CHAPTER 1

Tools of the Trade

To understand markets in a logical and objective manner, it is important to employ tools from mathematics and statistics in an appropriate way to unearth relevant patterns. This chapter highlights some of the basic tools used by practitioners to understand the volumes of data produced on a daily basis by financial markets. The tools described here allow an understanding of data in relatively simple terms and help forecast the direction of seemingly random series. The focus here is on the intuition behind mathematical tools rather than a deep dive into the formulas. Often finance books and academic papers rely on complex mathematical models to explain market behavior. As markets are constantly evolving, elaborate quantitative constructs can rapidly become irrelevant and provide false signals about market movements.

The frequent failure of complex quantitative models to forecast markets is not intended to discourage market investors from using mathematics; instead, *misuse* of math is to be avoided. Indeed, mathematics is crucial for objective analysis of market movements and for appropriately controlling risks. Although overuse of mathematics can be a problem, the opposite is just as dangerous. Market participants who only use qualitative assessments to trade markets may be exposed to numerous hidden risks. In the paragraphs to come, we consider some statistical techniques useful in finance, the appropriate situations in which to use them, and their embedded assumptions. These techniques are building blocks; creative thinking and building on the ideas behind these methods can lead to the discovery of new trading patterns as well as a more thorough evaluation of market

P1: OTA
c01 JWBT428-Jha January 20, 2011 14:51 Printer: Courier Westford

behavior. This chapter purposefully avoids mathematical proofs; instead, the focus is on developing a practical and intuitive grasp of the concepts. Readers are encouraged to consult mathematics or statistics texts for more detailed analysis of the concepts.

BASIC STATISTICS

Knowledge of basic statistics is a prerequisite for understanding today's markets, and especially the rates markets, where there is a preponderance of mathematically adept investors. We use the term "basic stats" to refer to initial descriptive statistics used to understand the data quantitatively. We build on these initial descriptions of the data using more advanced statistical techniques. To begin with, we consider the time series of a simple and commonly traded security: the 2-year Treasury yield. Figure 1.1 shows the 2-year Treasury yield time series over the past 20 years. We refer to this series repeatedly as we build our tools to think about how to trade this security. The figure also shows the average value and increments of

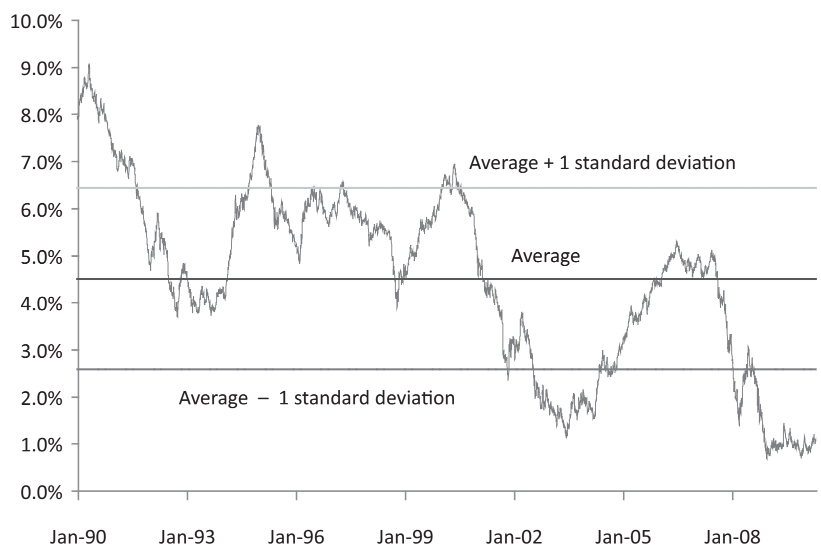

FIGURE 1.1 Two-Year Treasury Yield along with Average Value and One-Standard Deviation Bands
Source: Board of Governors of the Federal Reserve.

one standard deviation (SD) from the average value over the entire period shown. These quantities are discussed in more detail later in the chapter.

For a time series, basic descriptive statistical quantities are akin to descriptions of people's height and weight. In the realm of time series, the crucial descriptions are mean (or average), standard deviation, minimum, maximum, and, to a lesser extent, median. (Note: Mode is a commonly used descriptor in some fields but tends to have limited application in finance.) The median refers to the point in the middle when the data is sorted. The median tends to be less useful than the average since it is very resistant to outliers while profit and losses (P/Ls) are not. The mean of a string of data is the sum of the data series divided by the number of data points. The mean is susceptible to outliers since unusually high or low points can drag the average in either direction. Figure 1.1 shows the mean of the 2-year Treasury yield, which is 4.5% over the long run. This average, though, is susceptible to extreme values of the 2-year yields. At times, to deal with series where outliers can misrepresent data, outlying data can be trimmed, referred to as a trimmed mean. An example of this methodology is the calculation of the London Interbank Offered Rate (LIBOR) by the British Bankers' Association (BBA), where 16 banks submit their estimates for short-term rates and the top four and bottom four submissions are discarded to arrive at a trimmed mean. Frequently, averages are considered on a rolling basis over a fixed number of days; these are referred to as moving averages. Here, each point in the modified time series represents the average of a fixed number of preceding points from the original time series. A related concept to the average value is the weighted average. The simple average essentially multiplies each variable with an equal weight of $1/n$, where n is the number of elements. However, the weights on the average do not need to be equal, but can instead be altered based on different criteria. For example, the weight could be altered based on how close the data point is to today if more recent data points need to be given more weight.

Since most financial market entities are not deterministic, a concept called a random variable is useful. A random variable can be thought of as a variable that can take on a range of values, such as the commonly used x in algebra. However, the difference here is that instead of having a set value in an equation, a random variable essentially is a "package" of values, with each value occurring with some probability. The range of values of the random variable as well as the probabilities associated with the variable is known as its probability distribution (or just distribution). For a random variable, a special type of weighted average is known as the expected value. The expected value is calculated as the average payoff of the random variable weighted by the probability of its occurrence. For example, if the payoff from a dice roll is the rolled value, then the expected value is $1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6 = 3.5$, since each roll has 1/6 probability of taking place. Note that given the equal probability of each roll, the expected value is the same as the simple average of the payoffs. The notation for expected value of a random variable is generally $E(X)$.

Probability distributions as with the dice are known as discrete distributions because the payoffs occur in discrete amounts. In the finance field, continuous distributions, which have continuous payoffs, are used more frequently. One common probability distribution used in finance and other fields is the normal distribution, which represents the familiar bell curve. The center of the distribution, and also the most likely scenario, is the average. A continuous probability distribution's description is generally given through a mathematical function known as a probability density function. The probability density function gives the chance of an event occurring for this distribution around a single point—it can be approximately thought of as equivalent to the 1/6 probability of a given dice roll in a discrete distribution. For example, for the normal distribution, the probability density function (pdf) is

$$\mathrm{pdf}(x) = 1/\sqrt{}(2\pi\sigma^2) \times e^{-(x-\mu)\hat{}2/(2\sigma\hat{}2)}$$

where σ = standard deviation
μ = mean of the distribution

Taking this one step further, we arrive at a cumulative distribution function (cdf), which gives the probability of the random variable taking values *less than* a given value. For example, for a normal distribution, this would be denoted as $\Phi(k)$, which is the probability of a random normal variable x taking values less than k. Unlike the pdf, the normal distribution cdf does not have a closed-form formula, but instead numerical methods are needed to calculate the value. Geometrically, it is the area under the bell curve to left of the k point. The details of probability distributions are beyond the scope of this text, but the idea of a cumulative distribution arises in pricing options, which we discuss in Chapter 11. The concept behind the terminology is the important bit, as these functions can be calculated using software packages.

The expected value can be thought of as the average level of a random variable. The variance can be thought of as movement of the variable around its average. The variance is calculated as the sum of squared deviations of each data point from the average normalized by the number of data points. For the die described earlier, the variance would be calculated as $1/6 \times [(3.5-1)^2 + (3.5-2)^2 + (3.5-3)^2 + (3.5-4)^2 + (3.5-5)^2 + (3.5-6)^2]$.

In expected value notation, variance $= E(X - \mu)^2 = E(X^2) - E(X)^2$. The standard deviation, a related and more commonly considered measure, is defined as the square root of the variance. Variance and standard deviations are measures of dispersion in a data set; since they sum up squares of deviation, a negative or positive deviation is irrelevant in this case. Regardless of whether it is positive or negative, the size of the deviation is what matters. The weight of each deviation grows rapidly by how "unusual" it is (i.e., how far away it is from the average value). Figure 1.1 shows increments of one standard deviation above and below the average 2-year Treasury yield since 1990; the standard deviation here is 1.9%. Due to the link between standard deviation and dispersion, it is also used commonly as a risk measure in finance, and is commonly referred to as *volatility*. For example, volatility of a series of historical returns may be represented as 10% per year. If market returns are assumed to follow a normal distribution, the volatility has a further interpretation. For each standard deviation away from the mean in either direction, the chance of being outside the range drops off successively. For 1 standard deviation on either side of the mean of a normal distribution, the chance of an occurrence in that range is about 68.3% and 95% for 2 standard deviations. Therefore, a volatility of 10% per year implies that next year's return is likely to be 10% above or 10% below the average return about 68% of the time. Furthermore, the return has a 95% chance of being within $10\% \times 2 = 20\%$ of the average. Of course, if the distribution was different than normal, then the likelihood of outsized returns could be lower or higher. One area of continuing research in academic and market circles deals with the fact that normal distributions make "unlikely" events much rarer than reality (i.e., the frequency of multi–standard deviation moves in markets is higher than would be the case if markets were truly normal). There are other types of distributions that have "fat tails" (i.e., incorporate a higher likelihood of large moves), but the mathematics related to such distributions tends to be more complex, making them less prevalent in models.

The concept of volatility applies to a single variable and can be extended further to covariance between two or more variables. The covariance conveys how much two quantities move with respect to each other. In expected value notation, the covariance between two variables X and Y is calculated as $E(X \times Y) - E(X) \times E(Y)$. If two variables are independent, that is, they have no relation to each other, the covariance is 0. These expected values depend on the distributions of the two variables. Covariance also is used to calculate the variance of the sums of two variables. The variance of a sum of random variables A and B, expressed as $\text{Var}(A + B) = \text{Var}(A) + \text{Var}(B) + 2 \times \text{Cov}(A, B)$. The covariance is closely related to the concept of correlation, which will be discussed in more detail when we cover regression.

REGRESSION: THE FUNDAMENTALS

Now that the basic statistics have been covered, we move on to regression, which is one of the most extensively used, and at times misunderstood, tools to analyze empirical data. The purpose of regression is not really to price the instruments, but rather to analyze time series data and deduce relationships between them to predict their future behavior. The basic statistics discussed in the previous section concern a single time series. However, as might be expected, the financial world is full of intersecting relationships between variables, with various factors coming in and out of importance. Regression analysis provides a logical way to analyze such relationships.

The most commonly used type of regression is linear regression, which fits a "best-fit" line between the scatterplot of data points of two variables. By "scatterplot," we mean a graph that conveys this information: "the value of y when the value of x was __." Another way to think about this scatterplot is the value of a variable y at time t versus the value of a variable x also at time t (although time does not have to be the only common factor). An example of such a relationship would be between inflation and unemployment rate. Figure 1.2 shows the relationship from 1975 to 1977; the linear relationship between the two is evident. The y-axis on the figure is the inflation rate, the x-axis is the unemployment rate, and each point on the figure is the inflation rate that existed for a given level of unemployment rate. The points do not have to be unique—for example, at an unemployment level of 8%, we may see inflation at 2% in one time period and at 10%

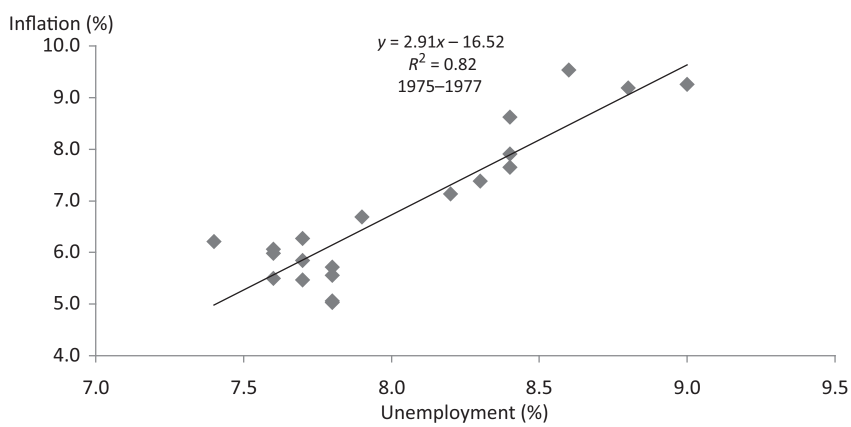

FIGURE 1.2 Inflation Regressed against Unemployment from 1975 to 1977
Source: Federal Reserve Bank of St. Louis.

in another time period, which would lead to two vertical points at 2% and 10% corresponding to 8% on the x-axis in the graph.

Once we have our data set defined, the regression line is fitted by finding a line such that the sum of squared distances from the line across the scatterplot points is minimized. This process is known as the least squares method and forms the basis for many simpler-fitting algorithms. The linear regression process results in an equation of the form:

$$y_{\text{model}} = \text{alpha} + \text{beta} \times x_t \tag{1.1}$$

where alpha and beta = constants (explained below)
x_t = explanatory variable at time t

The equation here can be thought of as the "average" relationship between y and x. Thus, by inputting an x value, we can arrive at a model y value. The meaning of the term "linear regression" becomes apparent, since our model equation is that of a straight line in geometry. The model y value is the best prediction of y *given* a value of x, which is referred to as a conditional expectation or conditional mean. Recall that earlier we referred to the mean of a series as a first-order prediction, but now we are conditioning that mean on the value of an explanatory variable, which we hope will improve prediction results. Given that Equation 1.1 is a best-fit line through the data, most data points will tend to lie at some distance from the line. This leads to the concept of the error, which can be found by considering a particular value of x. Now we consider the y value on the same day as x, referred to as the "actual" value, and subtract out the model value of y, which is found by plugging in x into Equation 1.1. To summarize:

$$\begin{aligned} y_{\text{model}} &= \text{alpha} + \text{beta} \times x_t \\ \text{error}_t &= y_t - y_{\text{model}} \end{aligned}$$

where error_t = error at time t

Before we analyze the errors in further detail, what are the alpha and beta in Equation 1.1? The alpha is referred to as the intercept, and it is the value of y if the value of x is 0. The alpha can be thought of as the "default" value of the dependent variable in case of lack of effect from x. Occasionally, the intercept can be forced to be 0 in situations when it is known beforehand that the default value of y when x is 0 is also 0, which is an option that is not commonly used. The beta here, as the reader may recall from geometry class, is the slope of the best-fit line. The slope's intuitive interpretation is the sensitivity of y to x; that is, the change in y for a unit change in x. The beta is a crucial variable to understanding relationships between variables, which in turn is an important factor when offsetting risks in trades (discussed in detail in Chapter 8). The error_t, shown as the difference between the actual value and the model value, is also known as the residual. The residual gives a quantitative measure of how effective the regression is at predicting the actual values for any given explanatory variable value. Of course, the smaller the residual, the more accurate the regression. Since the residual varies for each explanatory variable, a combined measure known as the standard error can give a quantitative measure of how well the regression works at fitting the data. The standard error is essentially the standard deviation of the residuals. It gives an indication of the average error for the regression's predictions—the larger the standard error, the less accurate the regression.

Finally, given the linear nature of this relationship, the fact that beta is constant implies that the sensitivity of y to changes in x remains the same regardless of the level of x; at times, this may be a desirable characteristic, but at other times, this may be too simplistic a model, and nonlinear methods may need to be considered (discussed later). In Figure 1.2, for example, the alpha is shown to be –16.52 and beta is 2.91. This beta implies that as unemployment rises 1%, the inflation rate in this period rises 2.91%.

The stylized example in Figure 1.2 focuses on the basic case of a single-variable linear regression. However, as the reader may have guessed, financial variables are rarely driven by a single factor, but instead are complex interactions of many factors. To consider a more realistic case, Equation 1.2 may be easily extended to the case of many variables:

$$y_{\text{model}} = \text{alpha} + \text{beta}_1 \times x_{1t} + \text{beta}_2 \times x_{2t} + \ldots + \text{beta}_n \times x_{nt} \qquad (1.2)$$

where alpha and $\text{beta}_{1\ldots n}$ = constants
$x_{1t}, x_{2t}, \ldots x_{nt} = n$ explanatory factors at time t

As with the one-variable case, the predicted y is linear with respect to its factors, but in this case, more than one explanatory variable is being used to predict y. The error here is computed using a similar calculation, by subtracting the actual value of y and the model value of y. The alpha, or intercept, also has the same interpretation, which represents the value of y if none of the factors had any contribution to the prediction. In multiple regression, the betas take on a slightly more subtle interpretation. Each beta is the *partial* sensitivity of y with regard to the corresponding factor (i.e., beta_1 is the partial sensitivity of x_1). The partial sensitivity can be thought of as the effect of one particular factor on y *after* controlling for the effects of all other variables considered. An example of this would be in the case of a medical study to consider the effect of weight on blood pressure, where it may not be enough to know the average increase in blood pressure for each one-pound gain in weight. Instead, other factors, such as family history or job stress levels, may need to be added. Once these other factors are added in the regression, the partial beta of weight will change since other relevant factors are being considered; the new, partial sensitivity of blood pressure to weight gives a more accurate relationship between the two. In the context of finance, knowledge of partial sensitivities is essential for hedging risks and accurately setting up trades. When important factors are missing from a model, a trader/researcher can have an inaccurate sense of the impact of various market factors on a trade, which can of course lead to large losses.

Multiple regressions are difficult to visualize given the multidimensional nature of the modeling. For a two-variable regression, for example, a three-dimensional figure would be required with two axes for the independent variables and an axis for the dependent variable. While impractical in a two-variable regression, for higher-order regressions, the figures are even less practical. To display multiple regressions on two-dimensional figures, the concept of a partial regression is used. Consider Equation 1.2. To display the relationship of the y variable to x_{1t}, the regression to display would be y – (beta$_2$ $\times$ x_{2t} + ... + beta$_n$ $\times$ x_{nt}) regressed against x_{1t}. This *partial* regression would preserve the same pattern and strength of association as multiple regression. To view the regression versus other x variables, the same procedure is repeated while keeping the target variable on the right-hand side of the equation.

So far we have considered only linear regression, in both single and multivariate settings. The linear aspect can be deduced by looking at the equations (they resemble the equations of a line in geometry) and also follows from the fact that the sensitivity of y to any factor x is a fixed constant beta. This constant beta does not have to be the case. In fact, linear regression is a very special case of general regression methods, and nearly any function may be used to relate x and y. The introduction of more complicated functions beyond a straight line vastly increases the complexity of the problem being considered, and should be used only if there is a compelling case to do so. As we will see in the goodness-of-fit section, an overly complicated model may improve the explanatory power of a model, but also subjects it to an increased risk of overfitting, which can actually lead to poor subsequent predictions.

Although general nonlinear regressions are a complex topic, there is a relatively simple extension of the linear regression just considered. Here, the alphas and betas are similar to the linear regression case (i.e., fixed constants), but the variables themselves may be squared, cubic, or higher powers of the factor variables. Why would we consider such a case? Generally, polynomials are considered when the relationship between y and x

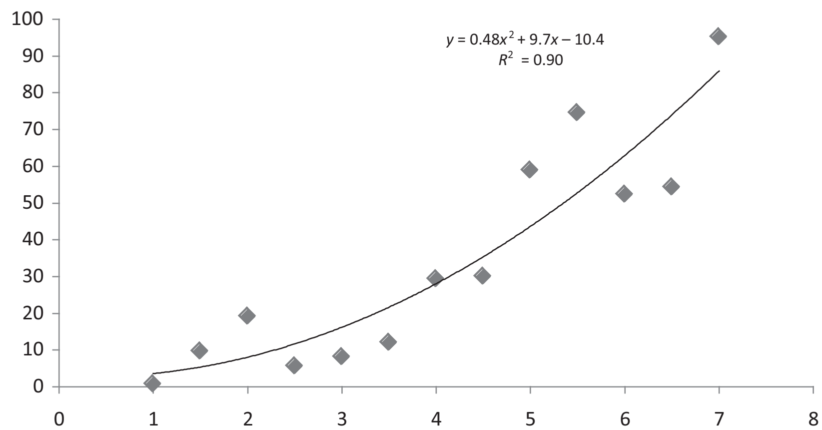

FIGURE 1.3 Example of a Quadratic Regression

is very plainly not a straight line, as seen in Figure 1.3, which shows two fictitious data sets with a quadratic relationship. It may be that the sensitivity of y grows as x grows, which would be the case in the quadratic regression depicted in Equation 1.3:

$$y_{\text{model}} = \text{alpha} + \text{beta}_1 \times x + \text{beta}_2 \times x^2 \tag{1.3}$$

Here, the equation is similar to the ones we covered earlier, but the dependence of y is on the square of x. The consequence of this is that y does not grow at a fixed beta rate as x grows; instead the growth in y becomes more rapid as x grows. Another way to think about a quadratic regression is that the *sensitivity of y to x itself grows in a linear fashion*, making y linked to x in a square relationship. For example, if we assume alpha $= 0$ and beta $= 2$, the value of y at $x = 1$ is $0 + 2 \times 1^2 = 2$; at $x = 2$, the value of y is $0 + 2 \times 2^2 = 8$. Here, the value of y rose by 6 when x increased from 1 to 2.

To explore these ideas in more concrete terms, we consider a simplified case of explaining the 2-year yield, which we considered earlier in examples of the mean and standard deviation. These quantities describe the series. In fact, if we were asked to predict the 2-year yield, we may use the long-term mean of 4.5% as our first guess. Although simple to compute, as Figure 1.1 shows, the mean is also a very poor predictor of future movements, given that deviations from the mean can be very large. How can we improve on this? To answer this question, we must first hypothesize which other, more fundamental variables may be driving the 2-year yield. As a matter of terminology, the 2-year yield then is referred to as the "dependent" variable while our mystery variable to explain the 2-year yield would be an "explanatory" or "independent" variable. By incorporating at

P1: OTA
c01 JWBT428-Jha January 20, 2011 14:51 Printer: Courier Westford

least some of the vast amount of information available to us about related variables, we may greatly improve the accuracy of our predictions. We explore drivers of interest rates in subsequent chapters and use regressions often to improve on forecasting as well as to understand which drivers matter more than others.

REGRESSION: HOW GOOD A FIT?

Although we can easily come up with a line to fit the data, not all fits are alike. We discussed the standard error earlier in relation to residuals as a way to quantitatively determine the prediction error for a regression. Although the standard error is a useful metric to understand the magnitude of errors to expect from a model, it can be difficult to compare standard errors across models, especially if the models are not very similar. Furthermore, when relying on linear regressions, the standard error measure does not easily tell us whether the linear regression model is appropriate to begin with—some data sets resemble linear patterns more closely than others. To quantify a goodness of fit, one of the most common, and also misunderstood, metrics is correlation. Correlation can be thought of as the closeness of association between two variables. It is a normalized measure allowing comparison of linear fit across different models. Correlation varies between –1 and 1. A correlation of –1 implies strong negative association, and a correlation of 1 implies strong positive association. A correlation of 0 implies little or no association. Correlation may be familiar from daily usage as a way to convey association, but in mathematical terms, the formula is:

$$\text{Correlation} = \text{cov}(X, Y)/[\text{stdev}(X) \times \text{stdev}(Y)]$$

where cov = covariance
stdev = standard deviation

Although covariance calculates association between variables X and Y, it is difficult to compare covariance across different data sets since it is dependent on their individual volatility levels. Essentially, correlation is a way of normalizing covariance for the volatilities of the individual series and results in a more easily comparable number. Correlation is not order dependent; that is, the y and x variables can be interchanged without affecting the measure.

A measure related to correlation is the R^2, which, as the name may imply, is the square of the correlation calculation above. The R^2 denotes the percentage of variation in the y variable explained by the x (or vice versa). Since it is the square of a number, it is always positive. Therefore, R^2 does

P1: OTA
c01 JWBT428-Jha January 20, 2011 14:51 Printer: Courier Westford

not specify the direction of the association, only the strength. There is no magic level of R^2; it is completely situation dependent, and for different models, different levels are tolerated. For example, for a model that is supposed to trade the market, a 50% R^2 would be poor. R^2 can also be increased by merely adding more variables, which is generally suboptimal. Therefore, an adjustment is made to the R^2 that inserts a penalty for inserting additional variables. The penalty for excess variables makes the adjusted R^2 the more commonly used metric. The R^2 can also determine the value of a variable in a model. A rough method is to remove a variable from a multiple regression and determine the impact on the overall R^2; if the R^2 does not change, the variable was likely not too important.

For the 2-year yield example, probably the most fundamental driver would be the prevailing Federal Reserve target interest rate. One point to note here is that we need to select an explanatory variable about which we may be able to form reasonably accurate predictions or use lags (more on this later); more importantly, we have a reasonable expectation of causality of the variable. We would not want to use a variable such as the lunar cycle in our prediction for 2-year yields; however, one can expect the federal funds rate, on the other hand, to have a direct causal relation to interest rates of farther out maturities. After the explanatory variable has been selected, we can calculate the correlation between the 2-year yield and the federal (fed) funds target rate, possibly using a software package. Another variable that may be relevant for the 2-year Treasury could be the Standard & Poor's (S&P) 500 index since money tends to flow between equities and Treasuries depending on risk appetite. Figure 1.4 shows the

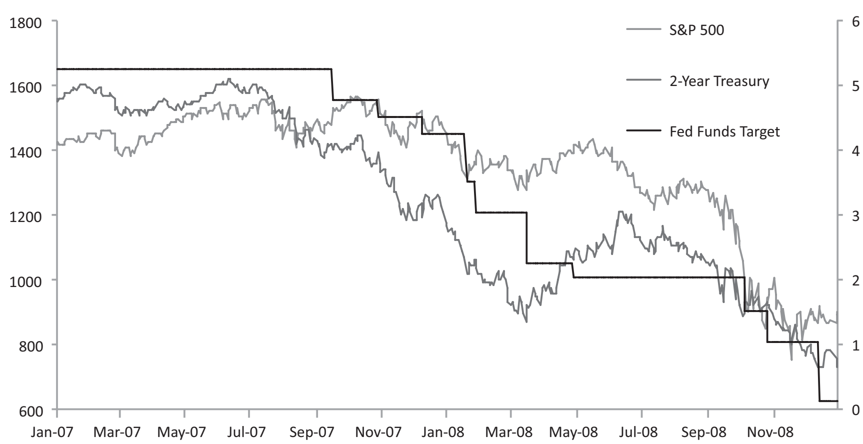

FIGURE 1.4 Path of Fed Funds, Treasury Yields, and Equities over Time
Source: Board of Governors of the Federal Reserve.

three variables alongside each other; the close relationship among the three variables is clear. The correlation between the 2-year Treasury yield and the fed funds target is 85%, while the correlation between the 2-year Treasury and S&P 500 is 61%. As a multiple regression, the combined R^2 for the multiple regression is 86%. Note that in the pairwise regressions, the fed funds versus 2-year Treasury regression is 85%; adding the S&P 500 increases it only to 86%, which means that the S&P 500 adds less "new information" in addition to the fed funds rate. Although this description is mathematically vague, this is a quick way to roughly figure out which variables matter in a multivariable regression.

The concept of correlation is relatively simple to grasp, and given the proliferation of computers, one rarely has to calculate it by hand. However, issues can arise when using correlation inappropriately in financial markets. Since understanding relationships between market variables is very important to understanding markets overall, misinterpreting correlation can be quite risky. One issue with correlation is that it is susceptible to erroneous signals in *nonstationary* variables. Stationary variables are those whose average and variance roughly stay the same over time. If a variable is trending higher consistently, for example, it is not stationary. Figure 1.5 shows a regression of two variables, the 2-year Treasury yield and the Japanese yen currency. The R^2 of the regression of the levels of the two variables is 70%, suggesting a strong link between the variables. To be sure, there may be some relationship between the variables dealing with cross-border flows and the search for higher interest rates. However, this correlation may arise in part from pure trending in the 2007 to 2008 time

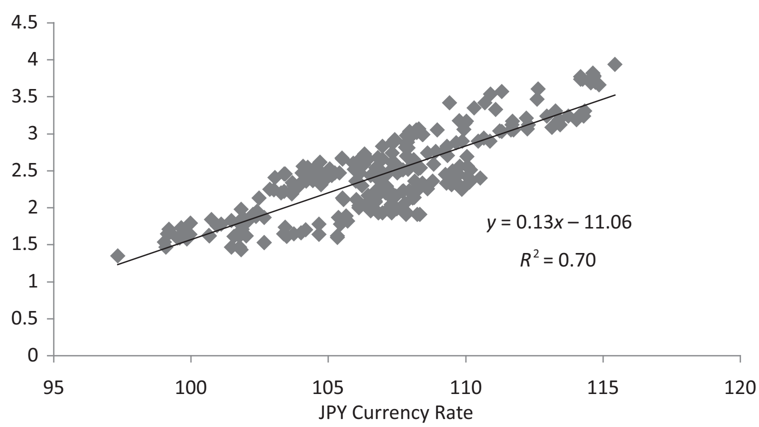

FIGURE 1.5 2Y Treasury Yield Level Regressed against the Japanese Yen Level
Sources: Board of Governors of the Federal Reserve, Bloomberg LP.

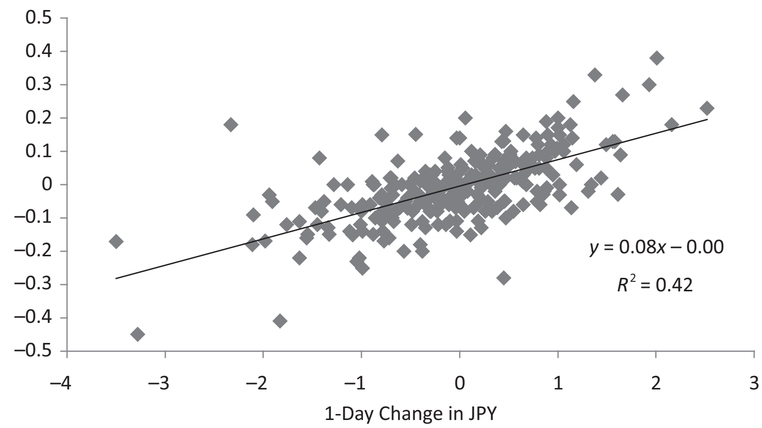

FIGURE 1.6 2Y Treasury Yield Changes Regressed against Japanese Yield Changes
Sources: Board of Governors of the Federal Reserve, Bloomberg LP.

period used for the regression, since the economic turmoil at the time caused many variables to move in similar trends even if they had little causal link. To test the link between the yen and 2-year yields in a more rigorous way, a better regression is between daily changes of the two variables. The use of daily changes reduces the chance of spurious trends creating false correlations. Figure 1.6 shows the regression of daily changes between the yen and 2-year Treasury yield. Now the R^2 is at 42%, which is still relatively high but significantly lower than the 70% in the levels regression. This suggests that some of the relationship in the levels of the two variables is due to trending and some to causality; in general for regressions, it is always prudent to check regressions between changes or percentage changes to control for spurious trends.

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis (PCA) is a statistical technique used to simplify multidimensional data sets that are highly correlated. In informal terms, PCA calculates a few underlying variables that describe complicated systems where the variables are closely related. PCA has become increasingly common in the rates market lately as it is an ideal technique to simplify the interest rate market. After all, the interest rate market is composed of interest rates of different time frames, such as 2-year, 5-year, and 10-year maturities, all of which move closely with each other. Instead of using regressions of each of these rates, PCA breaks down the system into a simpler two to three variables that describe the system almost as well as using the whole set of factors. The principal components are ranked in order of how much they explain the system. For example, the first principal component is the linear combination of interest rates, which describes the maximum amount of variation of the interest rate market; the second principal component describes the maximum amount of variation after the effect of the first principal component is taken out; and so on. These reduced variables tend to be linear weighted averages of the full set of variables, such as the weighted average of the 2-, 5-, 10-, and 30-year rates. By using PCAs to describe the various interest rates in the market, investors can also find those rates that seem unusually high or low, thus pointing to opportunities for profit. Thus, PCA can be used as a first step to find mispriced trades; once these are found, the fundamental and technical factors behind them are needed to eliminate trades that are mispriced for a good reason. A detailed discussion of the mathematics underlying PCA is beyond the scope of this book; for more information on PCA and other multivariable statistical methods, see *Applied Multivariate Statistical Analysis* by Härdle and Simar.

SCALING THROUGH TIME

A common assumption behind financial mathematics is that returns follow a Brownian motion. A random series of returns can be thought of as a random walk—that is, movements up or down based on an outcome from a probability distribution. Without digging deeply into the mathematics, Brownian motion can be understood as the random walk stemming from a normal distribution. One of the ideas that stems from the theory of Brownian motion is that variance scales with time; for instance, the passage of time grows the variance of the process by a proportional factor. Thus, the standard deviation grows with the square root of time. For example, if the daily standard deviation of a Brownian motion series is 1%, the three-month standard deviation would be $1\% \times \sqrt{60}$ since three months is 60 business days, while the one-year standard deviation is $1\% \times \sqrt{251}$, or 15.84% (since there are 251 business days in one year). These are the standard deviation scales for a Brownian motion—for example, if the one-year standard deviation is greater than 15.84%, the series is trending, while a one-year standard deviation less than 15.84% implies a stronger mean reversion tendency.

P1: OTA
c01 JWBT428-Jha January 20, 2011 14:51 Printer: Courier Westford

BACKTESTING STRATEGIES

Statistical tools can help formalize trading rules to provide objective signals. Even if signals produced by trading rules are not implemented blindly, they can provide a useful way of verifying whether certain rules have resulted in effective profits in the past. One very simple rule, for example, can be formulated as: "If bond prices fell yesterday, sell bonds today with the expectation of a further decline." Such a strategy is very simplistic and unlikely to be very effective, but it is an example of a trading rule whose performance can be tested in the past. Various statistical methods, such as regressions, combined with relevant fundamental driving factors, can be useful in building models that eventually may give trading signals. These signals can be tested in the past. Once a trading rule is used to generate past results, various statistical quantities can measure how good the rule is in producing profits. The simplest way to evaluate the results of an existing strategy or a prospective series of signals is to use average return. However, it is possible that two strategies with the same average return could have very different risks. For example, two strategies with 3% per year return could have 10% per year and 30% per year standard deviation, making the second one much less attractive due to its higher volatility. To account for this, a common metric is to take the ratio of the average to the standard deviation. In the example, the ratios are 3%/10% and 3%/30%, or 0.3 and 0.1, making the first one more attractive. In general, any profit or return needs to be scaled by volatility as it removes the "luck" factor, since increasing risk can generally increase returns. The ratio of average return to standard deviation is known as a Sharpe ratio. At times, the ratio is annualized using the frequency of the trades or observations of returns to enable better comparison with other strategies. For example, if a trading strategy gives a signal once a week, the annualized ratio would use $\sqrt{52}$ for 52 weeks in a year. Most often, at least a Sharpe ratio over 1.0 is needed to consider the strategy as viable, although this threshold can vary depending on the type of trade and risk tolerance. Furthermore, combining different strategies with relatively low individual Sharpe ratios can produce higher Sharpe ratio combinations. What determines how additive new strategies are? In short, it is correlation. Highly correlated strategies combined together will not be very additive for the combined Sharpe ratio, but combining strategies with low correlation can result in the Sharpes adding up, even if the individual ones are low.

The Sharpe ratio is just the starting point in evaluating a trading strategy. Even strategies with relatively high Sharpes can have large drawdowns. *Drawdown* refers to the largest peak-to-trough decline in returns and measures a sort of worst-case scenario for the strategy. A trader may not be willing to accept a high Sharpe strategy with a large drawdown that can deplete capital, leaving little for the strategy. In addition to the overall drawdown, periods where the strategy does poorly must be analyzed closely. In such periods, it is necessary to understand the driving factors contributing to the poor performance; regressions of strategy returns versus market variables can help determine these problem factors. Overall, building trading strategies is as much about thinking of trading rules as it is about controlling risk in an effective way. To do this, a variety of statistical measures should be employed to understand all aspects of a strategy's risk.

SUMMARY

This chapter covered the basic mathematical tools needed to understand data in a somewhat objective manner. In the end, statistics and mathematics are just tools to help understand and forecast relationships between different variables. Methods such as regression can help analyze relationships between multiple variables and to assess the strength of such relationships. Since financial variables tend to be increasingly linked with each other, statistical tools are indispensable in understanding markets. Statistical tools such as Sharpe ratios can also help to separate out the "luck" factor and enable a more objective assessment of the performance of a trading strategy. Finally, each mathematical tool is associated with a set of assumptions and flaws. It is important to be cognizant of these flaws and assumptions to use these tools effectively in analyzing data.