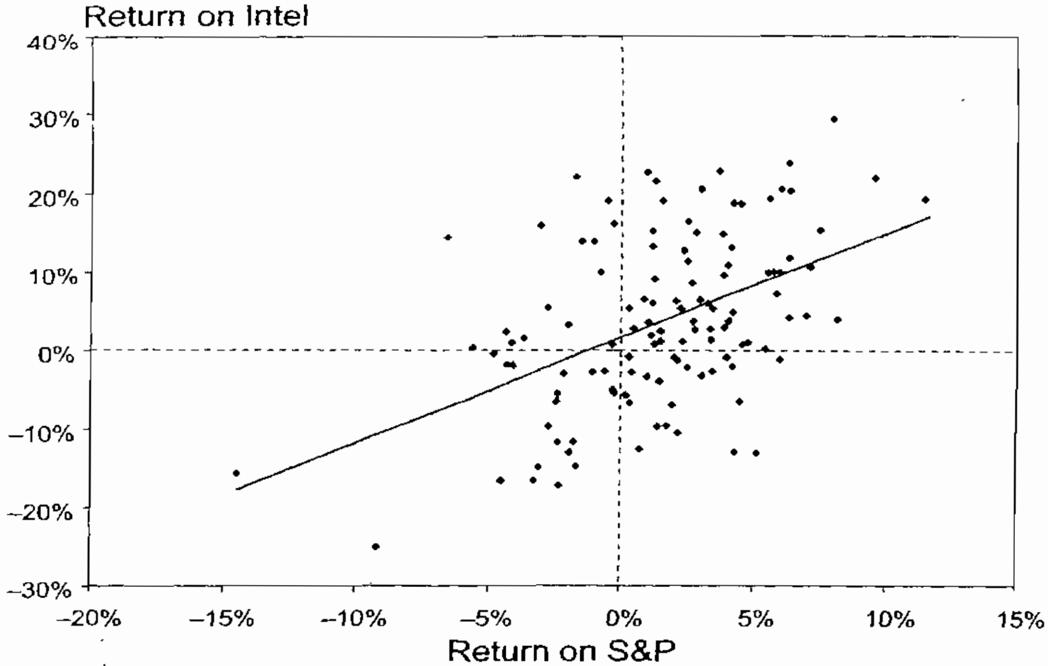

TABLE 3-1 Regression Results

$y = \alpha + \beta x$, $y =$ **Intel return,** $x =$ **S&P return**

R-square	0.228
Standard error of y	10.94%
Standard error of $\widehat{\epsilon}$	9.62%

Coefficient	Estimate	Standard Error	t-statistic	p-value
Intercept $\widehat{\alpha}$	0.0168	0.0094	1.78	0.77
Intercept $\widehat{\beta}$	1.349	0.229	5.90	0.00

This beta coefficient is also called **systematic risk**, or exposure to general market movements. Typically, technology stocks have greater-than-average systematic risk. Indeed, the slope in Intel's regression is greater than unity. To test whether β is significantly different from 1, we can compute a z-score as

$$z = \frac{(\widehat{\beta} - 1)}{s(\widehat{\beta})} = \frac{(1.349 - 1)}{0.229} = 1.53$$

This is less than the usual cutoff value of 2, so we cannot say for certain that Intel's systematic risk is greater than 1.

The R-square of 22.8% can be also interpreted by examining the reduction in dispersion from y to $\hat{\epsilon}$, which is from 10.94% to 9.62%. The R-square can be written as

$$R^2 = 1 - \frac{(9.62\%^2)}{(10.94\%^2)} = 22.8\%$$

Thus about 23% of the variance of Intel's returns can be attributed to the market.

3.3.5 Pitfalls with Regressions

As with any quantitative method, the usefulness of regression analysis depends on the underlying assumptions being fulfilled for the problem at hand. Potential problems of interpretation are now briefly mentioned.

The original OLS setup assumes that the X variables are predetermined (i.e., exogenous or fixed), as in a controlled experiment. In practice, regressions are performed on actual data that do not satisfy these strict conditions. In the previous regression, returns on the S&P are certainly not predetermined.

If the X variables are stochastic, however, most of the OLS results are still valid as long as the X variables are distributed independently of the errors and their distribution does not involve β and σ^2.

Violations of this assumption are serious because they create biases in the slope coefficients. Biases could lead the researcher to come to the wrong conclusion. For instance, we could have measurement error in the X variables, which causes the measured X to be correlated with ϵ. This so-called **errors in the variables** problem causes a downward bias, or reduces the estimated slope coefficients from their true values. Note that errors in the y variables are not an issue, because they are captured by the error component ϵ.

Another problem is **specification error**. Suppose the true model has N variables, but we use only a subset N_1. If the omitted variables are correlated with the included variables, the estimated coefficients will be biased. This is a very serious problem because it is difficult to identify.

Another class of problem is **multicollinearity**. This arises when the X variables are highly correlated. Some of the variables may be superfluous; for example, the two currencies used may be fixed to each other. As a result, the matrix in Equation (3.43) will be unstable and the estimated β unreliable. This problem will be evident in large standard errors, however. It can be fixed by discarding some of the variables that are highly correlated with others.

The third type of problem has to do with potential biases in the standard errors of the coefficients. This is are especially serious if the standard errors are underestimated, creating a sense of false precision in the regression results and perhaps leading to the wrong conclusions. The OLS approach assumes that the errors are independent across observations. This is generally the case for financial time series, but often not the case in cross-sectional setups. For instance, consider a cross-section of mutual fund returns on some attribute. Mutual fund families often have identical funds, except for the fee structure (e.g., called A for a front load, B for a deferred load). These funds, however, are invested in the same securities and have the same manager. Thus, their returns are certainly not independent. If we run a standard OLS regression with all funds, the standard errors will be too small. More generally, one has to check that there is no systematic correlation pattern in the residuals. Even with time series, problems can arise with **autocorrelation** in the errors. In addition, the residuals can have different variances across observations, in which case we have **heteroscedasticity**.[1] These problems can be identified by performing diagnostic checks on the residuals. For instance, the variance of residuals should not be related to other variables in the regression. If some relationship is found, then the model must be improved until the residuals are found to be independent.

Finally, even if all the OLS conditions are satisfied, one must be extremely careful about using a regression for forecasting. Unlike physical systems, which are inherently stable, financial markets are dynamic and relationships can change quickly. Indeed, financial anomalies, which show up as strongly significant coefficients in historical regressions, have an uncanny ability to disappear as soon as one tries to exploit them.

Example 3-7: FRM Exam 1999—Question 2

Under what circumstances could the explanatory power of regression analysis be overstated?

a) The explanatory variables are not correlated with one another.

b) The variance of the error term decreases as the value of the dependent variable increases.

c) The error term is normally distributed.

d) An important explanatory variable is omitted that influences the explanatory variables included, and the dependent variable.

[1]This is the opposite of the constant-variance case, or homoscedasticity.

Example 3-8: FRM Exam 1999—Question 20

What is the covariance between populations A and B?

A	17	14	12	13
B	22	26	31	29

a) −6.25
b) 6.50
c) −3.61
d) 3.61

Example 3-9: FRM Exam 1999—Question 6

It has been observed that daily returns on spot positions of the euro against the U.S. dollar are highly correlated with returns on spot holdings of the Japanese yen against the dollar. This implies that

a) When the euro strengthens against the dollar, the yen also tends to strengthen against the dollar. The two sets of returns are not necessarily equal.
b) The two sets of returns tend to be almost equal.
c) The two sets of returns tend to be almost equal in magnitude but opposite in sign.
d) None of the above are true.

Example 3-10: FRM Exam 1999—Question 10

An analyst wants to estimate the correlation between stocks on the Frankfurt and Tokyo exchanges. He collects closing prices for select securities on each exchange but notes that Frankfurt closes after Tokyo. How will this time discrepancy bias the computed volatilities for individual stocks and correlations between any pair of stocks, one from each market? There will be

a) Increased volatility with correlation unchanged
b) Lower volatility with lower correlation
c) Volatility unchanged with lower correlation
d) Volatility unchanged with correlation unchanged

Example 3-11: FRM Exam 2000—Question 125

If the F-test shows that the set of X variables explain a significant amount of variation in the Y variable, then

a) Another linear regression model should be tried.
b) A t-test should be used to test which of the individual X variables, if any, should be discarded.
c) A transformation of the Y variable should be made.
d) Another test could be done using an indicator variable to test the significance level of the model.

Example 3-12: FRM Exam 2000—Question 112

Positive autocorrelation in prices can be defined as

a) An upward movement in price is more than likely to be followed by another upward movement in price.

b) A downward movement in price is more than likely to be followed by another downward movement in price.

c) Both a) and b) are correct.

d) Historical prices have no correlation with future prices.

3.4 Answers to Chapter Examples

Example 3-1: FRM Exam 2002—Question 46

c) The return is given by the capital gain plus income, which is $(P_1 + D - P_0)/P_0 = (\$28 + \$3 - \$35)/\$35 = -11.4\%$. This assumes the dividend is not reinvested in the stock or in an interest-bearing account.

Example 3-2: FRM Exam 1999—Question 4

d) An efficient market implies that the distribution of future returns does not depend on past returns. Hence, returns cannot be correlated. It could happen, however, that return distributions are independent but that, just by chance, two successive returns are equal.

Example 3-3: FRM Exam 2002—Question 3

d) Assuming a random walk, we can use the square-root-of-time rule. The weekly volatility is then $34\% \times 1/\sqrt{52} = 4.71\%$.

Example 3-4: FRM Exam 1998—Question 7

b) The methodology is the same as for the time aggregation, except that the variance may not be constant over time. The total (annualized) variance is $0.15^2 \times 2 + 0.35^2 \times 1 = 0.1675$ for three months, or 0.0558 on average. Taking the square root, we get 0.236, or 24%.

Example 3-5: FRM Exam 1997—Question 15

b) This question assumes that VAR is obtained from the volatility using a normal distribution. With trends, or positive correlation between subsequent returns, the two-day variance is greater than that obtained from the square-root-of-time rule. See Equation (3.7).

Example 3-6: FRM Exam 2002—Question 2

a) With mean reversion, the volatility grows more slowly than the square root of time. This is the opposite of the case in the previous question.

Example 3-7: FRM Exam 1999—Question 2

d) If the true regression includes a third variable z that influences both y and x, the error term will not be conditionally independent of x, which violates one of the assumptions of the OLS model. This will artificially increase the explanatory power of the regression. Intuitively, the variable x will appear to explain more of the variation in y simply because it is correlated with z.

Example 3-8: FRM Exam 1999—Question 20

a) First, compute the averages of A and B, which are 14 and 27. Then construct a table as follows.

	A	B	$(A-14)$	$(B-27)$	$(A-14)(B-27)$
	17	22	3	−5	−15
	14	26	0	−1	0
	12	31	−2	4	8
	13	29	−1	2	2
Sum	56	108			−25

Summing the last column gives −25, or an average of −6.25.

Example 3-9: FRM Exam 1999—Question 6

a) Positive correlation means that, on average, a positive movement in one variable is associated with a positive movement in the other variable. Because correlation is scale-free, this has no implication for the actual size of movements.

Example 3-10: FRM Exam 1999—Question 10

c) The nonsynchronicity of prices does not alter the volatility, but will induce some error in the correlation coefficient across series. This is similar to the effect of errors in the variables, which biases downward the slope coefficient and the correlation.

Example 3-11: FRM Exam 2000—Question 125

b) The F-test applies to the group of variables but does not say which one is most significant. To identify which particular variable is significant, we use a t-test and discard the variables that do not appear significant.

Example 3-12: FRM Exam 2000—Question 112

c) Positive autocorrelation means that price movements in one direction are more likely to be followed by price movements in the same direction.