2
A brief overview of the classical linear regression model

Learning Outcomes

In this chapter, you will learn how to

- Derive the OLS formulae for estimating parameters and their standard errors
- Explain the desirable properties that a good estimator should have
- Discuss the factors that affect the sizes of standard errors
- Test hypotheses using the test of significance and confidence interval approaches
- Interpret p-values
- Estimate regression models and test single hypotheses in EViews

---

2.1 What is a regression model?

Regression analysis is almost certainly the most important tool at the econometrician's disposal. But what is regression analysis? In very general terms, regression is concerned with describing and evaluating the *relationship between a given variable and one or more other variables*. More specifically, regression is an attempt to explain movements in a variable by reference to movements in one or more other variables.

To make this more concrete, denote the variable whose movements the regression seeks to explain by y and the variables which are used to explain those variations by $x_1, x_2, \ldots, x_k$. Hence, in this relatively simple setup, it would be said that variations in k variables (the xs) cause changes in some other variable, y. This chapter will be limited to the case where the model seeks to explain changes in only one variable y (although this restriction will be removed in chapter 6).

Box 2.1 Names for y and xs in regression models	
Names for y	*Names for the xs*
Dependent variable	Independent variables
Regressand	Regressors
Effect variable	Causal variables
Explained variable	Explanatory variables

There are various completely interchangeable names for y and the xs, and all of these terms will be used synonymously in this book (see box 2.1).

2.2 Regression versus correlation

All readers will be aware of the notion and definition of *correlation*. The correlation between two variables measures the *degree of linear association* between them. If it is stated that y and x are correlated, it means that y and x are being treated in a completely symmetrical way. Thus, it is not implied that changes in x cause changes in y, or indeed that changes in y cause changes in x. Rather, it is simply stated that there is evidence for a linear relationship between the two variables, and that movements in the two are on average related to an extent given by the correlation coefficient.

In regression, the dependent variable (y) and the independent variable(s) (xs) are treated very differently. The y variable is assumed to be random or 'stochastic' in some way, i.e. to have a *probability distribution*. The x variables are, however, assumed to have fixed ('non-stochastic') values in repeated samples.[1] Regression as a tool is more flexible and more powerful than correlation.

2.3 Simple regression

For simplicity, suppose for now that it is believed that y depends on only one x variable. Again, this is of course a severely restricted case, but the case of more explanatory variables will be considered in the next chapter. Three examples of the kind of relationship that may be of interest include:

[1] Strictly, the assumption that the xs are non-stochastic is stronger than required, an issue that will be discussed in more detail in chapter 4.
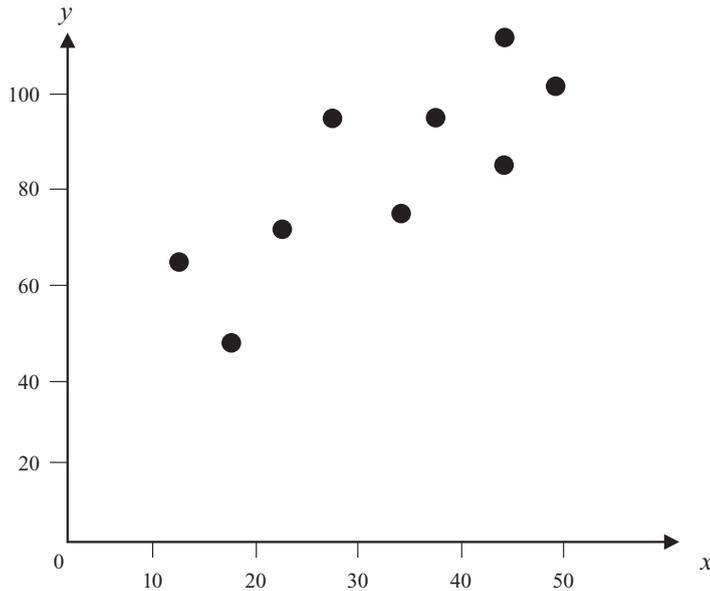
- How asset returns vary with their level of market risk
- Measuring the long-term relationship between stock prices and dividends
- Constructing an optimal hedge ratio.

Suppose that a researcher has some idea that there should be a relationship between two variables y and x, and that financial theory suggests that an increase in x will lead to an increase in y. A sensible first stage to testing whether there is indeed an association between the variables would be to form a scatter plot of them. Suppose that the outcome of this plot is figure 2.1.

In this case, it appears that there is an approximate positive linear relationship between x and y which means that increases in x are usually accompanied by increases in y, and that the relationship between them can be described approximately by a straight line. It would be possible to draw by hand onto the graph a line that appears to fit the data. The intercept and slope of the line fitted by eye could then be measured from the graph. However, in practice such a method is likely to be laborious and inaccurate.

It would therefore be of interest to determine to what extent this relationship can be described by an equation that can be estimated using a defined procedure. It is possible to use the general equation for a straight line

$$y = \alpha + \beta x \tag{2.1}$$

Box 2.2	Reasons for the inclusion of the disturbance term

- Even in the general case where there is more than one explanatory variable, some determinants of y_t will always in practice be omitted from the model. This might, for example, arise because the number of influences on y is too large to place in a single model, or because some determinants of y may be unobservable or not measurable.
- There may be errors in the way that y is measured which cannot be modelled.
- There are bound to be random outside influences on y that again cannot be modelled. For example, a terrorist attack, a hurricane or a computer failure could all affect financial asset returns in a way that cannot be captured in a model and cannot be forecast reliably. Similarly, many researchers would argue that human behaviour has an inherent randomness and unpredictability!

to get the line that best 'fits' the data. The researcher would then be seeking to find the values of the parameters or coefficients, α and β, which would place the line as close as possible to all of the data points taken together.

However, this equation ($y = \alpha + \beta x$) is an exact one. Assuming that this equation is appropriate, if the values of α and β had been calculated, then given a value of x, it would be possible to determine with certainty what the value of y would be. Imagine – a model which says with complete certainty what the value of one variable will be given any value of the other!

Clearly this model is not realistic. Statistically, it would correspond to the case where the model fitted the data perfectly – that is, all of the data points lay exactly on a straight line. To make the model more realistic, a random disturbance term, denoted by u, is added to the equation, thus

$$y_t = \alpha + \beta x_t + u_t \tag{2.2}$$

where the subscript $t (= 1, 2, 3, \ldots)$ denotes the observation number. The disturbance term can capture a number of features (see box 2.2).

So how are the appropriate values of α and β determined? α and β are chosen so that the (vertical) distances from the data points to the fitted lines are minimised (so that the line fits the data as closely as possible). The parameters are thus chosen to minimise collectively the (vertical) distances from the data points to the fitted line. This could be done by 'eye-balling' the data and, for each set of variables y and x, one could form a scatter plot and draw on a line that looks as if it fits the data well by hand, as in figure 2.2.

Note that the *vertical distances* are usually minimised rather than the horizontal distances or those taken perpendicular to the line. This arises

Figure 2.2

Scatter plot of two variables with a line of best fit chosen by eye
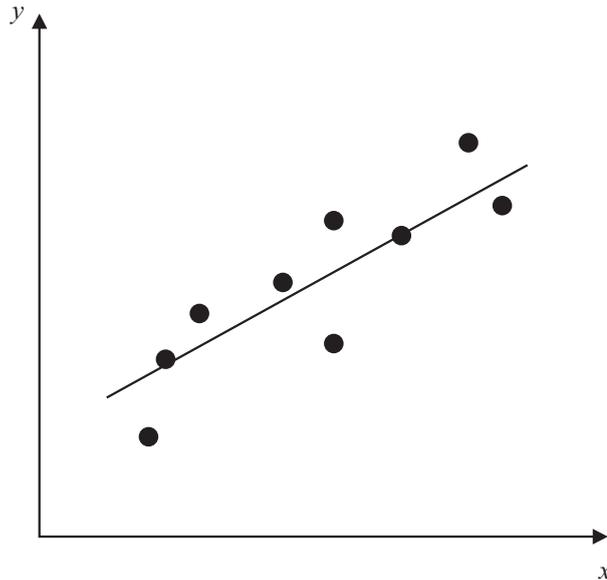
as a result of the assumption that x is fixed in repeated samples, so that the problem becomes one of determining the appropriate model for y given (or conditional upon) the observed values of x.

This 'eye-balling' procedure may be acceptable if only indicative results are required, but of course this method, as well as being tedious, is likely to be imprecise. The most common method used to fit a line to the data is known as ordinary least squares (OLS). This approach forms the workhorse of econometric model estimation, and will be discussed in detail in this and subsequent chapters.

Two alternative estimation methods (for determining the appropriate values of the coefficients α and β) are the method of moments and the method of maximum likelihood. A generalised version of the method of moments, due to Hansen (1982), is popular, but beyond the scope of this book. The method of maximum likelihood is also widely employed, and will be discussed in detail in chapter 8.

Suppose now, for ease of exposition, that the sample of data contains only five observations. The method of OLS entails taking each vertical distance from the point to the line, squaring it and then minimising the total sum of the areas of squares (hence 'least squares'), as shown in figure 2.3. This can be viewed as equivalent to minimising the sum of the areas of the squares drawn from the points to the line.

Tightening up the notation, let y_t denote the actual data point for observation t and let $\hat{y}_t$ denote the fitted value from the regression line – in

Method of OLS
fitting a line to the
data by minimising
the sum of squared
residuals
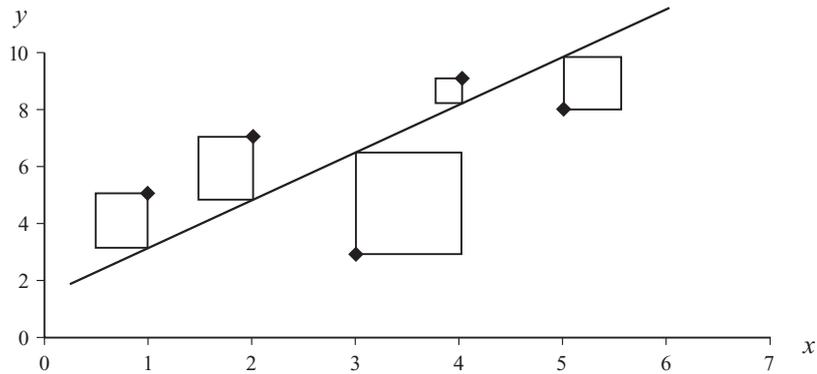
Figure 2.4
Plot of a single
observation,
together with the
line of best fit, the
residual and the
fitted value
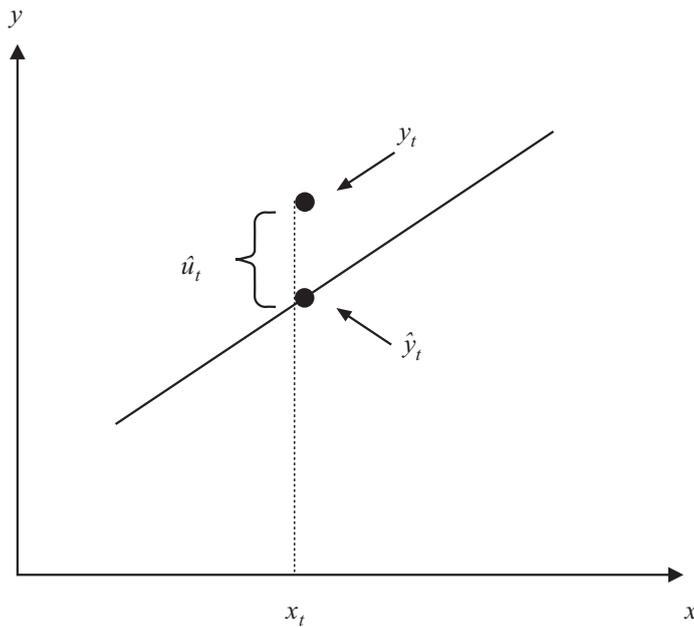
other words, for the given value of x of this observation t, $\hat{y}_t$ is the value
for y which the model would have predicted. Note that a hat (ˆ) over a
variable or parameter is used to denote a value estimated by a model.
Finally, let $\hat{u}_t$ denote the residual, which is the difference between the
actual value of y and the value fitted by the model for this data point –
i.e. $(y_t - \hat{y}_t)$. This is shown for just one observation t in figure 2.4.

   What is done is to minimise the sum of the $\hat{u}_t^2$. The reason that the sum
of the squared distances is minimised rather than, for example, finding
the sum of $\hat{u}_t$ that is as close to zero as possible, is that in the latter case
some points will lie above the line while others lie below it. Then, when
the sum to be made as close to zero as possible is formed, the points

above the line would count as positive values, while those below would count as negatives. So these distances will in large part cancel each other out, which would mean that one could fit virtually any line to the data, so long as the sum of the distances of the points above the line and the sum of the distances of the points below the line were the same. In that case, there would not be a unique solution for the estimated coefficients. In fact, any fitted line that goes through the mean of the observations (i.e. $\bar{x}$, $\bar{y}$) would set the sum of the $\hat{u}_t$ to zero. However, taking the squared distances ensures that all deviations that enter the calculation are positive and therefore do not cancel out.

So minimising the sum of squared distances is given by minimising $(\hat{u}_1^2 + \hat{u}_2^2 + \hat{u}_3^2 + \hat{u}_4^2 + \hat{u}_5^2)$, or minimising

$$\left(\sum_{t=1}^{5} \hat{u}_t^2\right)$$

This sum is known as the *residual sum of squares* (RSS) or the sum of squared residuals. But what is $\hat{u}_t$? Again, it is the difference between the actual point and the line, $y_t - \hat{y}_t$. So minimising $\sum_t \hat{u}_t^2$ is equivalent to minimising $\sum_t (y_t - \hat{y}_t)^2$.

Letting $\hat{\alpha}$ and $\hat{\beta}$ denote the values of α and β selected by minimising the RSS, respectively, the equation for the fitted line is given by $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$. Now let L denote the RSS, which is also known as a loss function. Take the summation over all of the observations, i.e. from $t = 1$ to T, where T is the number of observations

$$L = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta}x_t)^2. \tag{2.3}$$

L is minimised with respect to (w.r.t.) $\hat{\alpha}$ and $\hat{\beta}$, to find the values of α and β which minimise the residual sum of squares to give the line that is closest to the data. So L is differentiated w.r.t. $\hat{\alpha}$ and $\hat{\beta}$, setting the first derivatives to zero. A derivation of the ordinary least squares (OLS) estimator is given in the appendix to this chapter. The coefficient estimators for the slope and the intercept are given by

$$\hat{\beta} = \frac{\sum x_t y_t - T\overline{xy}}{\sum x_t^2 - T\bar{x}^2} \tag{2.4} \qquad\qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{2.5}$$

Equations (2.4) and (2.5) state that, given only the sets of observations x_t and y_t, it is always possible to calculate the values of the two parameters, $\hat{\alpha}$ and $\hat{\beta}$, that best fit the set of data. Equation (2.4) is the easiest formula

Introductory Econometrics for Finance

Table 2.1 Sample data on fund XXX to motivate OLS estimation

Year, t	Excess return on fund XXX $= r_{XXX,t} - rf_t$	Excess return on market index $= rm_t - rf_t$
1	17.8	13.7
2	39.0	23.2
3	12.8	6.9
4	24.2	16.8
5	17.2	12.3

to use to calculate the slope estimate, but the formula can also be written, more intuitively, as

$$\hat{\beta} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} \tag{2.6}$$

which is equivalent to the sample covariance between x and y divided by the sample variance of x.

To reiterate, this method of finding the optimum is known as OLS. It is also worth noting that it is obvious from the equation for $\hat{\alpha}$ that the regression line will go through the mean of the observations – i.e. that the point $(\bar{x}, \bar{y})$ lies on the regression line.

Example 2.1

Suppose that some data have been collected on the excess returns on a fund manager's portfolio ('fund XXX') together with the excess returns on a market index as shown in table 2.1.

The fund manager has some intuition that the beta (in the CAPM framework) on this fund is positive, and she therefore wants to find whether there appears to be a relationship between x and y given the data. Again, the first stage could be to form a scatter plot of the two variables (figure 2.5).

Clearly, there appears to be a positive, approximately linear relationship between x and y, although there is not much data on which to base this conclusion! Plugging the five observations in to make up the formulae given in (2.4) and (2.5) would lead to the estimates $\hat{\alpha} = -1.74$ and $\hat{\beta} = 1.64$. The fitted line would be written as

$$\hat{y}_t = -1.74 + 1.64 x_t \tag{2.7}$$

where x_t is the excess return of the market portfolio over the risk free rate (i.e. $rm - rf$), also known as the *market risk premium*.
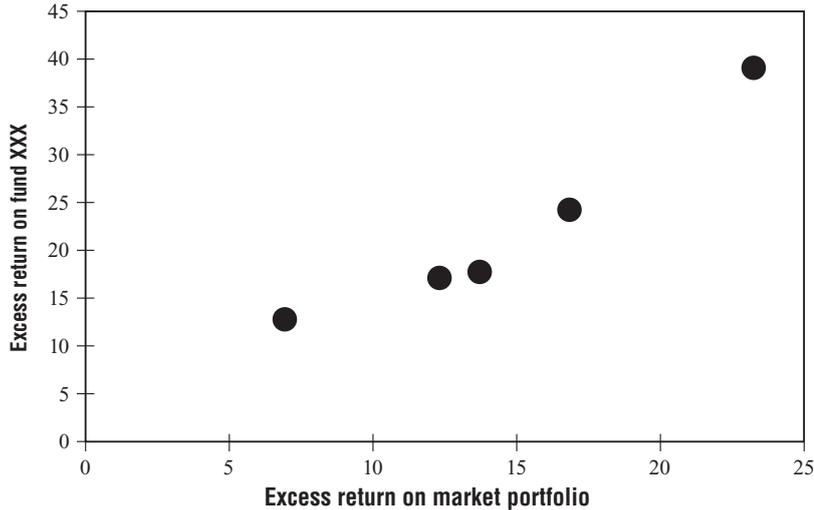
2.3.1 What are $\hat{\alpha}$ and $\hat{\beta}$ used for?

This question is probably best answered by posing another question. If an analyst tells you that she expects the market to yield a return 20% higher than the risk-free rate next year, what would you expect the return on fund XXX to be?

The expected value of $y =$ '$-1.74 + 1.64 \times$ value of x', so plug $x = 20$ into (2.7)

$$\hat{y}_t = -1.74 + 1.64 \times 20 = 31.06 \tag{2.8}$$

Thus, for a given expected market risk premium of 20%, and given its riskiness, fund XXX would be expected to earn an excess over the risk-free rate of approximately 31%. In this setup, the regression beta is also the CAPM beta, so that fund XXX has an estimated beta of 1.64, suggesting that the fund is rather risky. In this case, the residual sum of squares reaches its minimum value of 30.33 with these OLS coefficient values.

Although it may be obvious, it is worth stating that it is not advisable to conduct a regression analysis using only five observations! Thus the results presented here can be considered indicative and for illustration of the technique only. Some further discussions on appropriate sample sizes for regression analysis are given in chapter 4.

The coefficient estimate of 1.64 for β is interpreted as saying that, 'if x increases by 1 unit, y will be expected, everything else being equal, to increase by 1.64 units'. Of course, if $\hat{\beta}$ had been negative, a rise in x would on average cause a fall in y. $\hat{\alpha}$, the intercept coefficient estimate, is
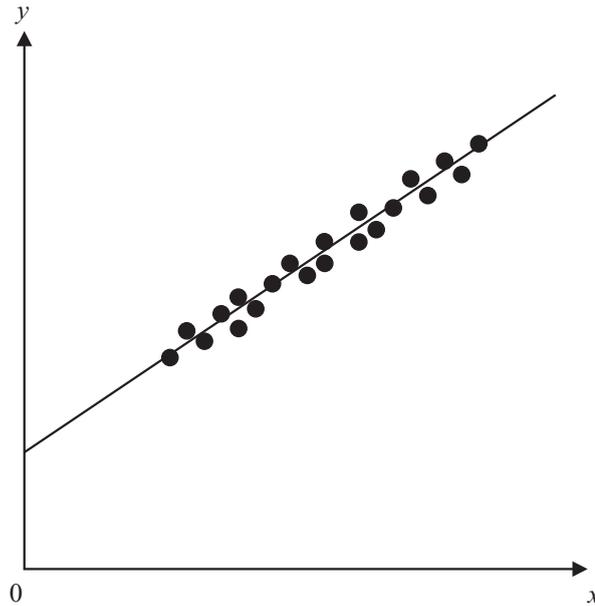
interpreted as the value that would be taken by the dependent variable y if the independent variable x took a value of zero. 'Units' here refer to the units of measurement of x_t and y_t. So, for example, suppose that $\hat{\beta} = 1.64$, x is measured in per cent and y is measured in thousands of US dollars. Then it would be said that if x rises by 1%, y will be expected to rise on average by $1.64 thousand (or $1,640). Note that changing the scale of y or x will make no difference to the overall results since the coefficient estimates will change by an off-setting factor to leave the overall relationship between y and x unchanged (see Gujarati, 2003, pp. 169–173 for a proof). Thus, if the units of measurement of y were hundreds of dollars instead of thousands, and everything else remains unchanged, the slope coefficient estimate would be 16.4, so that a 1% increase in x would lead to an increase in y of $16.4 hundreds (or $1,640) as before. All other properties of the OLS estimator discussed below are also invariant to changes in the scaling of the data.

A word of caution is, however, in order concerning the reliability of estimates of the constant term. Although the strict interpretation of the intercept is indeed as stated above, in practice, it is often the case that there are no values of x close to zero in the sample. In such instances, estimates of the value of the intercept will be unreliable. For example, consider figure 2.6, which demonstrates a situation where no points are close to the y-axis.

In such cases, one could not expect to obtain robust estimates of the value of y when x is zero as all of the information in the sample pertains to the case where x is considerably larger than zero.

A similar caution should be exercised when producing predictions for y using values of x that are a long way outside the range of values in the sample. In example 2.1, x takes values between 7% and 23% in the available data. So, it would not be advisable to use this model to determine the expected excess return on the fund if the expected excess return on the market were, say 1% or 30%, or -5% (i.e. the market was expected to fall).

2.4 Some further terminology

2.4.1 *The population and the sample*

The population is the *total collection of all objects or people to be studied*. For example, in the context of determining the relationship between risk and return for UK equities, the population of interest would be all time series observations on all stocks traded on the London Stock Exchange (LSE).

The population may be either finite or infinite, while a sample is a selection of *just some items from the population*. In general, either all of the observations for the entire population will not be available, or they may be so many in number that it is infeasible to work with them, in which case a sample of data is taken for analysis. The sample is usually *random*, and it should be *representative* of the population of interest. A random sample is a sample in which each individual item in the population is equally likely to be drawn. The size of the sample is the number of observations that are available, or that it is decided to use, in estimating the regression equation.

2.4.2 *The data generating process, the population regression function and the sample regression function*

The population regression function (PRF) is a description of the model that is thought to be generating the actual data and it represents the *true relationship between the variables*. The population regression function is also known as the data generating process (DGP). The PRF embodies the true values of α and β, and is expressed as

$$y_t = \alpha + \beta x_t + u_t \tag{2.9}$$

Note that there is a disturbance term in this equation, so that even if one had at one's disposal the entire population of observations on x and y,

it would still in general not be possible to obtain a perfect fit of the line to the data. In some textbooks, a distinction is drawn between the PRF (the underlying true relationship between y and x) and the DGP (the process describing the way that the actual observations on y come about), although in this book, the two terms will be used synonymously.

The sample regression function, SRF, is the relationship that has been estimated using the sample observations, and is often written as

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t \tag{2.10}$$

Notice that there is no error or residual term in (2.10); all this equation states is that given a particular value of x, multiplying it by $\hat{\beta}$ and adding $\hat{\alpha}$ will give the model fitted or expected value for y, denoted $\hat{y}$. It is also possible to write

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t \tag{2.11}$$

Equation (2.11) splits the observed value of y into two components: the fitted value from the model, and a residual term.

The SRF is used to infer likely values of the PRF. That is, the estimates $\hat{\alpha}$ and $\hat{\beta}$ are constructed, for the sample of data at hand, but what is really of interest is the true relationship between x and y – in other words, the PRF is what is really wanted, but all that is ever available is the SRF! However, what can be said is how likely it is, given the figures calculated for $\hat{\alpha}$ and $\hat{\beta}$, that the corresponding population parameters take on certain values.

2.4.3 Linearity and possible forms for the regression function

In order to use OLS, a model that is *linear* is required. This means that, in the simple bivariate case, the relationship between x and y must be capable of being expressed diagramatically using a straight line. More specifically, the model must be linear in the parameters (α and β), but it does not necessarily have to be linear in the variables (y and x). By 'linear in the parameters', it is meant that the parameters are not multiplied together, divided, squared, or cubed, etc.

Models that are not linear in the variables can often be made to take a linear form by applying a suitable transformation or manipulation. For example, consider the following exponential regression model

$$Y_t = A X_t^{\beta} e^{u_t} \tag{2.12}$$

Taking logarithms of both sides, applying the laws of logs and rearranging the right-hand side (RHS)

$$\ln Y_t = \ln(A) + \beta \ln X_t + u_t \tag{2.13}$$

where A and β are parameters to be estimated. Now let $\alpha = \ln(A)$, $y_t = \ln Y_t$ and $x_t = \ln X_t$

$$y_t = \alpha + \beta x_t + u_t \tag{2.14}$$

This is known as an *exponential regression model* since Y varies according to some exponent (power) function of X. In fact, when a regression equation is expressed in 'double logarithmic form', which means that both the dependent and the independent variables are natural logarithms, the coefficient estimates are interpreted as elasticities (strictly, they are unit changes on a logarithmic scale). Thus a coefficient estimate of 1.2 for $\hat{\beta}$ in (2.13) or (2.14) is interpreted as stating that 'a rise in X of 1% will lead on average, everything else being equal, to a rise in Y of 1.2%'. Conversely, for y and x in levels rather than logarithmic form (e.g. (2.9)), the coefficients denote unit changes as described above.

Similarly, if theory suggests that x should be inversely related to y according to a model of the form

$$y_t = \alpha + \frac{\beta}{x_t} + u_t \tag{2.15}$$

the regression can be estimated using OLS by setting

$$z_t = \frac{1}{x_t}$$

and regressing y on a constant and z. Clearly, then, a surprisingly varied array of models can be estimated using OLS by making suitable transformations to the variables. On the other hand, some models are *intrinsically non-linear*, e.g.

$$y_t = \alpha + \beta x_t^\gamma + u_t \tag{2.16}$$

Such models cannot be estimated using OLS, but might be estimable using a non-linear estimation method (see chapter 8).

2.4.4 Estimator or estimate?

Estimators are the formulae used to *calculate the coefficients* – for example, the expressions given in (2.4) and (2.5) above, while the estimates, on the other hand, are the *actual numerical values for the coefficients* that are obtained from the sample.

2.5 Simple linear regression in EViews – estimation of an optimal hedge ratio

This section shows how to run a bivariate regression using EViews. The example considers the situation where an investor wishes to hedge a long position in the S&P500 (or its constituent stocks) using a short position in futures contracts. Many academic studies assume that the objective of hedging is to minimise the variance of the hedged portfolio returns. If this is the case, then the appropriate hedge ratio (the number of units of the futures asset to sell per unit of the spot asset held) will be the slope estimate (i.e. $\hat{\beta}$) in a regression where the dependent variable is a time series of spot returns and the independent variable is a time series of futures returns.[2]

This regression will be run using the file 'SandPhedge.xls', which contains monthly returns for the S&P500 index (in column 2) and S&P500 futures (in column 3). As described in chapter 1, the first step is to open an appropriately dimensioned workfile. **Open EViews** and click on **File/New/Workfile**; choose **Dated – regular frequency** and **Monthly** frequency data. The start date is **2002:02** and the end date is **2007:07**. Then import the Excel file by clicking **Import** and **Read Text-Lotus-Excel**. The data start in B2 and as for the previous example in chapter 1, the first column contains only dates which we do not need to read in. In 'Names for series or Number if named in file', we can write **Spot Futures**. The two imported series will now appear as objects in the workfile and can be verified by checking a couple of entries at random against the original Excel file.

The first step is to transform the levels of the two series into percentage returns. It is common in academic research to use continuously compounded returns rather than simple returns. To achieve this (i.e. to produce continuously compounded returns), click on **Genr** and in the 'Enter Equation' dialog box, enter **dfutures=100*dlog(futures)**. Then click **Genr** again and do the same for the spot series: **dspot=100*dlog(spot)**. Do not forget to **Save the workfile**. Continue to re-save it at regular intervals to ensure that no work is lost!

Before proceeding to estimate the regression, now that we have imported more than one series, we can examine a number of descriptive statistics together and measures of association between the series. For example, click **Quick** and **Group Statistics**. From there you will see that it is possible to calculate the covariances or correlations between series and

---

[2] See chapter 8 for a detailed discussion of why this is the appropriate hedge ratio.

a number of other measures that will be discussed later in the book. For now, click on **Descriptive Statistics** and **Common Sample**.[3] In the dialog box that appears, type **rspot rfutures** and click **OK**. Some summary statistics for the spot and futures are presented, as displayed in screenshot 2.1, and these are quite similar across the two series, as one would expect.

	RSPOT	RFUTURES
Mean	0.421203	0.467466
Median	0.993048	0.907114
Maximum	8.291442	6.663863
Minimum	-11.65612	-8.647693
Std. Dev.	3.542992	3.313925
Skewness	-0.778888	-0.862431
Kurtosis	4.603577	3.985059
Jarque-Bera	13.53659	10.68570
Probability	0.001150	0.004782
Sum	27.37817	30.38530
Sum Sq. Dev.	803.3787	702.8542
Observations	65	65

Note that the number of observations has reduced from 66 for the levels of the series to 65 when we computed the returns (as one observation is 'lost' in constructing the $t - 1$ value of the prices in the returns formula). If you want to save the summary statistics, you must name them by clicking **Name** and then choose a name, e.g. **Descstats**. The default name is 'group01', which could have also been used. Click **OK**.

We can now proceed to estimate the regression. There are several ways to do this, but the easiest is to select **Quick** and then **Estimate Equation**. You

---

[3] 'Common sample' will use only the part of the sample that is available for all the series selected, whereas 'Individual sample' will use all available observations for each individual series. In this case, the number of observations is the same for both series and so identical results would be observed for both options.

Equation Estimation

Specification | Options

Equation specification

Dependent variable followed by list of regressors including ARMA
and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.

rspot c rfutures

Estimation settings

Method: LS - Least Squares (NLS and ARMA)

Sample: 2002M02 2007M07

OK   Cancel
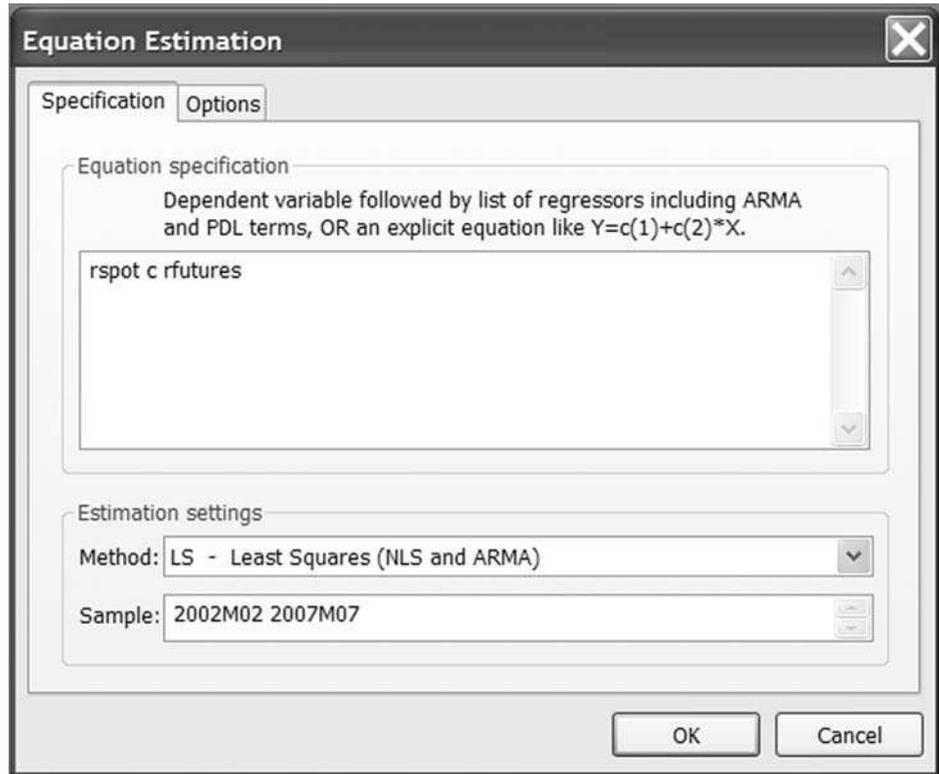
will be presented with a dialog box, which, when it has been completed,
will look like screenshot 2.2.

In the 'Equation Specification' window, you insert the list of variables
to be used, with the dependent variable (y) first, and including a constant
(c), so type **rspot c rfutures**. Note that it would have been possible to write
this in an equation format as rspot = c(1) + c(2)*rfutures, but this is more
cumbersome.

In the 'Estimation settings' box, the default estimation method is OLS
and the default sample is the whole sample, and these need not be modi-
fied. Click **OK** and the regression results will appear, as in screenshot 2.3.

The parameter estimates for the intercept ($\hat{\alpha}$) and slope ($\hat{\beta}$) are 0.36 and
0.12 respectively. Name the regression results **returnreg**, and it will now
appear as a new object in the list. A large number of other statistics are
also presented in the regression output – the purpose and interpretation
of these will be discussed later in this and subsequent chapters.

**Now estimate a regression for the levels of the series rather than
the returns** (i.e. run a regression of spot on a constant and futures) and
examine the parameter estimates. The return regression slope parame-
ter estimated above measures the optimal hedge ratio and also measures

Equation: UNTITLED   Workfile: SANDPHEDGE::Unti...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RSPOT
Method: Least Squares
Date: 08/09/07   Time: 10:17
Sample (adjusted): 2002M03 2007M07
Included observations: 65 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.363302	0.444369	0.817569	0.4167
RFUTURES	0.123860	0.133790	0.925781	0.3581

R-squared	0.013422	Mean dependent var	0.421203
Adjusted R-squared	-0.002238	S.D. dependent var	3.542992
S.E. of regression	3.546955	Akaike info criterion	5.400342
Sum squared resid	792.5960	Schwarz criterion	5.467246
Log likelihood	-173.5111	Hannan-Quinn criter.	5.426740
F-statistic	0.857070	Durbin-Watson stat	2.116689
Prob(F-statistic)	0.358093		

the short run relationship between the two series. By contrast, the slope parameter in a regression using the raw spot and futures indices (or the log of the spot series and the log of the futures series) can be interpreted as measuring the long run relationship between them. This issue of the long and short runs will be discussed in detail in chapter 4. For now, click **Quick/Estimate Equation** and enter the variables **spot c futures** in the Equation Specification dialog box, click **OK**, then **name the regression results 'levelreg'**. The intercept estimate $(\hat{\alpha})$ in this regression is 21.11 and the slope estimate $(\hat{\beta})$ is 0.98. The intercept can be considered to approximate the cost of carry, while as expected, the long-term relationship between spot and futures prices is almost 1:1 – see chapter 7 for further discussion of the estimation and interpretation of this long-term relationship. Finally, click the **Save** button to save the whole workfile.

2.6 The assumptions underlying the classical linear regression model

The model $y_t = \alpha + \beta x_t + u_t$ that has been derived above, together with the assumptions listed below, is known as the *classical linear regression model*

Box 2.3	Assumptions concerning disturbance terms and their interpretation

Technical notation	Interpretation
(1) $E(u_t) = 0$	The errors have zero mean
(2) $\text{var}(u_t) = \sigma^2 < \infty$	The variance of the errors is constant and finite over all values of x_t
(3) $\text{cov}(u_i, u_j) = 0$	The errors are linearly independent of one another
(4) $\text{cov}(u_t, x_t) = 0$	There is no relationship between the error and corresponding x variate

(CLRM). Data for x_t is observable, but since y_t also depends on u_t, it is necessary to be specific about how the u_t are generated. The set of assumptions shown in box 2.3 are usually made concerning the u_ts, the unobservable error or disturbance terms. Note that no assumptions are made concerning their observable counterparts, the estimated model's residuals.

As long as assumption 1 holds, assumption 4 can be equivalently written $E(x_t u_t) = 0$. Both formulations imply that the regressor is *orthogonal* to (i.e. unrelated to) the error term. An alternative assumption to 4, which is slightly stronger, is that the x_t are *non-stochastic* or fixed in repeated samples. This means that there is no sampling variation in x_t, and that its value is determined outside the model.

A fifth assumption is required to make valid inferences about the population parameters (the actual α and β) from the sample parameters ($\hat{\alpha}$ and $\hat{\beta}$) estimated using a finite amount of data:

(5)$u_t \sim N(0, \sigma^2)$ – i.e. that u_t is normally distributed

2.7 Properties of the OLS estimator

If assumptions 1–4 hold, then the estimators $\hat{\alpha}$ and $\hat{\beta}$ determined by OLS will have a number of desirable properties, and are known as Best Linear Unbiased Estimators (BLUE). What does this acronym stand for?

- 'Estimator' – $\hat{\alpha}$ and $\hat{\beta}$ are estimators of the true value of α and β
- 'Linear' – $\hat{\alpha}$ and $\hat{\beta}$ are linear estimators – that means that the formulae for $\hat{\alpha}$ and $\hat{\beta}$ are linear combinations of the random variables (in this case, y)
- 'Unbiased' – on average, the actual values of $\hat{\alpha}$ and $\hat{\beta}$ will be equal to their true values

- 'Best' – means that the OLS estimator $\hat{\beta}$ has minimum variance among the class of linear unbiased estimators; the Gauss–Markov theorem proves that the OLS estimator is best by examining an arbitrary alternative linear unbiased estimator and showing in all cases that it must have a variance no smaller than the OLS estimator.

Under assumptions 1–4 listed above, the OLS estimator can be shown to have the desirable properties that it is consistent, unbiased and efficient. Unbiasedness and efficiency have already been discussed above, and consistency is an additional desirable property. These three characteristics will now be discussed in turn.

2.7.1 Consistency

The least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ are consistent. One way to state this algebraically for $\hat{\beta}$ (with the obvious modifications made for $\hat{\alpha}$) is

$$\lim_{T \to \infty} \Pr[|\hat{\beta} - \beta| > \delta] = 0 \quad \forall \, \delta > 0 \tag{2.17}$$

This is a technical way of stating that the probability (Pr) that $\hat{\beta}$ is more than some arbitrary fixed distance δ away from its true value tends to zero as the sample size tends to infinity, for all positive values of δ. In the limit (i.e. for an infinite number of observations), the probability of the estimator being different from the true value is zero. That is, the estimates will converge to their true values as the sample size increases to infinity. Consistency is thus a large sample, or asymptotic property. The assumptions that $\mathrm{E}(x_t u_t) = 0$ and $E(u_t) = 0$ are sufficient to derive the consistency of the OLS estimator.

2.7.2 Unbiasedness

The least squares estimates of $\hat{\alpha}$ and $\hat{\beta}$ are unbiased. That is

$$E(\hat{\alpha}) = \alpha \tag{2.18}$$

and

$$E(\hat{\beta}) = \beta \tag{2.19}$$

Thus, on average, the estimated values for the coefficients will be equal to their true values. That is, there is no systematic overestimation or underestimation of the true coefficients. To prove this also requires the assumption that $\mathrm{cov}(u_t, \, x_t) = 0$. Clearly, unbiasedness is a stronger condition than consistency, since it holds for small as well as large samples (i.e. for all sample sizes).

2.7.3 *Efficiency*

An estimator $\hat{\beta}$ of a parameter β is said to be efficient if no other estimator has a smaller variance. Broadly speaking, if the estimator is efficient, it will be minimising the probability that it is a long way off from the true value of β. In other words, if the estimator is 'best', the uncertainty associated with estimation will be minimised for the class of linear unbiased estimators. A technical way to state this would be to say that an efficient estimator would have a probability distribution that is narrowly dispersed around the true value.

2.8 Precision and standard errors

Any set of regression estimates $\hat{\alpha}$ and $\hat{\beta}$ are specific to the sample used in their estimation. In other words, if a different sample of data was selected from within the population, the data points (the x_t and y_t) will be different, leading to different values of the OLS estimates.

Recall that the OLS estimators ($\hat{\alpha}$ and $\hat{\beta}$) are given by (2.4) and (2.5). It would be desirable to have an idea of how 'good' these estimates of α and β are in the sense of having some measure of the reliability or precision of the estimators ($\hat{\alpha}$ and $\hat{\beta}$). It is thus useful to know whether one can have confidence in the estimates, and whether they are likely to vary much from one sample to another sample within the given population. An idea of the sampling variability and hence of the precision of the estimates can be calculated using only the sample of data available. This estimate is given by its standard error. Given assumptions 1–4 above, valid estimators of the standard errors can be shown to be given by

$$SE(\hat{\alpha}) = s\sqrt{\frac{\sum x_t^2}{T\sum(x_t - \bar{x})^2}} = s\sqrt{\frac{\sum x_t^2}{T\left(\left(\sum x_t^2\right) - T\bar{x}^2\right)}} \tag{2.20}$$

$$SE(\hat{\beta}) = s\sqrt{\frac{1}{\sum(x_t - \bar{x})^2}} = s\sqrt{\frac{1}{\sum x_t^2 - T\bar{x}^2}} \tag{2.21}$$

where s is the estimated standard deviation of the residuals (see below). These formulae are derived in the appendix to this chapter.

It is worth noting that the standard errors give only a general indication of the likely accuracy of the regression parameters. They do not show how accurate a particular set of coefficient estimates is. If the standard errors are small, it shows that the coefficients are likely to be precise on average, not how precise they are for this particular sample. Thus standard errors give a measure of the *degree of uncertainty* in the estimated

values for the coefficients. It can be seen that they are a function of the actual observations on the explanatory variable, x, the sample size, T, and another term, s. The last of these is an estimate of the variance of the disturbance term. The actual variance of the disturbance term is usually denoted by σ^2. How can an estimate of σ^2 be obtained?

2.8.1 Estimating the variance of the error term (σ^2)

From elementary statistics, the variance of a random variable u_t is given by

$$\text{var}(u_t) = E[(u_t) - E(u_t)]^2 \tag{2.22}$$

Assumption 1 of the CLRM was that the expected or average value of the errors is zero. Under this assumption, (2.22) above reduces to

$$\text{var}(u_t) = E[u_t^2] \tag{2.23}$$

So what is required is an estimate of the average value of u_t^2, which could be calculated as

$$s^2 = \frac{1}{T} \sum u_t^2 \tag{2.24}$$

Unfortunately (2.24) is not workable since u_t is a series of population disturbances, which is not observable. Thus the sample counterpart to u_t, which is $\hat{u}_t$, is used

$$s^2 = \frac{1}{T} \sum \hat{u}_t^2 \tag{2.25}$$

But this estimator is a biased estimator of σ^2. An unbiased estimator, s^2, would be given by the following equation instead of the previous one

$$s^2 = \frac{\sum \hat{u}_t^2}{T-2} \tag{2.26}$$

where $\sum \hat{u}_t^2$ is the residual sum of squares, so that the quantity of relevance for the standard error formulae is the square root of (2.26)

$$s = \sqrt{\frac{\sum \hat{u}_t^2}{T-2}} \tag{2.27}$$

s is also known as the *standard error of the regression* or the standard error of the estimate. It is sometimes used as a broad measure of the fit of the regression equation. Everything else being equal, the smaller this quantity is, the closer is the fit of the line to the actual data.

2.8.2 Some comments on the standard error estimators

It is possible, of course, to derive the formulae for the standard errors of the coefficient estimates from first principles using some algebra, and

this is left to the appendix to this chapter. Some general intuition is now given as to why the formulae for the standard errors given by (2.20) and (2.21) contain the terms that they do and in the form that they do. The presentation offered in box 2.4 loosely follows that of Hill, Griffiths and Judge (1997), which is the clearest that this author has seen.

Box 2.4   Standard error estimators

(1) The larger the sample size, T, the smaller will be the coefficient standard errors. T appears explicitly in $SE(\hat{\alpha})$ and implicitly in $SE(\hat{\beta})$. T appears implicitly since the sum $\sum (x_t - \bar{x})^2$ is from $t = 1$ to T. The reason for this is simply that, at least for now, it is assumed that every observation on a series represents a piece of useful information which can be used to help determine the coefficient estimates. So the larger the size of the sample, the more information will have been used in estimation of the parameters, and hence the more confidence will be placed in those estimates.

(2) Both $SE(\hat{\alpha})$ and $SE(\hat{\beta})$ depend on s^2 (or s). Recall from above that s^2 is the estimate of the error variance. The larger this quantity is, the more dispersed are the residuals, and so the greater is the uncertainty in the model. If s^2 is large, the data points are collectively a long way away from the line.

(3) The sum of the squares of the x_t about their mean appears in both formulae – since $\sum (x_t - \bar{x})^2$ appears in the denominators. The larger the sum of squares, the smaller the coefficient variances. Consider what happens if $\sum (x_t - \bar{x})^2$ is small or large, as shown in figures 2.7 and 2.8, respectively.

In figure 2.7, the data are close together so that $\sum (x_t - \bar{x})^2$ is small. In this first case, it is more difficult to determine with any degree of certainty exactly where the line should be. On the other hand, in figure 2.8, the points are widely dispersed

---

Figure 2.7

Effect on the standard errors of the coefficient estimates when $(x_t - \bar{x})$ are narrowly dispersed
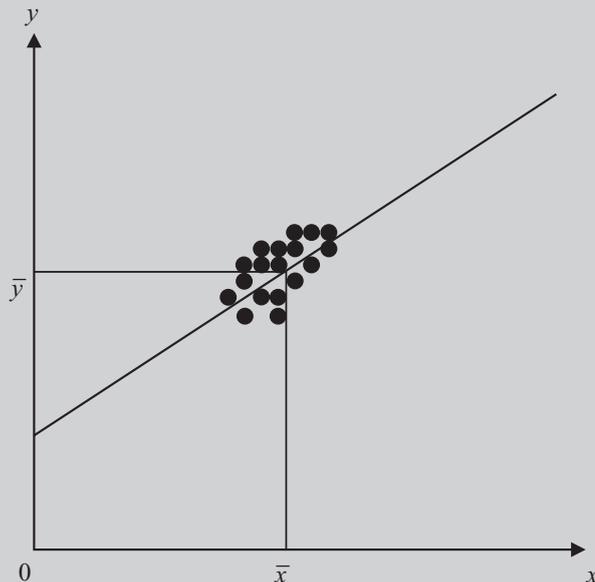
across a long section of the line, so that one could hold more confidence in the estimates in this case.

(4) The term $\sum x_t^2$ affects only the intercept standard error and not the slope standard error. The reason is that $\sum x_t^2$ measures how far the points are away from the *y*-axis. Consider figures 2.9 and 2.10.

In figure 2.9, all of the points are bunched a long way from the *y*-axis, which makes it more difficult to accurately estimate the point at which the estimated line crosses the *y*-axis (the intercept). In figure 2.10, the points collectively are closer to

Figure 2.8

Effect on the standard errors of the coefficient estimates when $(x_t - \bar{x})$ are widely dispersed
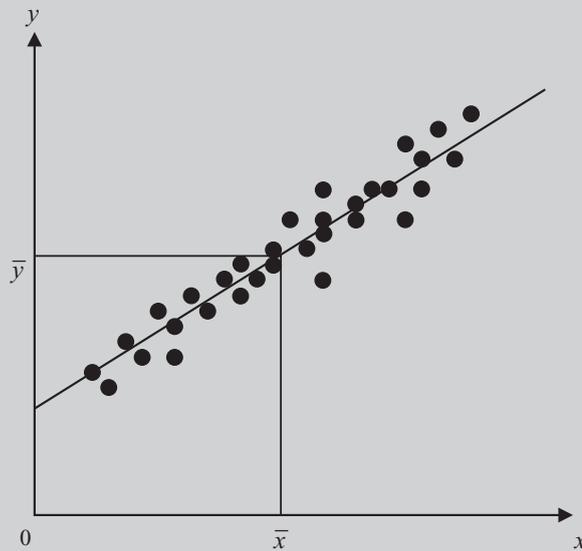
Figure 2.9

Effect on the standard errors of x_t^2 large
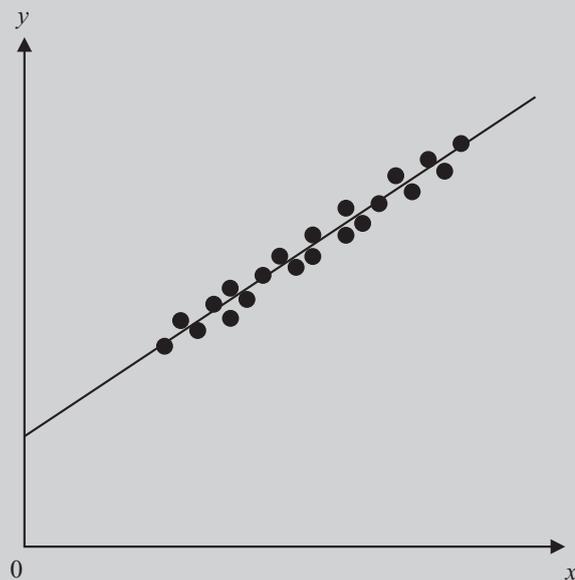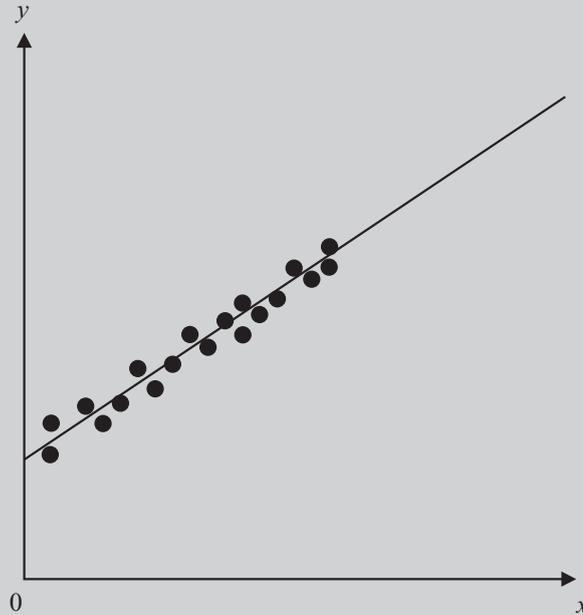
the *y*-axis and hence it will be easier to determine where the line actually crosses the axis. Note that this intuition will work only in the case where all of the x_t are positive!

Example 2.2

Assume that the following data have been calculated from a regression of *y* on a single variable *x* and a constant over 22 observations

$$\sum x_t y_t = 830102, \quad T = 22, \quad \bar{x} = 416.5, \quad \bar{y} = 86.65,$$
$$\sum x_t^2 = 3919654, \quad RSS = 130.6$$

Determine the appropriate values of the coefficient estimates and their standard errors.

This question can simply be answered by plugging the appropriate numbers into the formulae given above. The calculations are

$$\hat{\beta} = \frac{830102 - (22 \times 416.5 \times 86.65)}{3919654 - 22 \times (416.5)^2} = 0.35$$
$$\hat{\alpha} = 86.65 - 0.35 \times 416.5 = -59.12$$

The sample regression function would be written as

$$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$$
$$\hat{y}_t = -59.12 + 0.35 x_t$$

Now, turning to the standard error calculations, it is necessary to obtain an estimate, s, of the error variance

$$SE\ (\textit{regression}),\ s = \sqrt{\frac{\sum \hat{u}_t^2}{T-2}} = \sqrt{\frac{130.6}{20}} = 2.55$$

$$SE(\hat{\alpha}) = 2.55 \times \sqrt{\frac{3919654}{22 \times (3919654 - 22 \times 416.5^2)}} = 3.35$$

$$SE(\hat{\beta}) = 2.55 \times \sqrt{\frac{1}{3919654 - 22 \times 416.5^2}} = 0.0079$$

With the standard errors calculated, the results are written as

$$\hat{y}_t = -59.12 + 0.35 x_t$$
$$\quad\ (3.35)\quad (0.0079)$$

(2.28)

The standard error estimates are usually placed in parentheses under the relevant coefficient estimates.

2.9 An introduction to statistical inference

Often, financial theory will suggest that certain coefficients should take on particular values, or values within a given range. It is thus of interest to determine whether the relationships expected from financial theory are upheld by the data to hand or not. Estimates of α and β have been obtained from the sample, but these values are not of any particular interest; the population values that describe the true relationship between the variables would be of more interest, but are never available. Instead, inferences are made concerning the likely population values from the regression parameters that have been estimated from the sample of data to hand. In doing this, the aim is to determine whether the differences between the coefficient estimates that are actually obtained, and expectations arising from financial theory, are a long way from one another in a statistical sense.

Example 2.3

Suppose the following regression results have been calculated:

$$\hat{y}_t = 20.3 + 0.5091 x_t$$
$$\quad (14.38)\ (0.2561)$$

(2.29)

$\hat{\beta} = 0.5091$ is a single (point) estimate of the unknown population parameter, β. As stated above, the reliability of the point estimate is measured

by the coefficient's standard error. The information from one or more of the sample coefficients and their standard errors can be used to make inferences about the population parameters. So the estimate of the slope coefficient is $\hat{\beta} = 0.5091$, but it is obvious that this number is likely to vary to some degree from one sample to the next. It might be of interest to answer the question, 'Is it plausible, given this estimate, that the true population parameter, β, could be 0.5? Is it plausible that β could be 1?', etc. Answers to these questions can be obtained through *hypothesis testing*.

2.9.1 *Hypothesis testing: some concepts*

In the hypothesis testing framework, there are always two hypotheses that go together, known as the *null hypothesis* (denoted H_0 or occasionally H_N) and the *alternative hypothesis* (denoted H_1 or occasionally H_A). The null hypothesis is the statement or the statistical hypothesis that is actually being tested. The alternative hypothesis represents the remaining outcomes of interest.

For example, suppose that given the regression results above, it is of interest to test the hypothesis that the true value of β is in fact 0.5. The following notation would be used.

$$H_0 : \beta = 0.5$$
$$H_1 : \beta \neq 0.5$$

This states that the hypothesis that the true but unknown value of β could be 0.5 is being tested against an alternative hypothesis where β is not 0.5. This would be known as a two-sided test, since the outcomes of both $\beta < 0.5$ and $\beta > 0.5$ are subsumed under the alternative hypothesis.

Sometimes, some prior information may be available, suggesting for example that $\beta > 0.5$ would be expected rather than $\beta < 0.5$. In this case, $\beta < 0.5$ is no longer of interest to us, and hence a one-sided test would be conducted:

$$H_0 : \beta = 0.5$$
$$H_1 : \beta > 0.5$$

Here the null hypothesis that the true value of β is 0.5 is being tested against a one-sided alternative that β is more than 0.5.

On the other hand, one could envisage a situation where there is prior information that $\beta < 0.5$ is expected. For example, suppose that an investment bank bought a piece of new risk management software that is intended to better track the riskiness inherent in its traders' books and that β is some measure of the risk that previously took the value 0.5. Clearly, it would not make sense to expect the risk to have risen, and so

$\beta > 0.5$, corresponding to an increase in risk, is not of interest. In this case, the null and alternative hypotheses would be specified as

$H_0 : \beta = 0.5$
$H_1 : \beta < 0.5$

This prior information should come from the financial theory of the problem under consideration, and not from an examination of the estimated value of the coefficient. Note that there is always an equality under the null hypothesis. So, for example, $\beta < 0.5$ would not be specified under the null hypothesis.

There are two ways to conduct a hypothesis test: via the *test of significance* approach or via the *confidence interval* approach. Both methods centre on a statistical comparison of the estimated value of the coefficient, and its value under the null hypothesis. In very general terms, if the estimated value is a long way away from the hypothesised value, the null hypothesis is likely to be rejected; if the value under the null hypothesis and the estimated value are close to one another, the null hypothesis is less likely to be rejected. For example, consider $\hat{\beta} = 0.5091$ as above. A hypothesis that the true value of β is 5 is more likely to be rejected than a null hypothesis that the true value of β is 0.5. What is required now is a *statistical decision rule* that will permit the formal testing of such hypotheses.

2.9.2 The probability distribution of the least squares estimators

In order to test hypotheses, assumption 5 of the CLRM must be used, namely that $u_t \sim N(0, \sigma^2)$ – i.e. that the error term is normally distributed. The normal distribution is a convenient one to use for it involves only two parameters (its mean and variance). This makes the algebra involved in statistical inference considerably simpler than it otherwise would have been. Since y_t depends partially on u_t, it can be stated that if u_t is normally distributed, y_t will also be normally distributed.

Further, since the least squares estimators are linear combinations of the random variables, i.e. $\hat{\beta} = \sum w_t y_t$, where w_t are effectively weights, and since the weighted sum of normal random variables is also normally distributed, it can be said that the coefficient estimates will also be normally distributed. Thus

$$\hat{\alpha} \sim N(\alpha, \text{var}(\hat{\alpha})) \quad \text{and} \quad \hat{\beta} \sim N(\beta, \text{var}(\hat{\beta}))$$

Will the coefficient estimates still follow a normal distribution if the errors do not follow a normal distribution? Well, briefly, the answer is usually 'yes', provided that the other assumptions of the CLRM hold, and the sample size is sufficiently large. The issue of non-normality, how to test for it, and its consequences, will be further discussed in chapter 4.

Figure 2.11

The normal
distribution
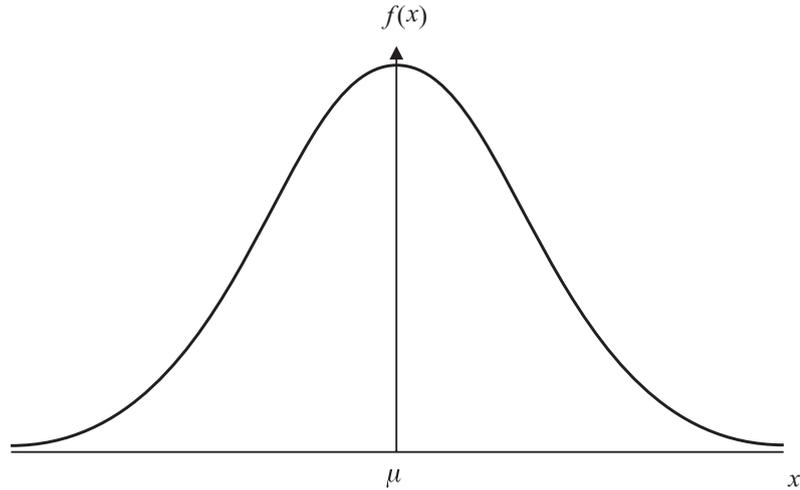
Standard normal variables can be constructed from $\hat{\alpha}$ and $\hat{\beta}$ by subtracting the mean and dividing by the square root of the variance

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\text{var}(\hat{\alpha})}} \sim N(0, 1) \quad \text{and} \quad \frac{\hat{\beta} - \beta}{\sqrt{\text{var}(\hat{\beta})}} \sim N(0, 1)$$

The square roots of the coefficient variances are the standard errors. Unfortunately, the standard errors of the true coefficient values under the PRF are never known – all that is available are their sample counterparts, the calculated standard errors of the coefficient estimates, $SE(\hat{\alpha})$ and $SE(\hat{\beta})$.[4]

Replacing the true values of the standard errors with the sample estimated versions induces another source of uncertainty, and also means that the standardised statistics follow a t-distribution with $T - 2$ degrees of freedom (defined below) rather than a normal distribution, so

$$\frac{\hat{\alpha} - \alpha}{SE(\hat{\alpha})} \sim t_{T-2} \quad \text{and} \quad \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \sim t_{T-2}$$

This result is not formally proved here. For a formal proof, see Hill, Griffiths and Judge (1997, pp. 88–90).

2.9.3 *A note on the t and the normal distributions*

The normal distribution, shown in figure 2.11, should be familiar to readers. Note its characteristic 'bell' shape and its symmetry around the mean (of zero for a standard normal distribution).

---

[4] Strictly, these are the estimated standard errors conditional on the parameter estimates, and so should be denoted $S\hat{E}(\hat{\alpha})$ and $S\hat{E}(\hat{\beta})$, but the additional layer of hats will be omitted here since the meaning should be obvious from the context.

Table 2.2 Critical values from the standard normal versus *t*-distribution

Significance level (%)	$N(0,1)$	t_{40}	t_4
50%	0	0	0
5%	1.64	1.68	2.13
2.5%	1.96	2.02	2.78
0.5%	2.57	2.70	4.60

Figure 2.12

The *t*-distribution versus the normal
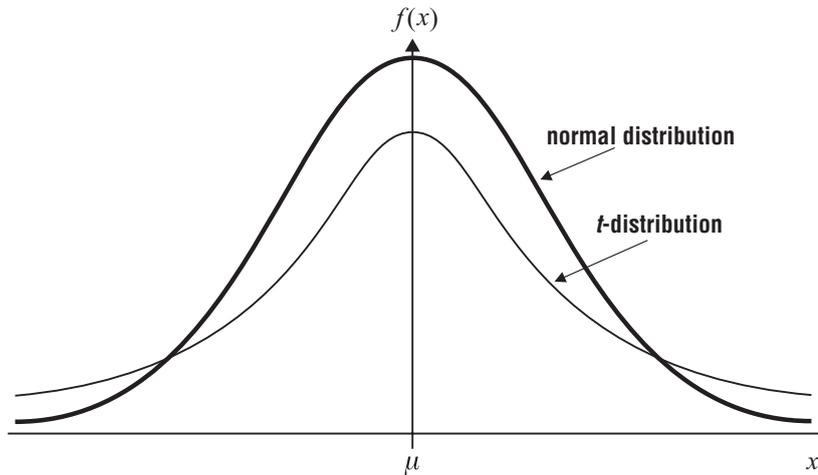
A normal variate can be scaled to have zero mean and unit variance by subtracting its mean and dividing by its standard deviation. There is a specific relationship between the *t*- and the standard normal distribution, and the *t*-distribution has another parameter, its degrees of freedom.

What does the *t*-distribution look like? It looks similar to a normal distribution, but with fatter tails, and a smaller peak at the mean, as shown in figure 2.12.

Some examples of the percentiles from the normal and *t*-distributions taken from the statistical tables are given in table 2.2. When used in the context of a hypothesis test, these percentiles become critical values. The values presented in table 2.2 would be those critical values appropriate for a one-sided test of the given significance level.

It can be seen that as the number of degrees of freedom for the *t*-distribution increases from 4 to 40, the critical values fall substantially. In figure 2.12, this is represented by a gradual increase in the height of the distribution at the centre and a reduction in the fatness of the tails as the number of degrees of freedom increases. In the limit, a *t*-distribution with an infinite number of degrees of freedom is a standard normal, i.e.

$t_\infty = N(0, 1)$, so the normal distribution can be viewed as a special case of the t.

Putting the limit case, t_∞, aside, the critical values for the t-distribution are larger in absolute value than those from the standard normal. This arises from the increased uncertainty associated with the situation where the error variance must be estimated. So now the t-distribution is used, and for a given statistic to constitute the same amount of reliable evidence against the null, it has to be bigger in absolute value than in circumstances where the normal is applicable.

There are broadly two approaches to testing hypotheses under regression analysis: the test of significance approach and the confidence interval approach. Each of these will now be considered in turn.

2.9.4 The test of significance approach

Assume the regression equation is given by $y_t = \alpha + \beta x_t + u_t, t = 1, 2, \ldots, T$. The steps involved in doing a test of significance are shown in box 2.5.

Box 2.5 Conducting a test of significance

(1) Estimate $\hat{\alpha}$, $\hat{\beta}$ and $SE(\hat{\alpha})$, $SE(\hat{\beta})$ in the usual way.
(2) Calculate the test statistic. This is given by the formula

$$test\ statistic = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \qquad (2.30)$$

where β^* is the value of β under the null hypothesis. The null hypothesis is $H_0 : \beta = \beta^*$ and the alternative hypothesis is $H_1 : \beta \neq \beta^*$ (for a two-sided test).
(3) A tabulated distribution with which to compare the estimated test statistics is required. Test statistics derived in this way can be shown to follow a t-distribution with $T - 2$ degrees of freedom.
(4) Choose a 'significance level', often denoted α (*not* the same as the regression intercept coefficient). It is conventional to use a significance level of 5%.
(5) Given a significance level, a *rejection region* and *non-rejection region* can be determined. If a 5% significance level is employed, this means that 5% of the total distribution (5% of the area under the curve) will be in the rejection region. That rejection region can either be split in half (for a two-sided test) or it can all fall on one side of the y-axis, as is the case for a one-sided test.

For a two-sided test, the 5% rejection region is split equally between the two tails, as shown in figure 2.13.

For a one-sided test, the 5% rejection region is located solely in one tail of the distribution, as shown in figures 2.14 and 2.15, for a test where the alternative is of the 'less than' form, and where the alternative is of the 'greater than' form, respectively.

Figure 2.13

Rejection regions for a two-sided 5% hypothesis test

$f(x)$

2.5% rejection region

95% non-rejection region

2.5% rejection region

x
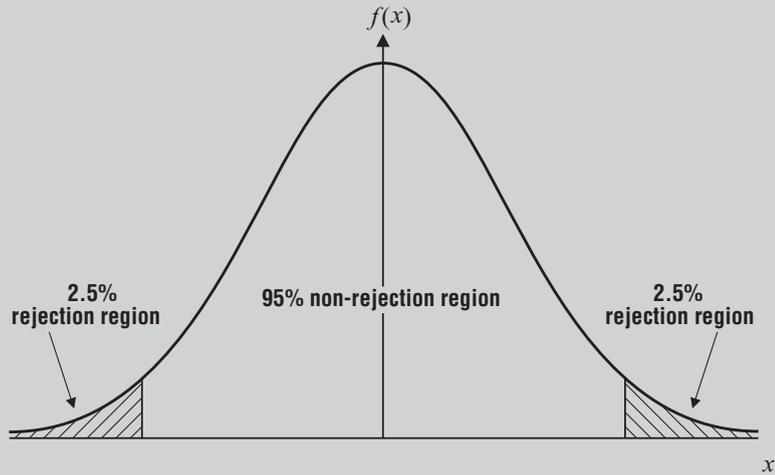
Figure 2.14

Rejection region for a one-sided hypothesis test of the form
$H_0 : \beta = \beta^*,$
$H_1 : \beta < \beta^*$

$f(x)$

5% rejection region

95% non-rejection region

x
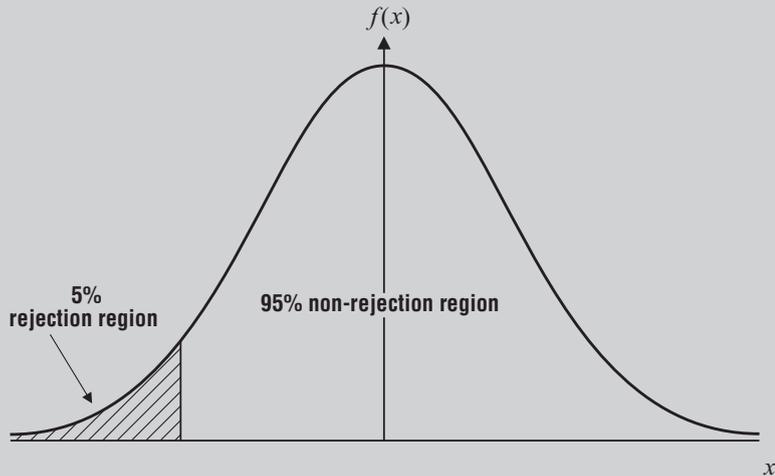
Figure 2.15

Rejection region for a one-sided hypothesis test of the form
$H_0 : \beta = \beta^*,$
$H_1 : \beta > \beta^*$

$f(x)$

95% non-rejection region

5% rejection region

x
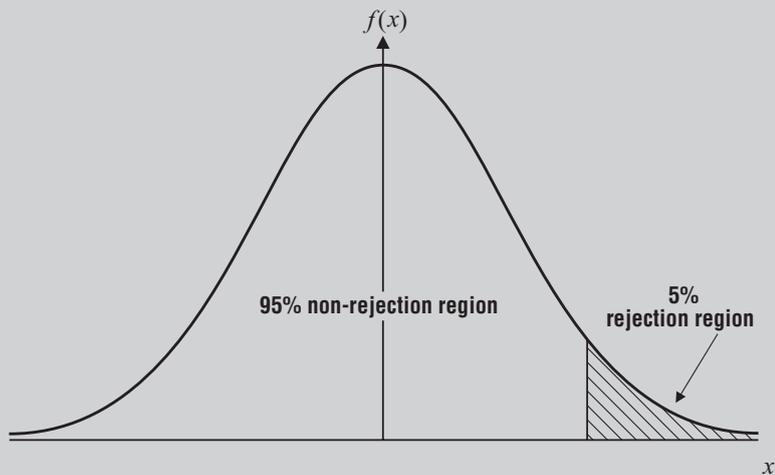
Box 2.5   contd.

(6) Use the t-tables to obtain a critical value or values with which to compare the test statistic. The critical value will be that value of x that puts 5% into the rejection region.

(7) Finally perform the test. If the test statistic lies in the rejection region then reject the null hypothesis (H_0), else do not reject H_0.

Steps 2–7 require further comment. In step 2, the estimated value of β is compared with the value that is subject to test under the null hypothesis, but this difference is 'normalised' or scaled by the standard error of the coefficient estimate. The standard error is a measure of how confident one is in the coefficient estimate obtained in the first stage. If a standard error is small, the value of the test statistic will be large relative to the case where the standard error is large. For a small standard error, it would not require the estimated and hypothesised values to be far away from one another for the null hypothesis to be rejected. Dividing by the standard error also ensures that, under the five CLRM assumptions, the test statistic follows a tabulated distribution.

In this context, the number of degrees of freedom can be interpreted as the number of pieces of additional information beyond the minimum requirement. If two parameters are estimated (α and β – the intercept and the slope of the line, respectively), a minimum of two observations is required to fit this line to the data. As the number of degrees of freedom increases, the critical values in the tables decrease in absolute terms, since less caution is required and one can be more confident that the results are appropriate.

The significance level is also sometimes called the *size of the test* (note that this is completely different from the size of the sample) and it determines the region where the null hypothesis under test will be rejected or not rejected. Remember that the distributions in figures 2.13–2.15 are for a random variable. Purely by chance, a random variable will take on extreme values (either large and positive values or large and negative values) occasionally. More specifically, a significance level of 5% means that a result as extreme as this or more extreme would be expected only 5% of the time as a consequence of chance alone. To give one illustration, if the 5% critical value for a one-sided test is 1.68, this implies that the test statistic would be expected to be greater than this only 5% of the time by chance alone. There is nothing magical about the test – all that is done is to specify an arbitrary cutoff value for the test statistic that determines whether the null hypothesis would be rejected or not. It is conventional to use a 5% size of test, but 10% and 1% are also commonly used.

However, one potential problem with the use of a fixed (e.g. 5%) size of test is that if the sample size is sufficiently large, any null hypothesis can be rejected. This is particularly worrisome in finance, where tens of thousands of observations or more are often available. What happens is that the standard errors reduce as the sample size increases, thus leading to an increase in the value of all *t*-test statistics. This problem is frequently overlooked in empirical work, but some econometricians have suggested that a lower size of test (e.g. 1%) should be used for large samples (see, for example, Leamer, 1978, for a discussion of these issues).

Note also the use of terminology in connection with hypothesis tests: it is said that the null hypothesis is either *rejected* or *not rejected*. It is incorrect to state that if the null hypothesis is not rejected, it is 'accepted' (although this error is frequently made in practice), and it is never said that the alternative hypothesis is accepted or rejected. One reason why it is not sensible to say that the null hypothesis is 'accepted' is that it is impossible to know whether the null is actually true or not! In any given situation, many null hypotheses will not be rejected. For example, suppose that $H_0 : \beta = 0.5$ and $H_0 : \beta = 1$ are separately tested against the relevant two-sided alternatives and neither null is rejected. Clearly then it would not make sense to say that '$H_0 : \beta = 0.5$ is accepted' and '$H_0 : \beta = 1$ is accepted', since the true (but unknown) value of β cannot be both 0.5 and 1. So, to summarise, the null hypothesis is either rejected or not rejected on the basis of the available evidence.

2.9.5 The confidence interval approach to hypothesis testing (box 2.6)

To give an example of its usage, one might estimate a parameter, say $\hat{\beta}$, to be 0.93, and a '95% confidence interval' to be (0.77, 1.09). This means that in many repeated samples, 95% of the time, the true value of β will be contained within this interval. Confidence intervals are almost invariably estimated in a two-sided form, although in theory a one-sided interval can be constructed. Constructing a 95% confidence interval is equivalent to using the 5% level in a test of significance.

2.9.6 The test of significance and confidence interval approaches always give the same conclusion

Under the test of significance approach, the null hypothesis that $\beta = \beta^*$ will not be rejected if the test statistic lies within the non-rejection region, i.e. if the following condition holds

$$-t_{crit} \leq \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \leq +t_{crit}$$

Box 2.6   Carrying out a hypothesis test using confidence intervals

(1) Calculate $\hat{\alpha}$, $\hat{\beta}$ and $SE(\hat{\alpha})$, $SE(\hat{\beta})$ as before.
(2) Choose a significance level, α (again the convention is 5%). This is equivalent to choosing a $(1 - \alpha)*100\%$ confidence interval

   i.e. 5% significance level $=$ 95% confidence interval

(3) Use the t-tables to find the appropriate critical value, which will again have $T-2$ degrees of freedom.
(4) The confidence interval for β is given by

   $$(\hat{\beta} - t_{crit} \cdot SE(\hat{\beta}), \hat{\beta} + t_{crit} \cdot SE(\hat{\beta}))$$

   Note that a centre dot ($\cdot$) is sometimes used instead of a cross ($\times$) to denote when two quantities are multiplied together.
(5) Perform the test: if the hypothesised value of β (i.e. β^*) lies outside the confidence interval, then reject the null hypothesis that $\beta = \beta^*$, otherwise do not reject the null.

Rearranging, the null hypothesis would not be rejected if

$$-t_{crit} \cdot SE(\hat{\beta}) \le \hat{\beta} - \beta^* \le + t_{crit} \cdot SE(\hat{\beta})$$

i.e. one would not reject if

$$\hat{\beta} - t_{crit} \cdot SE(\hat{\beta}) \le \beta^* \le \hat{\beta} + t_{crit} \cdot SE(\hat{\beta})$$

But this is just the rule for non-rejection under the confidence interval approach. So it will always be the case that, for a given significance level, the test of significance and confidence interval approaches will provide the same conclusion by construction. One testing approach is simply an algebraic rearrangement of the other.

Example 2.4

Given the regression results above

$$\hat{y}_t = 20.3 + 0.5091x_t \atop (14.38)\,(0.2561) \quad , \qquad T = 22 \qquad\qquad (2.31)$$

Using both the test of significance and confidence interval approaches, test the hypothesis that $\beta = 1$ against a two-sided alternative. This hypothesis might be of interest, for a unit coefficient on the explanatory variable implies a 1:1 relationship between movements in x and movements in y.

The null and alternative hypotheses are respectively:

$H_0 : \beta = 1$

$H_1 : \beta \ne 1$

Box 2.7 The test of significance and confidence interval approaches compared

Test of significance approach	*Confidence interval approach*

$$\text{test stat} = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$

Find $t_{crit} = t_{20;5\%} = \pm 2.086$

$$= \frac{0.5091 - 1}{0.2561} = -1.917$$

$$\hat{\beta} \pm t_{crit} \cdot SE(\hat{\beta})$$

Find $t_{crit} = t_{20;5\%} = \pm 2.086$

$$= 0.5091 \pm 2.086 \cdot 0.2561$$
$$= (-0.0251, 1.0433)$$

Do not reject H_0 since test statistic lies within non-rejection region

Do not reject H_0 since 1 lies within the confidence interval

The results of the test according to each approach are shown in box 2.7.

A couple of comments are in order. First, the critical value from the t-distribution that is required is for 20 degrees of freedom and at the 5% level. This means that 5% of the total distribution will be in the rejection region, and since this is a two-sided test, 2.5% of the distribution is required to be contained in each tail. From the symmetry of the t-distribution around zero, the critical values in the upper and lower tail will be equal in magnitude, but opposite in sign, as shown in figure 2.16.

What if instead the researcher wanted to test $H_0 : \beta = 0$ or $H_0 : \beta = 2$? In order to test these hypotheses using the test of significance approach, the test statistic would have to be reconstructed in each case, although the critical value would be the same. On the other hand, no additional work would be required if the confidence interval approach had been adopted,

Figure 2.16

Critical values and rejection regions for a $t_{20;5\%}$
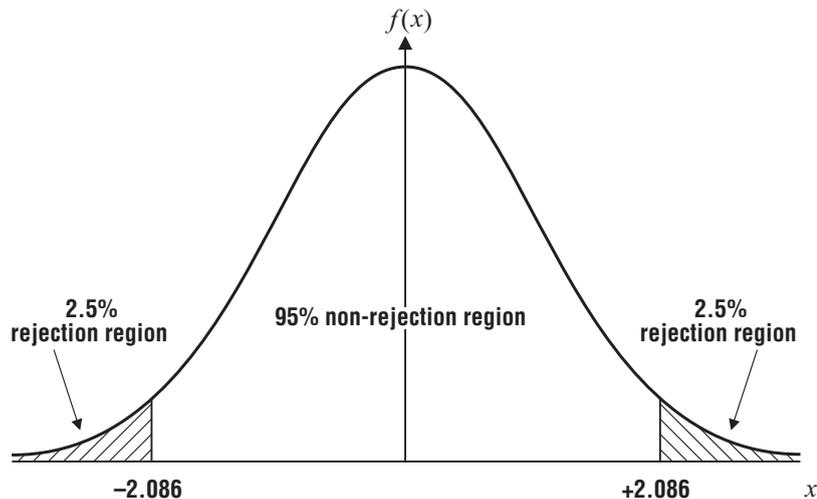
since it effectively permits the testing of an infinite number of hypotheses. So for example, suppose that the researcher wanted to test

$H_0 : \beta = 0$

versus

$H_1 : \beta \neq 0$

and

$H_0 : \beta = 2$

versus

$H_1 : \beta \neq 2$

In the first case, the null hypothesis (that $\beta = 0$) would not be rejected since 0 lies within the 95% confidence interval. By the same argument, the second null hypothesis (that $\beta = 2$) would be rejected since 2 lies outside the estimated confidence interval.

On the other hand, note that this book has so far considered only the results under a 5% size of test. In marginal cases (e.g. $H_0 : \beta = 1$, where the test statistic and critical value are close together), a completely different answer may arise if a different size of test was used. This is where the test of significance approach is preferable to the construction of a confidence interval.

For example, suppose that now a 10% size of test is used for the null hypothesis given in example 2.4. Using the test of significance approach,

$$\textit{test statistic} = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$
$$= \frac{0.5091 - 1}{0.2561} = -1.917$$

as above. The only thing that changes is the critical t-value. At the 10% level (so that 5% of the total distribution is placed in each of the tails for this two-sided test), the required critical value is $t_{20;10\%} = \pm 1.725$. So now, as the test statistic lies in the rejection region, H_0 would be rejected. In order to use a 10% test under the confidence interval approach, the interval itself would have to have been re-estimated since the critical value is embedded in the calculation of the confidence interval.

So the test of significance and confidence interval approaches both have their relative merits. The testing of a number of different hypotheses is easier under the confidence interval approach, while a consideration of

the effect of the size of the test on the conclusion is easier to address under the test of significance approach.

Caution should therefore be used when placing emphasis on or making decisions in the context of marginal cases (i.e. in cases where the null is only just rejected or not rejected). In this situation, the appropriate conclusion to draw is that the results are marginal and that no strong inference can be made one way or the other. A thorough empirical analysis should involve conducting a sensitivity analysis on the results to determine whether using a different size of test alters the conclusions. It is worth stating again that it is conventional to consider sizes of test of 10%, 5% and 1%. If the conclusion (i.e. 'reject' or 'do not reject') is robust to changes in the size of the test, then one can be more confident that the conclusions are appropriate. If the outcome of the test is qualitatively altered when the size of the test is modified, the conclusion must be that there is no conclusion one way or the other!

It is also worth noting that if a given null hypothesis is rejected using a 1% significance level, it will also automatically be rejected at the 5% level, so that there is no need to actually state the latter. Dougherty (1992, p. 100), gives the analogy of a high jumper. If the high jumper can clear 2 metres, it is obvious that the jumper could also clear 1.5 metres. The 1% significance level is a higher hurdle than the 5% significance level. Similarly, if the null is not rejected at the 5% level of significance, it will automatically not be rejected at any stronger level of significance (e.g. 1%). In this case, if the jumper cannot clear 1.5 metres, there is no way s/he will be able to clear 2 metres.

2.9.7 Some more terminology

If the null hypothesis is rejected at the 5% level, it would be said that the result of the test is 'statistically significant'. If the null hypothesis is not rejected, it would be said that the result of the test is 'not significant', or that it is 'insignificant'. Finally, if the null hypothesis is rejected at the 1% level, the result is termed 'highly statistically significant'.

Note that a statistically significant result may be of no practical significance. For example, if the estimated beta for a stock under a CAPM regression is 1.05, and a null hypothesis that $\beta = 1$ is rejected, the result will be statistically significant. But it may be the case that a slightly higher beta will make no difference to an investor's choice as to whether to buy the stock or not. In that case, one would say that the result of the test was statistically significant but financially or practically insignificant.

Table 2.3 Classifying hypothesis testing errors and correct conclusions

		Reality	
		H_0 is true	H_0 is false
Result of test	Significant (reject H_0)	Type I error $= \alpha$	✓
	Insignificant (do not reject H_0)	✓	Type II error $= \beta$

2.9.8 Classifying the errors that can be made using hypothesis tests

H_0 is usually rejected if the test statistic is statistically significant at a chosen significance level. There are two possible errors that could be made:

(1) Rejecting H_0 when it was really true; this is called a *type I error*.
(2) Not rejecting H_0 when it was in fact false; this is called a *type II error*.

The possible scenarios can be summarised in table 2.3.

The probability of a type I error is just α, the significance level or size of test chosen. To see this, recall what is meant by 'significance' at the 5% level: it is only 5% likely that a result as or more extreme as this could have occurred purely by chance. Or, to put this another way, it is only 5% likely that this null would be rejected when it was in fact true.

Note that there is no chance for a free lunch (i.e. a cost-less gain) here! What happens if the size of the test is reduced (e.g. from a 5% test to a 1% test)? The chances of making a type I error would be reduced . . . but so would the probability that the null hypothesis would be rejected at all, so increasing the probability of a type II error. The two competing effects of reducing the size of the test can be shown in box 2.8.

So there always exists, therefore, a direct trade-off between type I and type II errors when choosing a significance level. The only way to

Box 2.8 Type I and Type II errors

Reduce size→More strict →Reject null↗reject Less likely to falsely →Lower chance of type I error
of test (e.g. criterion for hypothesis↘ More likely to incorrectly →Higher chance of type II error
5% to 1%) rejection less often not reject

reduce the chances of both is to increase the sample size or to select a sample with more variation, thus increasing the amount of information upon which the results of the hypothesis test are based. In practice, up to a certain level, type I errors are usually considered more serious and hence a small size of test is usually chosen (5% or 1% are the most common).

The probability of a type I error is the probability of incorrectly rejecting a correct null hypothesis, which is also the size of the test. Another important piece of terminology in this area is the *power of a test*. The power of a test is defined as the probability of (appropriately) rejecting an incorrect null hypothesis. The power of the test is also equal to one minus the probability of a type II error.

An optimal test would be one with an actual test size that matched the nominal size and which had as high a power as possible. Such a test would imply, for example, that using a 5% significance level would result in the null being rejected exactly 5% of the time by chance alone, and that an incorrect null hypothesis would be rejected close to 100% of the time.

2.10 A special type of hypothesis test: the *t*-ratio

Recall that the formula under a test of significance approach to hypothesis testing using a *t*-test for the slope parameter was

$$\text{test statistic} = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \tag{2.32}$$

with the obvious adjustments to test a hypothesis about the intercept. If the test is

$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$

i.e. a test that the population parameter is zero against a two-sided alternative, this is known as a *t*-ratio test. Since $\beta^* = 0$, the expression in (2.32) collapses to

$$\text{test statistic} = \frac{\hat{\beta}}{SE(\hat{\beta})} \tag{2.33}$$

Thus the ratio of the coefficient to its standard error, given by this expression, is known as the *t-ratio* or *t-statistic*.

Example 2.5

Suppose that we have calculated the estimates for the intercept and the slope (1.10 and -19.88 respectively) and their corresponding standard errors (1.35 and 1.98 respectively). The t-ratios associated with each of the intercept and slope coefficients would be given by

	$\hat{\alpha}$	$\hat{\beta}$
Coefficient	1.10	-19.88
SE	1.35	1.98
t-ratio	0.81	-10.04

Note that if a coefficient is negative, its t-ratio will also be negative. In order to test (separately) the null hypotheses that $\alpha = 0$ and $\beta = 0$, the test statistics would be compared with the appropriate critical value from a t-distribution. In this case, the number of degrees of freedom, given by $T - k$, is equal to $15-3=12$. The 5% critical value for this two-sided test (remember, 2.5% in each tail for a 5% test) is 2.179, while the 1% two-sided critical value (0.5% in each tail) is 3.055. Given these t-ratios and critical values, would the following null hypotheses be rejected?

$H_0 : \alpha = 0?$ *(No)*
$H_0 : \beta = 0?$ *(Yes)*

If H_0 is rejected, it would be said that the test statistic is *significant*. If the variable is not 'significant', it means that while the estimated value of the coefficient is not exactly zero (e.g. 1.10 in the example above), the coefficient is indistinguishable statistically from zero. If a zero were placed in the fitted equation instead of the estimated value, this would mean that whatever happened to the value of that explanatory variable, the dependent variable would be unaffected. This would then be taken to mean that the variable is not helping to explain variations in y, and that it could therefore be removed from the regression equation. For example, if the t-ratio associated with x had been -1.04 rather than -10.04 (assuming that the standard error stayed the same), the variable would be classed as insignificant (i.e. not statistically different from zero). The only insignificant term in the above regression is the intercept. There are good statistical reasons for always retaining the constant, even if it is not significant; see chapter 4.

It is worth noting that, for degrees of freedom greater than around 25, the 5% two-sided critical value is approximately ± 2. So, as a rule of thumb (i.e. a rough guide), the null hypothesis would be rejected if the t-statistic exceeds 2 in absolute value.

Some authors place the *t*-ratios in parentheses below the corresponding coefficient estimates rather than the standard errors. One thus needs to check which convention is being used in each particular application, and also to state this clearly when presenting estimation results.

There will now follow two finance case studies that involve only the estimation of bivariate linear regression models and the construction and interpretation of *t*-ratios.

2.11 An example of the use of a simple *t*-test to test a theory in finance: can US mutual funds beat the market?

Jensen (1968) was the first to systematically test the performance of mutual funds, and in particular examine whether any 'beat the market'. He used a sample of annual returns on the portfolios of 115 mutual funds from 1945–64. Each of the 115 funds was subjected to a separate OLS time series regression of the form

$$R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{mt} - R_{ft}) + u_{jt} \tag{2.52}$$

where R_{jt} is the return on portfolio j at time t, R_{ft} is the return on a risk-free proxy (a 1-year government bond), R_{mt} is the return on a market portfolio proxy, u_{jt} is an error term, and α_j, β_j are parameters to be estimated. The quantity of interest is the significance of α_j, since this parameter defines whether the fund outperforms or underperforms the market index. Thus the null hypothesis is given by: $H_0 : \alpha_j = 0$. A positive and significant α_j for a given fund would suggest that the fund is able to earn significant abnormal returns in excess of the market-required return for a fund of this given riskiness. This coefficient has become known as 'Jensen's alpha'. Some summary statistics across the 115 funds for the estimated regression results for (2.52) are given in table 2.4.

Table 2.4 Summary statistics for the estimated regression results for (2.52)

			Extremal values	
Item	Mean value	Median value	Minimum	Maximum
$\hat{\alpha}$	−0.011	−0.009	−0.080	0.058
$\hat{\beta}$	0.840	0.848	0.219	1.405
Sample size	17	19	10	20

Source: Jensen (1968). Reprinted with the permission of Blackwell Publishers.

Figure 2.17
Frequency
distribution of
t-ratios of mutual
fund alphas (gross
of transactions
costs) Source:
Jensen (1968).
Reprinted with the
permission of
Blackwell Publishers
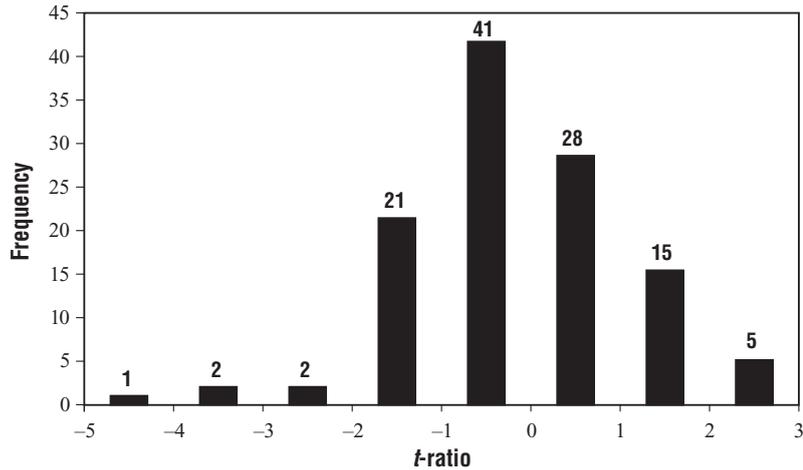
Figure 2.18
Frequency
distribution of
t-ratios of mutual
fund alphas (net of
transactions costs)
Source: Jensen
(1968). Reprinted
with the permission
of Blackwell
Publishers
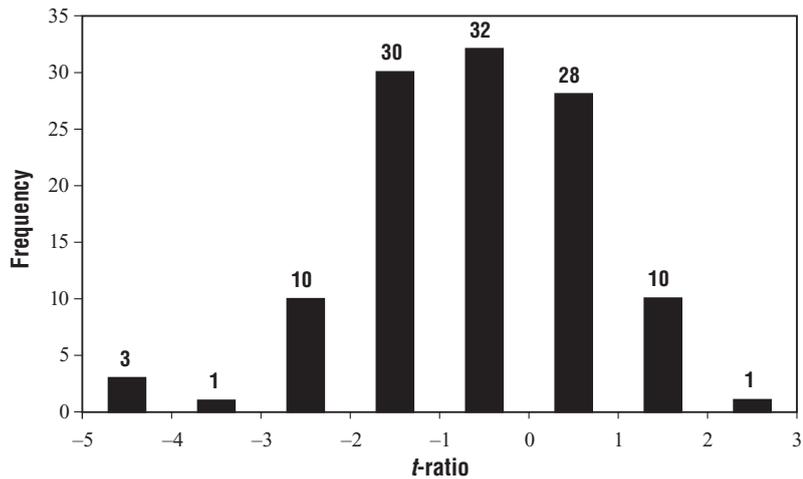
As table 2.4 shows, the average (defined as either the mean or the me-
dian) fund was unable to 'beat the market', recording a negative alpha
in both cases. There were, however, some funds that did manage to per-
form significantly better than expected given their level of risk, with the
best fund of all yielding an alpha of 0.058. Interestingly, the average fund
had a beta estimate of around 0.85, indicating that, in the CAPM context,
most funds were less risky than the market index. This result may be
attributable to the funds investing predominantly in (mature) blue chip
stocks rather than small caps.

The most visual method of presenting the results was obtained by plot-
ting the number of mutual funds in each *t*-ratio category for the alpha
coefficient, first gross and then net of transactions costs, as in figure 2.17
and figure 2.18, respectively.

Table 2.5 Summary statistics for unit trust returns, January 1979–May 2000

	Mean (%)	Minimum (%)	Maximum (%)	Median (%)
Average monthly return, 1979–2000	1.0	0.6	1.4	1.0
Standard deviation of returns over time	5.1	4.3	6.9	5.0

The appropriate critical value for a two-sided test of $\alpha_j = 0$ is approximately 2.10 (assuming 20 years of annual data leading to 18 degrees of freedom). As can be seen, only five funds have estimated t-ratios greater than 2 and are therefore implied to have been able to outperform the market before transactions costs are taken into account. Interestingly, five firms have also significantly underperformed the market, with t-ratios of –2 or less.

When transactions costs are taken into account (figure 2.18), only one fund out of 115 is able to significantly outperform the market, while 14 significantly underperform it. Given that a nominal 5% two-sided size of test is being used, one would expect two or three funds to 'significantly beat the market' by chance alone. It would thus be concluded that, during the sample period studied, US fund managers appeared unable to systematically generate positive abnormal returns.

2.12 Can UK unit trust managers beat the market?

Jensen's study has proved pivotal in suggesting a method for conducting empirical tests of the performance of fund managers. However, it has been criticised on several grounds. One of the most important of these in the context of this book is that only between 10 and 20 annual observations were used for each regression. Such a small number of observations is really insufficient for the asymptotic theory underlying the testing procedure to be validly invoked.

A variant on Jensen's test is now estimated in the context of the UK market, by considering monthly returns on 76 equity unit trusts. The data cover the period January 1979–May 2000 (257 observations for each fund). Some summary statistics for the funds are presented in table 2.5.

From these summary statistics, the average continuously compounded return is 1.0% per month, although the most interesting feature is the

Table 2.6 CAPM regression results for unit trust returns, January 1979–May 2000

Estimates of	Mean	Minimum	Maximum	Median
$\alpha(\%)$	−0.02	−0.54	0.33	−0.03
β	0.91	0.56	1.09	0.91
t-ratio on α	−0.07	−2.44	3.11	−0.25

Figure 2.19

Performance of UK unit trusts, 1979–2000
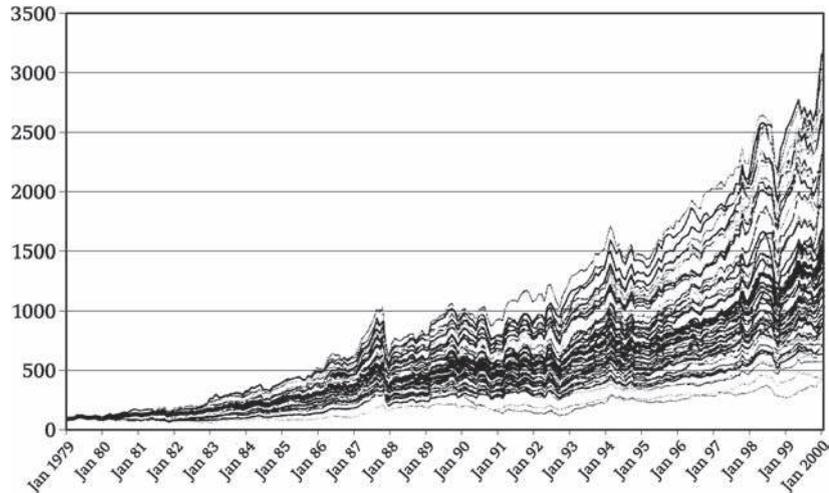
wide variation in the performances of the funds. The worst-performing fund yields an average return of 0.6% per month over the 20-year period, while the best would give 1.4% per month. This variability is further demonstrated in figure 2.19, which plots over time the value of £100 invested in each of the funds in January 1979.

A regression of the form (2.52) is applied to the UK data, and the summary results presented in table 2.6. A number of features of the regression results are worthy of further comment. First, most of the funds have estimated betas less than one again, perhaps suggesting that the fund managers have historically been risk-averse or investing disproportionately in blue chip companies in mature sectors. Second, gross of transactions costs, nine funds of the sample of 76 were able to significantly outperform the market by providing a significant positive alpha, while seven funds yielded significant negative alphas. The average fund (where 'average' is measured using either the mean or the median) is not able to earn any excess return over the required rate given its level of risk.

Box 2.9 Reasons for stock market overreactions

(1) *That the 'overreaction effect' is just another manifestation of the 'size effect'*. The size effect is the tendency of small firms to generate on average, superior returns to large firms. The argument would follow that the losers were small firms and that these small firms would subsequently outperform the large firms. DeBondt and Thaler did not believe this a sufficient explanation, but Zarowin (1990) found that allowing for firm size did reduce the subsequent return on the losers.

(2) *That the reversals of fortune reflect changes in equilibrium required returns*. The losers are argued to be likely to have considerably higher CAPM betas, reflecting investors' perceptions that they are more risky. Of course, betas can change over time, and a substantial fall in the firms' share prices (for the losers) would lead to a rise in their leverage ratios, leading in all likelihood to an increase in their perceived riskiness. Therefore, the required rate of return on the losers will be larger, and their *ex post* performance better. Ball and Kothari (1989) find the CAPM betas of losers to be considerably higher than those of winners.

2.13 The overreaction hypothesis and the UK stock market

2.13.1 Motivation

Two studies by DeBondt and Thaler (1985, 1987) showed that stocks experiencing a poor performance over a 3–5-year period subsequently tend to outperform stocks that had previously performed relatively well. This implies that, on average, stocks which are 'losers' in terms of their returns subsequently become 'winners', and vice versa. This chapter now examines a paper by Clare and Thomas (1995) that conducts a similar study using monthly UK stock returns from January 1955 to 1990 (36 years) on all firms traded on the London Stock exchange.

This phenomenon seems at first blush to be inconsistent with the efficient markets hypothesis, and Clare and Thomas propose two explanations (box 2.9).

Zarowin (1990) also finds that 80% of the extra return available from holding the losers accrues to investors in January, so that almost all of the 'overreaction effect' seems to occur at the start of the calendar year.

2.13.2 Methodology

Clare and Thomas take a random sample of 1,000 firms and, for each, they calculate the monthly excess return of the stock for the market over a 12-, 24- or 36-month period for each stock i

$$U_{it} = R_{it} - R_{mt} \quad t = 1, \ldots, n; \quad i = 1, \ldots, 1000;$$

$$n = 12, 24 \text{ or } 36 \tag{2.53}$$

Box 2.10   Ranking stocks and forming portfolios

Portfolio	Ranking
Portfolio 1	Best performing 20% of firms
Portfolio 2	Next 20%
Portfolio 3	Next 20%
Portfolio 4	Next 20%
Portfolio 5	Worst performing 20% of firms

Box 2.11   Portfolio monitoring

Estimate $\bar{R}_i$ for year 1
Monitor portfolios for year 2
Estimate $\bar{R}_i$ for year 3

$\vdots$

Monitor portfolios for year 36

Then the average monthly return over each stock i for the first 12-, 24-, or 36-month period is calculated:

$$\bar{R}_i = \frac{1}{n} \sum_{t=1}^{n} U_{it} \tag{2.54}$$

The stocks are then ranked from highest average return to lowest and from these 5 portfolios are formed and returns are calculated assuming an equal weighting of stocks in each portfolio (box 2.10).

The same sample length n is used to monitor the performance of each portfolio. Thus, for example, if the portfolio formation period is one, two or three years, the subsequent portfolio tracking period will also be one, two or three years, respectively. Then another portfolio formation period follows and so on until the sample period has been exhausted. How many samples of length n will there be? $n = 1$, 2, or 3 years. First, suppose $n = 1$ year. The procedure adopted would be as shown in box 2.11.

So if $n = 1$, there are 18 independent (non-overlapping) observation periods and 18 independent tracking periods. By similar arguments, $n = 2$ gives 9 independent periods and $n = 3$ gives 6 independent periods. The mean return for each month over the 18, 9, or 6 periods for the winner and loser portfolios (the top 20% and bottom 20% of firms in the portfolio formation period) are denoted by $\bar{R}_{pt}^{W}$ and $\bar{R}_{pt}^{L}$, respectively. Define the difference between these as $\bar{R}_{Dt} = \bar{R}_{pt}^{L} - \bar{R}_{pt}^{W}$.

Table 2.7 Is there an overreaction effect in the UK stock market?

Panel A: All Months			
	$n = 12$	$n = 24$	$n = 36$
Return on loser	0.0033	0.0011	0.0129
Return on winner	0.0036	−0.0003	0.0115
Implied annualised return difference	−0.37%	1.68%	1.56%
Coefficient for (2.55): $\hat{\alpha}_1$	−0.00031	0.0014**	0.0013
	(0.29)	(2.01)	(1.55)
Coefficients for (2.56): $\hat{\alpha}_2$	−0.00034	0.00147**	0.0013*
	(−0.30)	(2.01)	(1.41)
Coefficients for (2.56): $\hat{\beta}$	−0.022	0.010	−0.0025
	(−0.25)	(0.21)	(−0.06)
Panel B: all months except January			
Coefficient for (2.55): $\hat{\alpha}_1$	−0.0007	0.0012*	0.0009
	(−0.72)	(1.63)	(1.05)

Notes: *t*-ratios in parentheses; * and ** denote significance at the 10% and 5% levels, respectively.
Source: Clare and Thomas (1995). Reprinted with the permission of Blackwell Publishers.

The first regression to be performed is of the excess return of the losers over the winners on a constant only

$$\bar{R}_{Dt} = \alpha_1 + \eta_t \tag{2.55}$$

where η_t is an error term. The test is of whether α_1 is significant and positive. However, a significant and positive α_1 is not a sufficient condition for the overreaction effect to be confirmed because it could be owing to higher returns being required on loser stocks owing to loser stocks being more risky. The solution, Clare and Thomas (1995) argue, is to allow for risk differences by regressing against the market risk premium

$$\bar{R}_{Dt} = \alpha_2 + \beta(R_{mt} - R_{ft}) + \eta_t \tag{2.56}$$

where R_{mt} is the return on the FTA All-share, and R_{ft} is the return on a UK government three-month Treasury Bill. The results for each of these two regressions are presented in table 2.7.

As can be seen by comparing the returns on the winners and losers in the first two rows of table 2.7, 12 months is not a sufficiently long time for losers to become winners. By the two-year tracking horizon, however, the losers have become winners, and similarly for the three-year samples. This translates into an average 1.68% higher return on the losers than the

winners at the two-year horizon, and 1.56% higher return at the three-year horizon. Recall that the estimated value of the coefficient in a regression of a variable on a constant only is equal to the average value of that variable. It can also be seen that the estimated coefficients on the constant terms for each horizon are exactly equal to the differences between the returns of the losers and the winners. This coefficient is statistically significant at the two-year horizon, and marginally significant at the three-year horizon.

In the second test regression, $\hat{\beta}$ represents the difference between the market betas of the winner and loser portfolios. None of the beta coefficient estimates are even close to being significant, and the inclusion of the risk term makes virtually no difference to the coefficient values or significances of the intercept terms.

Removal of the January returns from the samples reduces the subsequent degree of overperformance of the loser portfolios, and the significances of the $\hat{\alpha}_1$ terms is somewhat reduced. It is concluded, therefore, that only a part of the overreaction phenomenon occurs in January. Clare and Thomas then proceed to examine whether the overreaction effect is related to firm size, although the results are not presented here.

2.13.3 Conclusions

The main conclusions from Clare and Thomas' study are:

(1) There appears to be evidence of overreactions in UK stock returns, as found in previous US studies.
(2) These over-reactions are unrelated to the CAPM beta.
(3) Losers that subsequently become winners tend to be small, so that most of the overreaction in the UK can be attributed to the size effect.

2.14 The exact significance level

The exact significance level is also commonly known as the p-value. It gives the *marginal significance level* where one would be indifferent between rejecting and not rejecting the null hypothesis. If the test statistic is 'large' in absolute value, the p-value will be small, and vice versa. For example, consider a test statistic that is distributed as a t_{62} and takes a value of 1.47. Would the null hypothesis be rejected? It would depend on the size of the test. Now, suppose that the p-value for this test is calculated to be 0.12:

- Is the null rejected at the 5% level?   *No*
- Is the null rejected at the 10% level?   *No*
- Is the null rejected at the 20% level?   *Yes*

Table 2.8 Part of the EViews regression output revisited

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.363302	0.444369	0.817569	0.4167
RFUTURES	0.123860	0.133790	0.925781	0.3581

In fact, the null would have been rejected at the 12% level or higher. To see this, consider conducting a series of tests with size 0.1%, 0.2%, 0.3%, 0.4%, ... 1%, ..., 5%, ... 10%, ... Eventually, the critical value and test statistic will meet and this will be the *p*-value. *p*-values are almost always provided automatically by software packages. Note how useful they are! They provide all of the information required to conduct a hypothesis test without requiring of the researcher the need to calculate a test statistic or to find a critical value from a table – both of these steps have already been taken by the package in producing the *p*-value. The *p*-value is also useful since it avoids the requirement of specifying an arbitrary significance level (α). Sensitivity analysis of the effect of the significance level on the conclusion occurs automatically.

Informally, the *p*-value is also often referred to as the probability of being wrong when the null hypothesis is rejected. Thus, for example, if a *p*-value of 0.05 or less leads the researcher to reject the null (equivalent to a 5% significance level), this is equivalent to saying that if the probability of incorrectly rejecting the null is more than 5%, do not reject it. The *p*-value has also been termed the 'plausibility' of the null hypothesis; so, the smaller is the *p*-value, the less plausible is the null hypothesis.

2.15 Hypothesis testing in EViews – example 1: hedging revisited

Reload the 'hedge.wf1' EViews work file that was created above. If we re-examine the results table from the returns regression (screenshot 2.3 on p. 43), it can be seen that as well as the parameter estimates, EViews automatically calculates the standard errors, the *t*-ratios, and the *p*-values associated with a two-sided test of the null hypothesis that the true value of a parameter is zero. Part of the results table is replicated again here (table 2.8) for ease of interpretation.

The third column presents the *t*-ratios, which are the test statistics for testing the null hypothesis that the true values of these parameters are zero against a two sided alternative – i.e. these statistics test $H_0 : \alpha = 0$ versus $H_1 : \alpha \neq 0$ in the first row of numbers and $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$

in the second. The fact that these test statistics are both very small is indicative that neither of these null hypotheses is likely to be rejected. This conclusion is confirmed by the *p*-values given in the final column. Both *p*-values are considerably larger than 0.1, indicating that the corresponding test statistics are not even significant at the 10% level.

Suppose now that we wanted to test the null hypothesis that $H_0 : \beta = 1$ rather than $H_0 : \beta = 0$. We could test this, or any other hypothesis about the coefficients, by hand, using the information we already have. But it is easier to let EViews do the work by typing **View** and then **Coefficient Tests/Wald – Coefficient Restrictions** .... EViews defines all of the parameters in a vector C, so that C(1) will be the intercept and C(2) will be the slope. Type **C(2)=1** and click **OK**. Note that using this software, it is possible to test multiple hypotheses, which will be discussed in chapter 3, and also non-linear restrictions, which cannot be tested using the standard procedure for inference described above.

Wald Test:
Equation: LEVELREG

Test Statistic	Value	df	Probability
F-statistic	0.565298	(1, 64)	0.4549
Chi-square	0.565298	1	0.4521

Null Hypothesis Summary:

Normalised Restriction (= 0)	Value	Std. Err.
$-1 + C(2)$	-0.017777	0.023644

Restrictions are linear in coefficients.

The test is performed in two different ways, but results suggest that the null hypothesis should clearly be rejected as the *p*-value for the test is zero to four decimal places. Since we are testing a hypothesis about only one parameter, the two test statistics ('*F*-statistic' and 'χ-square') will always be identical. These are equivalent to conducting a *t*-test, and these alternative formulations will be discussed in detail in chapter 4. EViews also reports the 'normalised restriction', although this can be ignored for the time being since it merely reports the regression slope parameter (in a different form) and its standard error.

Now go back to the regression in levels (i.e. with the raw prices rather than the returns) and test the null hypothesis that $\beta = 1$ in this regression. You should find in this case that the null hypothesis is not rejected (table below).

Wald Test:
Equation: RETURNREG

Test Statistic	Value	df	Probability
F-statistic	42.88455	(1, 63)	0.0000
Chi-square	42.88455	1	0.0000

Null Hypothesis Summary:

Normalised Restriction (= 0)	Value	Std. Err.
$-1 + C(2)$	-0.876140	0.133790

Restrictions are linear in coefficients.

2.16 Estimation and hypothesis testing in EViews – example 2: the CAPM

This exercise will estimate and test some hypotheses about the CAPM beta for several US stocks. First, **Open a new workfile** to accommodate monthly data commencing in January 2002 and ending in April 2007. Then **import the Excel file 'capm.xls'**. The file is organised by observation and contains six columns of numbers plus the dates in the first column, so in the 'Names for series or Number if named in file' box, type **6**. As before, do not import the dates so the data start in cell B2. The monthly stock prices of four companies (Ford, General Motors, Microsoft and Sun) will appear as objects, along with index values for the S&P500 ('sandp') and three-month US-Treasury bills ('ustb3m'). **Save the EViews workfile as 'capm.wk1'**.

In order to estimate a CAPM equation for the Ford stock, for example, we need to first transform the price series into returns and then the excess returns over the risk free rate. To transform the series, click on the Generate button (**Genr**) in the workfile window. In the new window, type

RSANDP=100*LOG(SANDP/SANDP(−1))

This will create a new series named RSANDP that will contain the returns of the S&P500. The operator (−1) is used to instruct EViews to use the one-period lagged observation of the series. To estimate percentage returns on the Ford stock, press the **Genr** button again and type

RFORD=100*LOG(FORD/FORD(−1))

This will yield a new series named RFORD that will contain the returns of the Ford stock. EViews allows various kinds of transformations to the

series. For example

X2=X/2	creates a new variable called X2 that is half of X
XSQ=Xˆ2	creates a new variable XSQ that is X squared
LX=LOG(X)	creates a new variable LX that is the log of X
LAGX=X(−1)	creates a new variable LAGX containing X lagged by one period
LAGX2=X(−2)	creates a new variable LAGX2 containing X lagged by two periods

Other functions include:

d(X)	first difference of X
d(X,n)	nth order difference of X
dlog(X)	first difference of the logarithm of X
dlog(X,n)	nth order difference of the logarithm of X
abs(X)	absolute value of X

If, in the transformation, the new series is given the same name as the old series, then the old series will be overwritten. Note that the returns for the S&P index could have been constructed using a simpler command in the 'Genr' window such as

RSANDP=100*DLOG(SANDP)

as we used in chapter 1. Before we can transform the returns into excess returns, we need to be slightly careful because the stock returns are monthly, but the Treasury bill yields are annualised. We could run the whole analysis using monthly data or using annualised data and it should not matter which we use, but the two series must be measured consistently. So, to turn the T-bill yields into monthly figures and to write over the original series, press the **Genr** button again and type

USTB3M=USTB3M/12

Now, to compute the excess returns, click **Genr** again and type

ERSANDP=RSANDP-USTB3M

where 'ERSANDP' will be used to denote the excess returns, so that the original raw returns series will remain in the workfile. The Ford returns can similarly be transformed into a set of excess returns.

Now that the excess returns have been obtained for the two series, before running the regression, plot the data to examine visually whether

the series appear to move together. To do this, create a new object by clicking on the **Object/New Object** menu on the menu bar. Select **Graph**, provide a name (call the graph **Graph1**) and then in the new window provide the names of the series to plot. In this new window, type

ERSANDP ERFORD

Then press **OK** and screenshot 2.4 will appear.
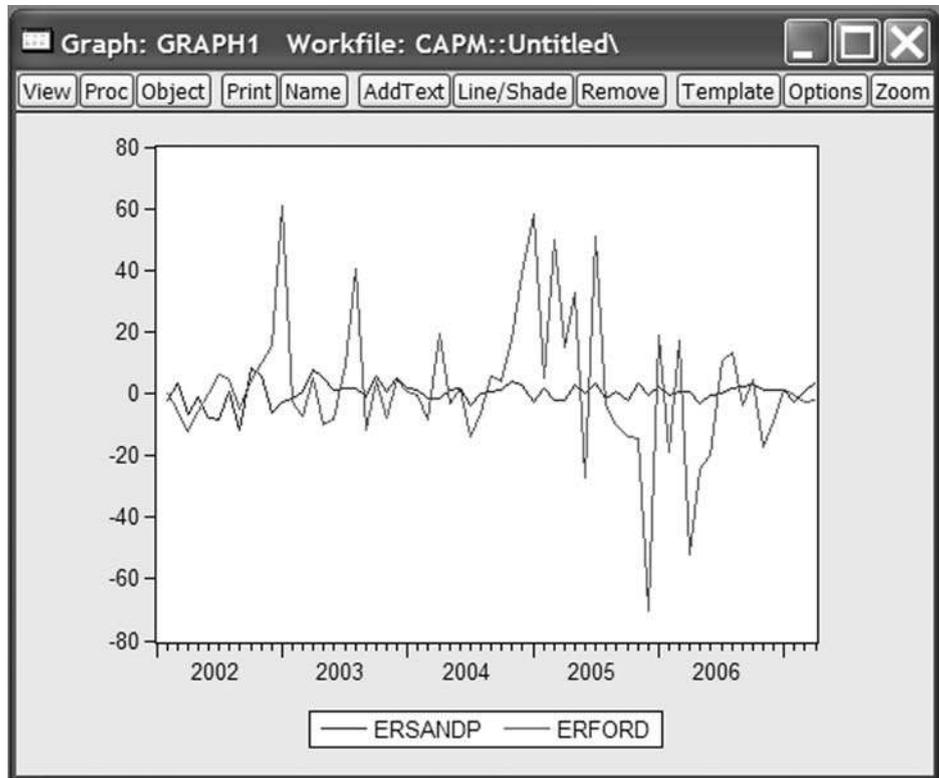
This is a time-series plot of the two variables, but a scatter plot may be more informative. To examine a scatter plot, Click **Options**, choose the **Type** tab, then select **Scatter** from the list and click **OK**. There appears to be a weak association between ERFTAS and ERFORD. Close the window of the graph and return to the workfile window.

To estimate the CAPM equation, click on **Object/New Objects**. In the new window, select **Equation** and name the object **CAPM**. Click on **OK**. In the window, specify the regression equation. The regression equation takes the form

$$(R_{Ford} - r_f)_t = \alpha + \beta(R_M - r_f)_t + u_t$$

Since the data have already been transformed to obtain the excess returns, in order to specify this regression equation, type in the equation window

ERFORD C ERSANDP

To use all the observations in the sample and to estimate the regression using LS – Least Squares (NLS and ARMA), click on **OK**. The results screen appears as in the following table. Make sure that you **save the Workfile** again to include the transformed series and regression results!

Dependent Variable: ERFORD
Method: Least Squares
Date: 08/21/07 Time: 15:02
Sample (adjusted): 2002M02 2007M04
Included observations: 63 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	2.020219	2.801382	0.721151	0.4736
ERSANDP	0.359726	0.794443	0.452803	0.6523
R-squared	0.003350	Mean dependent var		2.097445
Adjusted R-squared	−0.012989	S.D. dependent var		22.05129
S.E. of regression	22.19404	Akaike info criterion		9.068756
Sum squared resid	30047.09	Schwarz criterion		9.136792
Log likelihood	−283.6658	Hannan-Quinn criter.		9.095514
F-statistic	0.205031	Durbin-Watson stat		1.785699
Prob(F-statistic)	0.652297			

Take a couple of minutes to examine the results of the regression. What is the slope coefficient estimate and what does it signify? Is this coefficient statistically significant? The beta coefficient (the slope coefficient) estimate is 0.3597. The *p*-value of the *t*-ratio is 0.6523, signifying that the excess return on the market proxy has no significant explanatory power for the variability of the excess returns of Ford stock. What is the interpretation of the intercept estimate? Is it statistically significant?

In fact, there is a considerably quicker method for using transformed variables in regression equations, and that is to write the transformation directly into the equation window. In the CAPM example above, this could be done by typing

DLOG(FORD)-USTB3M C DLOG(SANDP)-USTB3M

into the equation window. As well as being quicker, an advantage of this approach is that the output will show more clearly the regression that has actually been conducted, so that any errors in making the transformations can be seen more clearly.

How could the hypothesis that the value of the population coefficient is equal to 1 be tested? The answer is to click on **View/Coefficient Tests/Wald – Coefficient Restrictions**… and then in the box that appears, **Type C(2)=1**. The conclusion here is that the null hypothesis that the CAPM beta of Ford stock is 1 cannot be rejected and hence the estimated beta of 0.359 is not significantly different from 1.[5]

Key concepts

The key terms to be able to define and explain from this chapter are

- regression model
- population
- linear model
- unbiasedness
- standard error
- null hypothesis
- *t*-distribution
- test statistic
- type I error
- size of a test
- *p*-value
- asymptotic

- disturbance term
- sample
- consistency
- efficiency
- statistical inference
- alternative hypothesis
- confidence interval
- rejection region
- type II error
- power of a test
- data mining

Appendix: Mathematical derivations of CLRM results

2A.1 *Derivation of the OLS coefficient estimator in the bivariate case*

$$L = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}(y_t - \hat{\alpha} - \hat{\beta}x_t)^2 \tag{2A.1}$$

It is necessary to minimise L w.r.t. $\hat{\alpha}$ and $\hat{\beta}$, to find the values of α and β that give the line that is closest to the data. So L is differentiated w.r.t. $\hat{\alpha}$ and $\hat{\beta}$, and the first derivatives are set to zero. The first derivatives are given by

$$\frac{\partial L}{\partial \hat{\alpha}} = -2 \sum_t (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \tag{2A.2}$$

$$\frac{\partial L}{\partial \hat{\beta}} = -2 \sum_t x_t(y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \tag{2A.3}$$

---

[5] Although the value 0.359 may seem a long way from 1, considered purely from an econometric perspective, the sample size is quite small and this has led to a large parameter standard error, which explains the failure to reject both $H_0 : \beta = 0$ and $H_0 : \beta = 1$.

The next step is to rearrange (2A.2) and (2A.3) in order to obtain expressions for $\hat{\alpha}$ and $\hat{\beta}$. From (2A.2)

$$\sum_t (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \tag{2A.4}$$

Expanding the parentheses and recalling that the sum runs from 1 to T so that there will be T terms in $\hat{\alpha}$

$$\sum y_t - T\hat{\alpha} - \hat{\beta}\sum x_t = 0 \tag{2A.5}$$

But $\sum y_t = T\bar{y}$ and $\sum x_t = T\bar{x}$, so it is possible to write (2A.5) as

$$T\bar{y} - T\hat{\alpha} - T\hat{\beta}\bar{x} = 0 \tag{2A.6}$$

or

$$\bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} = 0 \tag{2A.7}$$

From (2A.3)

$$\sum_t x_t(y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \tag{2A.8}$$

From (2A.7)

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{2A.9}$$

Substituting into (2A.8) for $\hat{\alpha}$ from (2A.9)

$$\sum_t x_t(y_t - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_t) = 0 \tag{2A.10}$$

$$\sum_t x_t y_t - \bar{y}\sum x_t + \hat{\beta}\bar{x}\sum x_t - \hat{\beta}\sum x_t^2 = 0 \tag{2A.11}$$

$$\sum_t x_t y_t - T\bar{x}\bar{y} + \hat{\beta}T\bar{x}^2 - \hat{\beta}\sum x_t^2 = 0 \tag{2A.12}$$

Rearranging for $\hat{\beta}$,

$$\hat{\beta}\left(T\bar{x}^2 - \sum x_t^2\right) = T\overline{xy} - \sum x_t y_t \tag{2A.13}$$

Dividing both sides of (2A.13) by $\left(T\bar{x}^2 - \sum x_t^2\right)$ gives

$$\hat{\beta} = \frac{\sum x_t y_t - T\overline{xy}}{\sum x_t^2 - T\bar{x}^2} \quad and \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{2A.14}$$

2A.2 Derivation of the OLS standard error estimators for the intercept and slope in the bivariate case

Recall that the variance of the random variable $\hat{\alpha}$ can be written as

$$\text{var}(\hat{\alpha}) = E(\hat{\alpha} - E(\hat{\alpha}))^2 \tag{2A.15}$$

and since the OLS estimator is unbiased

$$\text{var}(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2 \tag{2A.16}$$

By similar arguments, the variance of the slope estimator can be written as

$$\text{var}(\hat{\beta}) = E(\hat{\beta} - \beta)^2 \tag{2A.17}$$

Working first with (2A.17), replacing $\hat{\beta}$ with the formula for it given by the OLS estimator

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} - \beta\right)^2 \tag{2A.18}$$

Replacing y_t with $\alpha + \beta x_t + u_t$, and replacing $\bar{y}$ with $\alpha + \beta\bar{x}$ in (2A.18)

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum (x_t - \bar{x})(\alpha + \beta x_t + u_t - \alpha - \beta\bar{x})}{\sum (x_t - \bar{x})^2} - \beta\right)^2 \tag{2A.19}$$

Cancelling α and multiplying the last β term in (2A.19) by $\dfrac{\sum (x_t - \bar{x})^2}{\sum (x_t - \bar{x})^2}$

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum (x_t - \bar{x})(\beta x_t + u_t - \beta\bar{x}) - \beta \sum (x_t - \bar{x})^2}{\sum (x_t - \bar{x})^2}\right)^2 \tag{2A.20}$$

Rearranging

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum (x_t - \bar{x})\beta(x_t - \bar{x}) + \sum u_t(x_t - \bar{x}) - \beta \sum (x_t - \bar{x})^2}{\sum (x_t - \bar{x})^2}\right)^2 \tag{2A.21}$$

$$\text{var}(\hat{\beta}) = E\left(\frac{\beta \sum (x_t - \bar{x})^2 + \sum u_t(x_t - \bar{x}) - \beta \sum (x_t - \bar{x})^2}{\sum (x_t - \bar{x})^2}\right)^2 \tag{2A.22}$$

Now the β terms in (2A.22) will cancel to give

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum u_t(x_t - \bar{x})}{\sum (x_t - \bar{x})^2}\right)^2 \tag{2A.23}$$

Now let x_t^* denote the mean-adjusted observation for x_t, i.e. $(x_t - \bar{x})$. Equation (2A.23) can be written

$$\text{var}(\hat{\beta}) = E\left(\frac{\sum u_t x_t^*}{\sum x_t^{*2}}\right)^2 \tag{2A.24}$$

The denominator of (2A.24) can be taken through the expectations operator under the assumption that x is fixed or non-stochastic

$$\text{var}(\hat{\beta}) = \frac{1}{\left(\sum x_t^{*2}\right)^2} E\left(\sum u_t x_t^*\right)^2 \tag{2A.25}$$

Writing the terms out in the last summation of (2A.25)

$$\text{var}(\hat{\beta}) = \frac{1}{\left(\sum x_t^{*2}\right)^2} E\left(u_1 x_1^* + u_2 x_2^* + \cdots + u_T x_T^*\right)^2 \tag{2A.26}$$

Now expanding the brackets of the squared term in the expectations operator of (2A.26)

$$\text{var}(\hat{\beta}) = \frac{1}{\left(\sum x_t^{*2}\right)^2} E\left(u_1^2 x_1^{*2} + u_2^2 x_2^{*2} + \cdots + u_T^2 x_T^{*2} + \text{cross-products}\right) \tag{2A.27}$$

where '*cross-products*' in (2A.27) denotes all of the terms $u_i x_i^* u_j x_j^*$ $(i \neq j)$. These cross-products can be written as $u_i u_j x_i^* x_j^*$ $(i \neq j)$ and their expectation will be zero under the assumption that the error terms are uncorrelated with one another. Thus, the '*cross-products*' term in (2A.27) will drop out. Recall also from the chapter text that $E(u_t^2)$ is the error variance, which is estimated using s^2

$$\text{var}(\hat{\beta}) = \frac{1}{\left(\sum x_t^{*2}\right)^2} \left(s^2 x_1^{*2} + s^2 x_2^{*2} + \cdots + s^2 x_T^{*2}\right) \tag{2A.28}$$

which can also be written

$$\text{var}(\hat{\beta}) = \frac{s^2}{\left(\sum x_t^{*2}\right)^2} \left(x_1^{*2} + x_2^{*2} + \cdots + x_T^{*2}\right) = \frac{s^2 \sum x_t^{*2}}{\left(\sum x_t^{*2}\right)^2} \tag{2A.29}$$

A term in $\sum x_t^{*2}$ can be cancelled from the numerator and denominator of (2A.29), and recalling that $x_t^* = (x_t - \bar{x})$, this gives the variance of the slope coefficient as

$$\text{var}(\hat{\beta}) = \frac{s^2}{\sum (x_t - \bar{x})^2} \tag{2A.30}$$

so that the standard error can be obtained by taking the square root of (2A.30)

$$SE(\hat{\beta}) = s\sqrt{\frac{1}{\sum(x_t - \bar{x})^2}} \tag{2A.31}$$

Turning now to the derivation of the intercept standard error, this is in fact much more difficult than that of the slope standard error. In fact, both are very much easier using matrix algebra as shown below. Therefore, this derivation will be offered in summary form. It is possible to express $\hat{\alpha}$ as a function of the true α and of the disturbances, u_t

$$\hat{\alpha} = \alpha + \frac{\sum u_t\left[\sum x_t^2 - x_t \sum x_t\right]}{\left[T\sum x_t^2 - \left(\sum x_t\right)^2\right]} \tag{2A.32}$$

Denoting all of the elements in square brackets as g_t, (2A.32) can be written

$$\hat{\alpha} - \alpha = \sum u_t g_t \tag{2A.33}$$

From (2A.15), the intercept variance would be written

$$\text{var}(\hat{\alpha}) = E\left(\sum u_t g_t\right)^2 = \sum g_t^2 E(u_t^2) = s^2 \sum g_t^2 \tag{2A.34}$$

Writing (2A.34) out in full for g_t^2 and expanding the brackets

$$\text{var}(\hat{\alpha}) = \frac{s^2\left[T\left(\sum x_t^2\right)^2 - 2\sum x_t\left(\sum x_t^2\right)\sum x_t + \left(\sum x_t^2\right)\left(\sum x_t\right)^2\right]}{\left[T\sum x_t^2 - \left(\sum x_t\right)^2\right]^2} \tag{2A.35}$$

This looks rather complex, but fortunately, if we take $\sum x_t^2$ outside the square brackets in the numerator, the remaining numerator cancels with a term in the denominator to leave the required result

$$SE(\hat{\alpha}) = s\sqrt{\frac{\sum x_t^2}{T\sum(x_t - \bar{x})^2}} \tag{2A.36}$$

Review questions

1. (a) Why does OLS estimation involve taking vertical deviations of the points to the line rather than horizontal distances?
   (b) Why are the vertical distances squared before being added together?

(c) Why are the squares of the vertical distances taken rather than the absolute values?

2. Explain, with the use of equations, the difference between the sample regression function and the population regression function.

3. What is an estimator? Is the OLS estimator superior to all other estimators? Why or why not?

4. What five assumptions are usually made about the unobservable error terms in the classical linear regression model (CLRM)? Briefly explain the meaning of each. Why are these assumptions made?

5. Which of the following models can be estimated (following a suitable rearrangement if necessary) using ordinary least squares (OLS), where X, y, Z are variables and α, β, γ are parameters to be estimated? (*Hint*: the models need to be linear in the parameters.)

$$y_t = \alpha + \beta x_t + u_t \tag{2.57}$$

$$y_t = e^{\alpha} x_t^{\beta} e^{u_t} \tag{2.58}$$

$$y_t = \alpha + \beta \gamma x_t + u_t \tag{2.59}$$

$$\ln(y_t) = \alpha + \beta \ln(x_t) + u_t \tag{2.60}$$

$$y_t = \alpha + \beta x_t z_t + u_t \tag{2.61}$$

6. The capital asset pricing model (CAPM) can be written as

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] \tag{2.62}$$

using the standard notation.

The first step in using the CAPM is to estimate the stock's beta using the market model. The market model can be written as

$$R_{it} = \alpha_i + \beta_i R_{mt} + u_{it} \tag{2.63}$$

where R_{it} is the excess return for security i at time t, R_{mt} is the excess return on a proxy for the market portfolio at time t, and u_t is an iid random disturbance term. The cofficient beta in this case is also the CAPM beta for security i.

Suppose that you had estimated (2.63) and found that the estimated value of beta for a stock, $\hat{\beta}$ was 1.147. The standard error associated with this coefficient $SE(\hat{\beta})$ is estimated to be 0.0548.

A city analyst has told you that this security closely follows the market, but that it is no more risky, on average, than the market. This can be tested by the null hypotheses that the value of beta is one. The model is estimated over 62 daily observations. Test this hypothesis against a one-sided alternative that the security is more risky than the

market, at the 5% level. Write down the null and alternative hypothesis. What do you conclude? Are the analyst's claims empirically verified?

7. The analyst also tells you that shares in Chris Mining PLC have no systematic risk, in other words that the returns on its shares are completely unrelated to movements in the market. The value of beta and its standard error are calculated to be 0.214 and 0.186, respectively. The model is estimated over 38 quarterly observations. Write down the null and alternative hypotheses. Test this null hypothesis against a two-sided alternative.

8. Form and interpret a 95% and a 99% confidence interval for beta using the figures given in question 7.

9. Are hypotheses tested concerning the actual values of the coefficients (i.e. β) or their estimated values (i.e. $\hat{\beta}$) and why?

10. Using EViews, select one of the other stock series from the 'capm.wk1' file and estimate a CAPM beta for that stock. Test the null hypothesis that the true beta is one and also test the null hypothesis that the true alpha (intercept) is zero. What are your conclusions?