



Modelling long-run relationships in finance

Learning Outcomes

In this chapter, you will learn how to

- Highlight the problems that may occur if non-stationary data are used in their levels form
 - Test for unit roots
 - Examine whether systems of variables are cointegrated
 - Estimate error correction and vector error correction models
 - Explain the intuition behind Johansen's test for cointegration
 - Describe how to test hypotheses in the Johansen framework
 - Construct models for long-run relationships between variables in EViews
-

7.1 Stationarity and unit root testing

7.1.1 Why are tests for non-stationarity necessary?

There are several reasons why the concept of non-stationarity is important and why it is essential that variables that are non-stationary be treated differently from those that are stationary. Two definitions of non-stationarity were presented at the start of chapter 5. For the purpose of the analysis in this chapter, a stationary series can be defined as one with a *constant mean*, *constant variance* and *constant autocovariances* for each given lag. Therefore, the discussion in this chapter relates to the concept of weak stationarity. An examination of whether a series can be viewed as stationary or not is essential for the following reasons:

- The stationarity or otherwise of a series can *strongly influence its behaviour and properties*. To offer one illustration, the word 'shock' is usually used

to denote a change or an unexpected change in a variable or perhaps simply the value of the error term during a particular time period. For a stationary series, ‘shocks’ to the system will gradually die away. That is, a shock during time t will have a smaller effect in time $t + 1$, a smaller effect still in time $t + 2$, and so on. This can be contrasted with the case of non-stationary data, where the persistence of shocks will always be infinite, so that for a non-stationary series, the effect of a shock during time t will not have a smaller effect in time $t + 1$, and in time $t + 2$, etc.

- The use of non-stationary data can lead to *spurious regressions*. If two stationary variables are generated as independent random series, when one of those variables is regressed on the other, the t -ratio on the slope coefficient would be expected not to be significantly different from zero, and the value of R^2 would be expected to be very low. This seems obvious, for the variables are not related to one another. However, if two variables are trending over time, a regression of one on the other could have a high R^2 even if the two are totally unrelated. So, if standard regression techniques are applied to non-stationary data, the end result could be a regression that ‘looks’ good under standard measures (significant coefficient estimates and a high R^2), but which is really valueless. Such a model would be termed a ‘spurious regression’.

To give an illustration of this, two independent sets of non-stationary variables, y and x , were generated with sample size 500, one regressed on the other and the R^2 noted. This was repeated 1,000 times to obtain 1,000 R^2 values. A histogram of these values is given in figure 7.1.

As figure 7.1 shows, although one would have expected the R^2 values for each regression to be close to zero, since the explained and

Figure 7.1

Value of R^2 for 1,000 sets of regressions of a non-stationary variable on another independent non-stationary variable

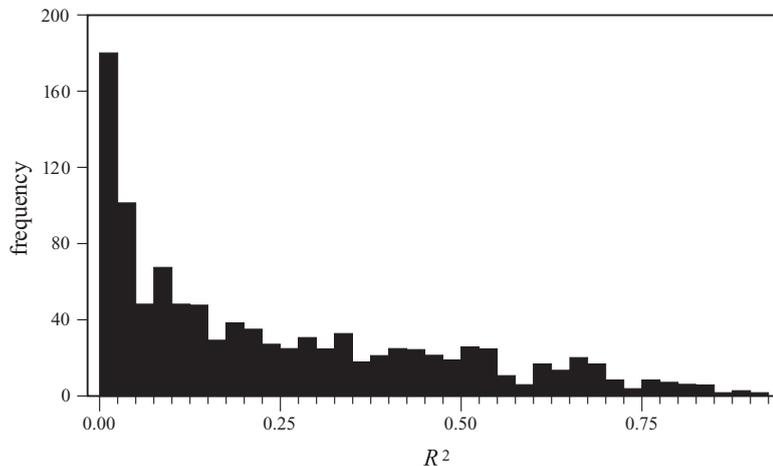
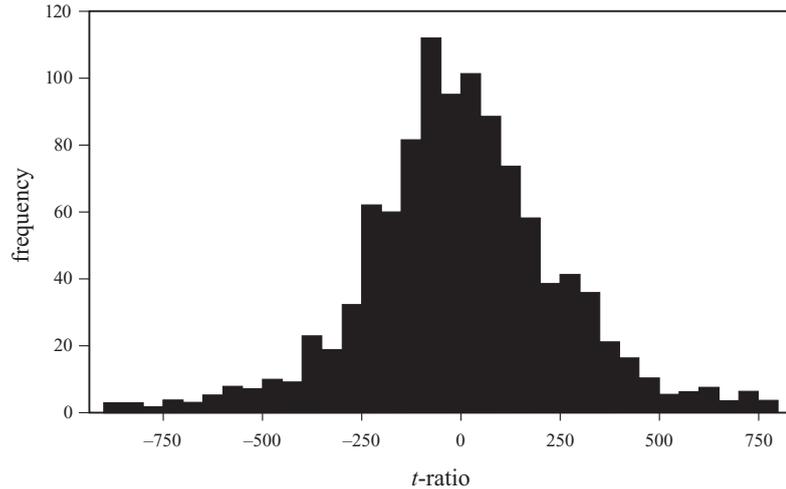


Figure 7.2

Value of t -ratio of slope coefficient for 1,000 sets of regressions of a non-stationary variable on another independent non-stationary variable



explanatory variables in each case are independent of one another, in fact R^2 takes on values across the whole range. For one set of data, R^2 is bigger than 0.9, while it is bigger than 0.5 over 16% of the time!

- If the variables employed in a regression model are *not stationary*, then it can be proved that the standard assumptions for asymptotic analysis will not be valid. In other words, the usual ' t -ratios' will not follow a t -distribution, and the F -statistic will not follow an F -distribution, and so on. Using the same simulated data as used to produce figure 7.1, figure 7.2 plots a histogram of the estimated t -ratio on the slope coefficient for each set of data.

In general, if one variable is regressed on another unrelated variable, the t -ratio on the slope coefficient will follow a t -distribution. For a sample of size 500, this implies that 95% of the time, the t -ratio will lie between ± 2 . As figure 7.2 shows quite dramatically, however, the standard t -ratio in a regression of non-stationary variables can take on enormously large values. In fact, in the above example, the t -ratio is bigger than 2 in absolute value over 98% of the time, when it should be bigger than 2 in absolute value only approximately 5% of the time! Clearly, it is therefore not possible to validly undertake hypothesis tests about the regression parameters if the data are non-stationary.

7.1.2 Two types of non-stationarity

There are two models that have been frequently used to characterise the non-stationarity, the *random walk model with drift*

$$y_t = \mu + y_{t-1} + u_t \quad (7.1)$$

and the *trend-stationary process* – so-called because it is stationary around a linear trend

$$y_t = \alpha + \beta t + u_t \quad (7.2)$$

where u_t is a white noise disturbance term in both cases.

Note that the model (7.1) could be generalised to the case where y_t is an explosive process

$$y_t = \mu + \phi y_{t-1} + u_t \quad (7.3)$$

where $\phi > 1$. Typically, this case is ignored and $\phi = 1$ is used to characterise the non-stationarity because $\phi > 1$ does not describe many data series in economics and finance, but $\phi = 1$ has been found to describe accurately many financial and economic time series. Moreover, $\phi > 1$ has an intuitively unappealing property: shocks to the system are not only persistent through time, they are propagated so that a given shock will have an increasingly large influence. In other words, the effect of a shock during time t will have a larger effect in time $t + 1$, a larger effect still in time $t + 2$, and so on. To see this, consider the general case of an AR(1) with no drift

$$y_t = \phi y_{t-1} + u_t \quad (7.4)$$

Let ϕ take any value for now. Lagging (7.4) one and then two periods

$$y_{t-1} = \phi y_{t-2} + u_{t-1} \quad (7.5)$$

$$y_{t-2} = \phi y_{t-3} + u_{t-2} \quad (7.6)$$

Substituting into (7.4) from (7.5) for y_{t-1} yields

$$y_t = \phi(\phi y_{t-2} + u_{t-1}) + u_t \quad (7.7)$$

$$y_t = \phi^2 y_{t-2} + \phi u_{t-1} + u_t \quad (7.8)$$

Substituting again for y_{t-2} from (7.6)

$$y_t = \phi^2(\phi y_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t \quad (7.9)$$

$$y_t = \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \quad (7.10)$$

T successive substitutions of this type lead to

$$y_t = \phi^{T+1} y_{t-(T+1)} + \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \cdots + \phi^T u_{t-T} + u_t \quad (7.11)$$

There are three possible cases:

(1) $\phi < 1 \Rightarrow \phi^T \rightarrow 0$ as $T \rightarrow \infty$

So the shocks to the system gradually die away – this is the *stationary case*.

$$(2) \phi = 1 \Rightarrow \phi^T = 1 \quad \forall T$$

So shocks persist in the system and never die away. The following is obtained

$$y_t = y_0 + \sum_{i=0}^{\infty} u_t \text{ as } T \rightarrow \infty \quad (7.12)$$

So the current value of y is just an infinite sum of past shocks plus some starting value of y_0 . This is known as the *unit root case*, for the root of the characteristic equation would be unity.

(3) $\phi > 1$. Now given shocks become more influential as time goes on, since if $\phi > 1$, $\phi^3 > \phi^2 > \phi$, etc. This is the *explosive case* which, for the reasons listed above, will not be considered as a plausible description of the data.

Going back to the two characterisations of non-stationarity, the random walk with drift

$$y_t = \mu + y_{t-1} + u_t \quad (7.13)$$

and the trend-stationary process

$$y_t = \alpha + \beta t + u_t \quad (7.14)$$

The two will require different treatments to induce stationarity. The second case is known as *deterministic non-stationarity* and de-trending is required. In other words, if it is believed that only this class of non-stationarity is present, a regression of the form given in (7.14) would be run, and any subsequent estimation would be done on the residuals from (7.14), which would have had the linear trend removed.

The first case is known as stochastic non-stationarity, where there is a stochastic trend in the data. Letting $\Delta y_t = y_t - y_{t-1}$ and $Ly_t = y_{t-1}$ so that $(1 - L)y_t = y_t - Ly_t = y_t - y_{t-1}$. If (7.13) is taken and y_{t-1} subtracted from both sides

$$y_t - y_{t-1} = \mu + u_t \quad (7.15)$$

$$(1 - L)y_t = \mu + u_t \quad (7.16)$$

$$\Delta y_t = \mu + u_t \quad (7.17)$$

There now exists a new variable Δy_t , which will be stationary. It would be said that stationarity has been induced by 'differencing once'. It should also be apparent from the representation given by (7.16) why y_t is also known as a *unit root process*: i.e. that the root of the characteristic equation $(1 - z) = 0$, will be unity.

Although trend-stationary and difference-stationary series are both ‘trending’ over time, the correct approach needs to be used in each case. If first differences of a trend-stationary series were taken, it would ‘remove’ the non-stationarity, but at the expense of introducing an MA(1) structure into the errors. To see this, consider the trend-stationary model

$$y_t = \alpha + \beta t + u_t \quad (7.18)$$

This model can be expressed for time $t - 1$, which would be obtained by removing 1 from all of the time subscripts in (7.18)

$$y_{t-1} = \alpha + \beta(t - 1) + u_{t-1} \quad (7.19)$$

Subtracting (7.19) from (7.18) gives

$$\Delta y_t = \beta + u_t - u_{t-1} \quad (7.20)$$

Not only is this a moving average in the errors that has been created, it is a non-invertible MA (i.e. one that cannot be expressed as an autoregressive process). Thus the series, Δy_t would in this case have some very undesirable properties.

Conversely if one tried to de-trend a series which has stochastic trend, then the non-stationarity would not be removed. Clearly then, it is not always obvious which way to proceed. One possibility is to nest both cases in a more general model and to test that. For example, consider the model

$$\Delta y_t = \alpha_0 + \alpha_1 t + (\gamma - 1)y_{t-1} + u_t \quad (7.21)$$

Although again, of course the t -ratios in (7.21) will not follow a t -distribution. Such a model could allow for both deterministic and stochastic non-stationarity. However, this book will now concentrate on the stochastic stationarity model since it is the model that has been found to best describe most non-stationary financial and economic time series. Consider again the simplest stochastic trend model

$$y_t = y_{t-1} + u_t \quad (7.22)$$

or

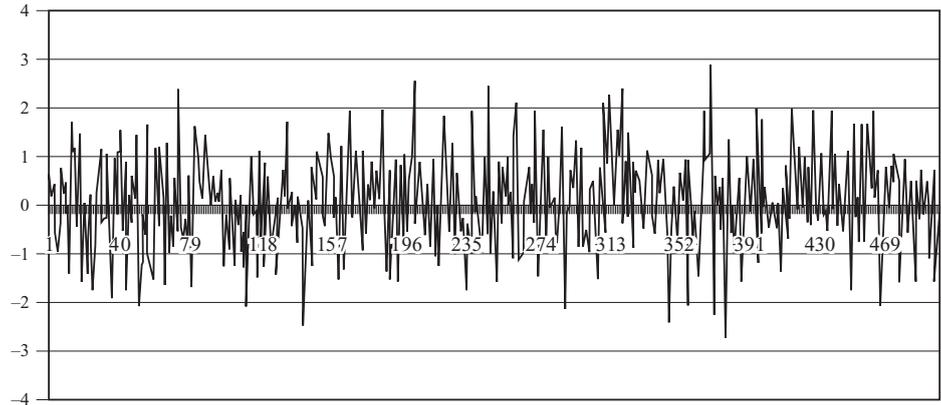
$$\Delta y_t = u_t \quad (7.23)$$

This concept can be generalised to consider the case where the series contains more than one ‘unit root’. That is, the first difference operator, Δ , would need to be applied more than once to induce stationarity. This situation will be described later in this chapter.

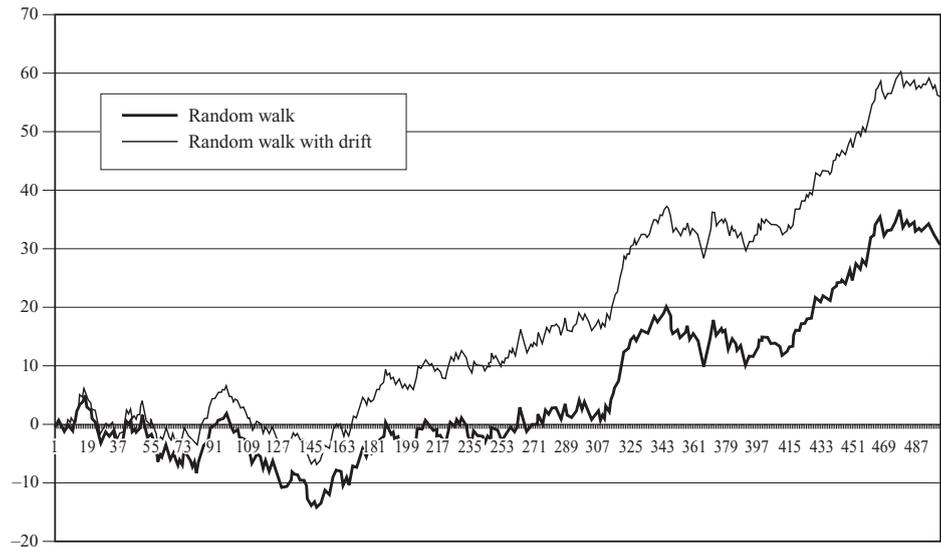
Arguably the best way to understand the ideas discussed above is to consider some diagrams showing the typical properties of certain relevant

Figure 7.3

Example of a white noise process

**Figure 7.4**

Time series plot of a random walk versus a random walk with drift

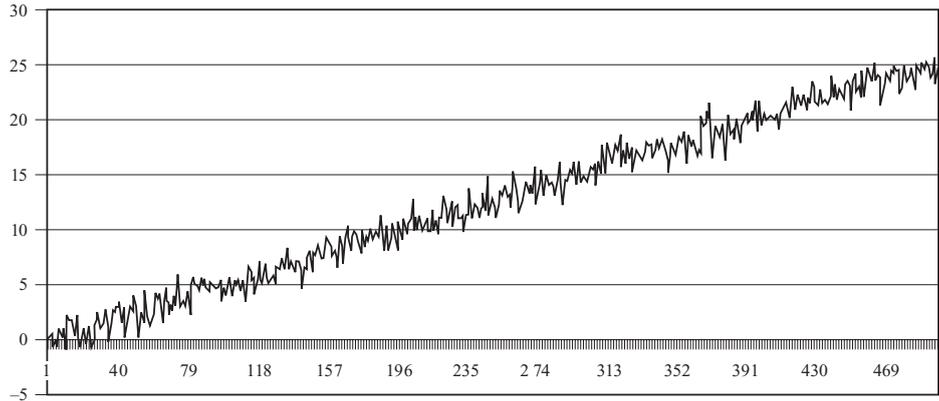


types of processes. Figure 7.3 plots a white noise (pure random) process, while figures 7.4 and 7.5 plot a random walk versus a random walk with drift and a deterministic trend process, respectively.

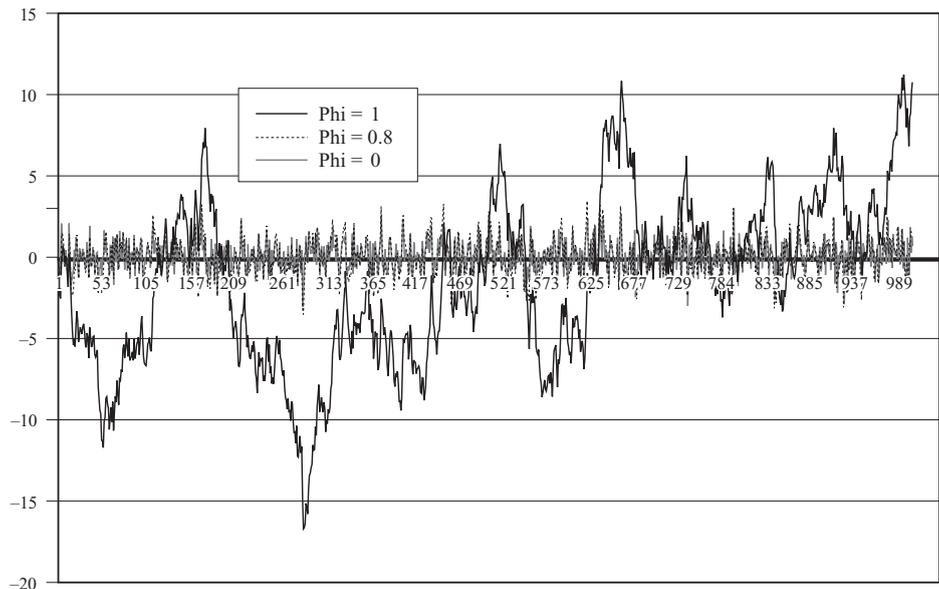
Comparing these three figures gives a good idea of the differences between the properties of a stationary, a stochastic trend and a deterministic trend process. In figure 7.3, a white noise process visibly has no trending behaviour, and it frequently crosses its mean value of zero. The random walk (thick line) and random walk with drift (faint line) processes of figure 7.4 exhibit 'long swings' away from their mean value, which they cross very rarely. A comparison of the two lines in this graph reveals that the positive drift leads to a series that is more likely to rise over time than to fall; obviously, the effect of the drift on the series becomes greater and

Figure 7.5

Time series plot of a deterministic trend process

**Figure 7.6**

Autoregressive processes with differing values of ϕ (0, 0.8, 1)



greater the further the two processes are tracked. Finally, the deterministic trend process of figure 7.5 clearly does not have a constant mean, and exhibits completely random fluctuations about its upward trend. If the trend were removed from the series, a plot similar to the white noise process of figure 7.3 would result. In this author's opinion, more time series in finance and economics look like figure 7.4 than either figure 7.3 or 7.5. Consequently, as stated above, the stochastic trend model will be the focus of the remainder of this chapter.

Finally, figure 7.6 plots the value of an autoregressive process of order 1 with different values of the autoregressive coefficient as given by (7.4).

Values of $\phi = 0$ (i.e. a white noise process), $\phi = 0.8$ (i.e. a stationary AR(1)) and $\phi = 1$ (i.e. a random walk) are plotted over time.

7.1.3 Some more definitions and terminology

If a non-stationary series, y_t must be differenced d times before it becomes stationary, then it is said to be integrated of order d . This would be written $y_t \sim I(d)$. So if $y_t \sim I(d)$ then $\Delta^d y_t \sim I(0)$. This latter piece of terminology states that applying the difference operator, Δ , d times, leads to an $I(0)$ process, i.e. a process with no unit roots. In fact, applying the difference operator more than d times to an $I(d)$ process will still result in a stationary series (but with an MA error structure). An $I(0)$ series is a stationary series, while an $I(1)$ series contains one unit root. For example, consider the random walk

$$y_t = y_{t-1} + u_t \quad (7.24)$$

An $I(2)$ series contains two unit roots and so would require differencing twice to induce stationarity. $I(1)$ and $I(2)$ series can wander a long way from their mean value and cross this mean value rarely, while $I(0)$ series should cross the mean frequently. The majority of financial and economic time series contain a single unit root, although some are stationary and some have been argued to possibly contain two unit roots (series such as nominal consumer prices and nominal wages). The efficient markets hypothesis together with rational expectations suggest that asset prices (or the natural logarithms of asset prices) should follow a random walk or a random walk with drift, so that their differences are unpredictable (or only predictable to their long-term average value).

To see what types of data generating process could lead to an $I(2)$ series, consider the equation

$$y_t = 2y_{t-1} - y_{t-2} + u_t \quad (7.25)$$

taking all of the terms in y over to the LHS, and then applying the lag operator notation

$$y_t - 2y_{t-1} + y_{t-2} = u_t \quad (7.26)$$

$$(1 - 2L + L^2)y_t = u_t \quad (7.27)$$

$$(1 - L)(1 - L)y_t = u_t \quad (7.28)$$

It should be evident now that this process for y_t contains two unit roots, and would require differencing twice to induce stationarity.

What would happen if y_t in (7.25) were differenced only once? Taking first differences of (7.25), i.e. subtracting y_{t-1} from both sides

$$y_t - y_{t-1} = y_{t-1} - y_{t-2} + u_t \quad (7.29)$$

$$y_t - y_{t-1} = (y_t - y_{t-1})_{-1} + u_t \quad (7.30)$$

$$\Delta y_t = \Delta y_{t-1} + u_t \quad (7.31)$$

$$(1 - L)\Delta y_t = u_t \quad (7.32)$$

First differencing would therefore have removed one of the unit roots, but there is still a unit root remaining in the new variable, Δy_t .

7.1.4 Testing for a unit root

One immediately obvious (but inappropriate) method that readers may think of to test for a unit root would be to examine the autocorrelation function of the series of interest. However, although shocks to a unit root process will remain in the system indefinitely, the acf for a unit root process (a random walk) will often be seen to decay away very slowly to zero. Thus, such a process may be mistaken for a highly persistent but stationary process. Hence it is not possible to use the acf or pacf to determine whether a series is characterised by a unit root or not. Furthermore, even if the true data generating process for y_t contains a unit root, the results of the tests for a given sample could lead one to believe that the process is stationary. Therefore, what is required is some kind of formal hypothesis testing procedure that answers the question, ‘given the sample of data to hand, is it plausible that the true data generating process for y contains one or more unit roots?’

The early and pioneering work on testing for a unit root in time series was done by Dickey and Fuller (Fuller, 1976; Dickey and Fuller, 1979). The basic objective of the test is to examine the null hypothesis that $\phi = 1$ in

$$y_t = \phi y_{t-1} + u_t \quad (7.33)$$

against the one-sided alternative $\phi < 1$. Thus the hypotheses of interest are H_0 : series contains a unit root versus H_1 : series is stationary.

In practice, the following regression is employed, rather than (7.33), for ease of computation and interpretation

$$\Delta y_t = \psi y_{t-1} + u_t \quad (7.34)$$

so that a test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (since $\phi - 1 = \psi$).

Dickey–Fuller (DF) tests are also known as τ -tests, and can be conducted allowing for an intercept, or an intercept and deterministic trend, or

Table 7.1 Critical values for DF tests (Fuller, 1976, p. 373)

Significance level	10%	5%	1%
CV for constant but no trend	-2.57	-2.86	-3.43
CV for constant and trend	-3.12	-3.41	-3.96

neither, in the test regression. The model for the unit root test in each case is

$$y_t = \phi y_{t-1} + \mu + \lambda t + u_t \quad (7.35)$$

The tests can also be written, by subtracting y_{t-1} from each side of the equation, as

$$\Delta y_t = \psi y_{t-1} + \mu + \lambda t + u_t \quad (7.36)$$

In another paper, Dickey and Fuller (1981) provide a set of additional test statistics and their critical values for joint tests of the significance of the lagged y , and the constant and trend terms. These are not examined further here. The test statistics for the original DF tests are defined as

$$\text{test statistic} = \frac{\hat{\psi}}{SE(\hat{\psi})} \quad (7.37)$$

The test statistics do not follow the usual t -distribution under the null hypothesis, since the null is one of non-stationarity, but rather they follow a non-standard distribution. Critical values are derived from simulations experiments in, for example, Fuller (1976); see also chapter 12 in this book. Relevant examples of the distribution are shown in table 7.1. A full set of Dickey–Fuller (DF) critical values is given in the appendix of statistical tables at the end of this book. A discussion and example of how such critical values (CV) are derived using simulations methods are presented in chapter 12.

Comparing these with the standard normal critical values, it can be seen that the DF critical values are much bigger in absolute terms (i.e. more negative). Thus more evidence against the null hypothesis is required in the context of unit root tests than under standard t -tests. This arises partly from the inherent instability of the unit root process, the fatter distribution of the t -ratios in the context of non-stationary data (see figure 7.2), and the resulting uncertainty in inference. The null hypothesis of a unit root is rejected in favour of the stationary alternative in each case if the test statistic is more negative than the critical value.

The tests above are valid only if u_t is white noise. In particular, u_t is assumed not to be autocorrelated, but would be so if there was autocorrelation in the dependent variable of the regression (Δy_t) which has not been modelled. If this is the case, the test would be ‘oversized’, meaning that the true size of the test (the proportion of times a correct null hypothesis is incorrectly rejected) would be higher than the nominal size used (e.g. 5%). The solution is to ‘augment’ the test using p lags of the dependent variable. The alternative model in case (i) is now written

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + u_t \quad (7.38)$$

The lags of Δy_t now ‘soak up’ any dynamic structure present in the dependent variable, to ensure that u_t is not autocorrelated. The test is known as an augmented Dickey–Fuller (ADF) test and is still conducted on ψ , and the same critical values from the DF tables are used as before.

A problem now arises in determining the optimal number of lags of the dependent variable. Although several ways of choosing p have been proposed, they are all somewhat arbitrary, and are thus not presented here. Instead, the following two simple rules of thumb are suggested. First, the *frequency of the data* can be used to decide. So, for example, if the data are monthly, use 12 lags, if the data are quarterly, use 4 lags, and so on. Clearly, there would not be an obvious choice for the number of lags to use in a regression containing higher frequency financial data (e.g. hourly or daily)! Second, an *information criterion* can be used to decide. So choose the number of lags that minimises the value of an information criterion, as outlined in chapter 6.

It is quite important to attempt to use an optimal number of lags of the dependent variable in the test regression, and to examine the sensitivity of the outcome of the test to the lag length chosen. In most cases, hopefully the conclusion will not be qualitatively altered by small changes in p , but sometimes it will. Including too few lags will not remove all of the autocorrelation, thus biasing the results, while using too many will increase the coefficient standard errors. The latter effect arises since an increase in the number of parameters to estimate uses up degrees of freedom. Therefore, everything else being equal, the absolute values of the test statistics will be reduced. This will result in a reduction in the power of the test, implying that for a stationary process the null hypothesis of a unit root will be rejected less frequently than would otherwise have been the case.

7.1.5 Testing for higher orders of integration

Consider the simple regression

$$\Delta y_t = \psi y_{t-1} + u_t \quad (7.39)$$

$H_0: \psi = 0$ is tested against $H_1: \psi < 0$.

If H_0 is rejected, it would simply be concluded that y_t does not contain a unit root. But what should be the conclusion if H_0 is not rejected? The series contains a unit root, but is that it? No! What if $y_t \sim I(2)$? The null hypothesis would still not have been rejected. It is now necessary to perform a test of

$$H_0 : y_t \sim I(2) \text{ vs. } H_1 : y_t \sim I(1)$$

$\Delta^2 y_t (= \Delta y_t - \Delta y_{t-1})$ would now be regressed on Δy_{t-1} (plus lags of $\Delta^2 y_t$ to augment the test if necessary). Thus, testing $H_0: \Delta y_t \sim I(1)$ is equivalent to $H_0: y_t \sim I(2)$. So in this case, if H_0 is not rejected (very unlikely in practice), it would be concluded that y_t is at least $I(2)$. If H_0 is rejected, it would be concluded that y_t contains a single unit root. The tests should continue for a further unit root until H_0 is rejected.

Dickey and Pantula (1987) have argued that an ordering of the tests as described above (i.e. testing for $I(1)$, then $I(2)$, and so on) is, strictly speaking, invalid. The theoretically correct approach would be to start by assuming some highest plausible order of integration (e.g. $I(2)$), and to test $I(2)$ against $I(1)$. If $I(2)$ is rejected, then test $I(1)$ against $I(0)$. In practice, however, to the author's knowledge, no financial time series contain more than a single unit root, so that this matter is of less concern in finance.

7.1.6 Phillips–Perron (PP) tests

Phillips and Perron have developed a more comprehensive theory of unit root non-stationarity. The tests are similar to ADF tests, but they incorporate an automatic correction to the DF procedure to allow for autocorrelated residuals. The tests often give the same conclusions as, and suffer from most of the same important limitations as, the ADF tests.

7.1.7 Criticisms of Dickey–Fuller- and Phillips–Perron-type tests

The most important criticism that has been levelled at unit root tests is that their power is low if the process is stationary but with a root close to the non-stationary boundary. So, for example, consider an AR(1) data generating process with coefficient 0.95. If the true data generating process is

$$y_t = 0.95y_{t-1} + u_t \quad (7.40)$$

Box 7.1 Stationarity tests

Stationarity tests have stationarity under the null hypothesis, thus reversing the null and alternatives under the Dickey–Fuller approach. Thus, under stationarity tests, the data will appear stationary by default if there is little information in the sample. One such stationarity test is the KPSS test (Kwiatkowski *et al.*, 1992). The computation of the test statistic is not discussed here but the test is available within the EViews software. The results of these tests can be compared with the ADF/PP procedure to see if the same conclusion is obtained. The null and alternative hypotheses under each testing approach are as follows:

ADF/PP	KPSS
$H_0 : y_t \sim I(1)$	$H_0 : y_t \sim I(0)$
$H_1 : y_t \sim I(0)$	$H_1 : y_t \sim I(1)$

There are four possible outcomes:

- | | | |
|-------------------------|-----|---------------------|
| (1) Reject H_0 | and | Do not reject H_0 |
| (2) Do not Reject H_0 | and | Reject H_0 |
| (3) Reject H_0 | and | Reject H_0 |
| (4) Do not reject H_0 | and | Do not reject H_0 |

For the conclusions to be robust, the results should fall under outcomes 1 or 2, which would be the case when both tests concluded that the series is stationary or non-stationary, respectively. Outcomes 3 or 4 imply conflicting results. The joint use of stationarity and unit root tests is known as *confirmatory data analysis*.

the null hypothesis of a unit root should be rejected. It has been thus argued that the tests are poor at deciding, for example, whether $\phi = 1$ or $\phi = 0.95$, especially with small sample sizes. The source of this problem is that, under the classical hypothesis-testing framework, the null hypothesis is never accepted, it is simply stated that it is either rejected or not rejected. This means that a failure to reject the null hypothesis could occur either because the null was correct, or because there is insufficient information in the sample to enable rejection. One way to get around this problem is to use a stationarity test as well as a unit root test, as described in box 7.1.

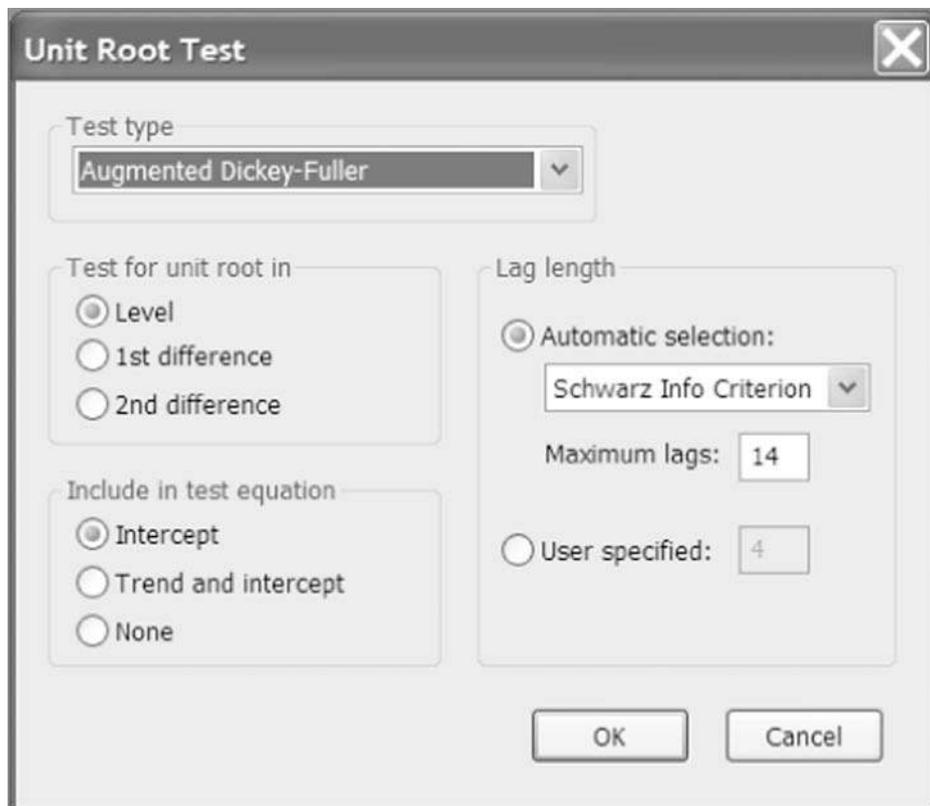
7.2 Testing for unit roots in EViews

This example uses the same data on UK house prices as employed in chapter 5. Assuming that the data have been loaded, and the variables are defined as in chapter 5, double click on the icon next to the name of the series that you want to perform the unit root test on, so that a spreadsheet

appears containing the observations on that series. Open the raw house price series, 'hp' by clicking on the **hp** icon. Next, click on the **View** button on the button bar above the spreadsheet and then **Unit Root Test...** You will then be presented with a menu containing various options, as in screenshot 7.1.

Screenshot 7.1

Options menu for unit root tests



From this, choose the following options:

- | | |
|------------------------------|-------------------------|
| (1) Test Type | Augmented Dickey-Fuller |
| (2) Test for Unit Root in | Levels |
| (3) Include in test equation | Intercept |
| (4) Maximum lags | 12 |

and click **OK**.

This will obviously perform an augmented Dickey-Fuller (ADF) test with up to 12 lags of the dependent variable in a regression equation on the raw data series with a constant but no trend in the test equation. EViews presents a large number of options here – for example, instead of the

Dickey-Fuller series, we could run the Phillips-Perron or KPSS tests as described above. Or, if we find that the levels of the series are non-stationary, we could repeat the analysis on the first differences directly from this menu rather than having to create the first differenced series separately. We can also choose between various methods for determining the optimum lag length in an augmented Dickey-Fuller test, with the Schwarz criterion being the default. The results for the raw house price series would appear as in the following table.

Null Hypothesis: HP has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	2.707012	1.0000
Test critical values:	1% level	-3.464101
	5% level	-2.876277
	10% level	-2.574704

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(HP)

Method: Least Squares

Date: 09/05/07 Time: 21:15

Sample (adjusted): 1991M04 2007M05

Included observations: 194 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
HP(-1)	0.004890	0.001806	2.707012	0.0074
D(HP(-1))	0.220916	0.070007	3.155634	0.0019
D(HP(-2))	0.291059	0.070711	4.116164	0.0001
C	-99.91536	155.1872	-0.643838	0.5205
R-squared	0.303246	Mean dependent var		663.3590
Adjusted R-squared	0.292244	S.D. dependent var		1081.701
S.E. of regression	910.0161	Akaike info criterion		16.48520
Sum squared resid	1.57E+08	Schwarz criterion		16.55258
Log likelihood	-1595.065	Hannan-Quinn criter.		16.51249
F-statistic	27.56430	Durbin-Watson stat		2.010299
Prob(F-statistic)	0.000000			

The value of the test statistic and the relevant critical values given the type of test equation (e.g. whether there is a constant and/or trend included) and sample size, are given in the first panel of the output above.

Schwarz's criterion has in this case chosen to include 2 lags of the dependent variable in the test regression. Clearly, the test statistic is not more negative than the critical value, so the null hypothesis of a unit root in the house price series cannot be rejected. The remainder of the output presents the estimation results. Since the dependent variable in this regression is non-stationary, it is not appropriate to examine the coefficient standard errors or their t -ratios in the test regression.

Now repeat all of the above steps for the **first difference of the house price series** (use the 'First Difference' option in the unit root testing window rather than using the level of the dhp series). The output would appear as in the following table

Null Hypothesis: D(HP) has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic based on SIC, MAXLAG=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.112531	0.0000
Test critical values:		
1% level	-3.464101	
5% level	-2.876277	
10% level	-2.574704	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(HP,2)
 Method: Least Squares
 Date: 09/05/07 Time: 21:20
 Sample (adjusted): 1991M04 2007M05
 Included observations: 194 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(HP(-1))	-0.374773	0.073305	-5.112531	0.0000
D(HP(-1),2)	-0.346556	0.068786	-5.038192	0.0000
C	259.6274	81.58188	3.182415	0.0017
R-squared	0.372994	Mean dependent var		9.661185
Adjusted R-squared	0.366429	S.D. dependent var		1162.061
S.E. of regression	924.9679	Akaike info criterion		16.51274
Sum squared resid	1.63E+08	Schwarz criterion		16.56327
Log likelihood	-1598.736	Hannan-Quinn criter.		16.53320
F-statistic	56.81124	Durbin-Watson stat		2.045299
Prob(F-statistic)	0.000000			

In this case, as one would expect, the test statistic is more negative than the critical value and hence the null hypothesis of a unit root in the first differences is convincingly rejected. For completeness, run a unit root test on the **levels of the dhp series**, which are the percentage changes rather than the absolute differences in prices. You should find that these are also stationary.

Finally, run the KPSS test on the hp levels series by selecting it from the 'Test Type' box in the unit root testing window. You should observe now that the test statistic exceeds the critical value, even at the 1% level, so that the null hypothesis of a *stationary series* is strongly rejected, thus confirming the result of the unit root test previously conducted on the same series.

7.3 Cointegration

In most cases, if two variables that are I(1) are linearly combined, then the combination will also be I(1). More generally, if variables with differing orders of integration are combined, the combination will have an order of integration equal to the largest. If $X_{i,t} \sim I(d_i)$ for $i = 1, 2, 3, \dots, k$ so that there are k variables each integrated of order d_i , and letting

$$z_t = \sum_{i=1}^k \alpha_i X_{i,t} \quad (7.41)$$

Then $z_t \sim I(\max d_i)$. z_t in this context is simply a linear combination of the k variables X_i . Rearranging (7.41)

$$X_{1,t} = \sum_{i=2}^k \beta_i X_{i,t} + z'_t \quad (7.42)$$

where $\beta_i = -\frac{\alpha_i}{\alpha_1}$, $z'_t = \frac{z_t}{\alpha_1}$, $i = 2, \dots, k$. All that has been done is to take one of the variables, $X_{1,t}$, and to rearrange (7.41) to make it the subject. It could also be said that the equation has been normalised on $X_{1,t}$. But viewed another way, (7.42) is just a regression equation where z'_t is a disturbance term. These disturbances would have some very undesirable properties: in general, z'_t will not be stationary and is autocorrelated if all of the X_i are I(1).

As a further illustration, consider the following regression model containing variables y_t, x_{2t}, x_{3t} which are all I(1)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad (7.43)$$

For the estimated model, the SRF would be written

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{u}_t \quad (7.44)$$

Taking everything except the residuals to the LHS

$$y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t} = \hat{u}_t \quad (7.45)$$

Again, the residuals when expressed in this way can be considered a linear combination of the variables. Typically, this linear combination of I(1) variables will itself be I(1), but it would obviously be desirable to obtain residuals that are I(0). Under what circumstances will this be the case? The answer is that a linear combination of I(1) variables will be I(0), in other words stationary, if the variables are *cointegrated*.

7.3.1 Definition of cointegration (Engle and Granger, 1987)

Let w_t be a $k \times 1$ vector of variables, then the components of w_t are integrated of order (d, b) if:

- (1) All components of w_t are I(d)
- (2) There is at least one vector of coefficients α such that

$$\alpha' w_t \sim I(d - b)$$

In practice, many financial variables contain one unit root, and are thus I(1), so that the remainder of this chapter will restrict analysis to the case where $d = b = 1$. In this context, a set of variables is defined as cointegrated if a linear combination of them is stationary. Many time series are non-stationary but ‘move together’ over time – that is, there exist some influences on the series (for example, market forces), which imply that the two series are bound by some relationship in the long run. A cointegrating relationship may also be seen as a long-term or equilibrium phenomenon, since it is possible that cointegrating variables may deviate from their relationship in the short run, but their association would return in the long run.

7.3.2 Examples of possible cointegrating relationships in finance

Financial theory should suggest where two or more variables would be expected to hold some long-run relationship with one another. There are many examples in finance of areas where cointegration might be expected to hold, including:

- Spot and futures prices for a given commodity or asset
- Ratio of relative prices and an exchange rate
- Equity prices and dividends.

In all three cases, market forces arising from no-arbitrage conditions suggest that there should be an equilibrium relationship between the

series concerned. The easiest way to understand this notion is perhaps to consider what would be the effect if the series were not cointegrated. If there were no cointegration, there would be no long-run relationship binding the series together, so that the series could wander apart without bound. Such an effect would arise since all linear combinations of the series would be non-stationary, and hence would not have a constant mean that would be returned to frequently.

Spot and futures prices may be expected to be cointegrated since they are obviously prices for the same asset at different points in time, and hence will be affected in very similar ways by given pieces of information. The long-run relationship between spot and futures prices would be given by the cost of carry.

Purchasing power parity (PPP) theory states that a given representative basket of goods and services should cost the same wherever it is bought when converted into a common currency. Further discussion of PPP occurs in section 7.9, but for now suffice it to say that PPP implies that the ratio of relative prices in two countries and the exchange rate between them should be cointegrated. If they did not cointegrate, assuming zero transactions costs, it would be profitable to buy goods in one country, sell them in another, and convert the money obtained back to the currency of the original country.

Finally, if it is assumed that some stock in a particular company is held to perpetuity (i.e. for ever), then the only return that would accrue to that investor would be in the form of an infinite stream of future dividend payments. Hence the discounted dividend model argues that the appropriate price to pay for a share today is the present value of all future dividends. Hence, it may be argued that one would not expect current prices to 'move out of line' with future anticipated dividends in the long run, thus implying that share prices and dividends should be cointegrated.

An interesting question to ask is whether a potentially cointegrating regression should be estimated using the levels of the variables or the logarithms of the levels of the variables. Financial theory may provide an answer as to the more appropriate functional form, but fortunately even if not, Hendry and Juselius (2000) note that if a set of series is cointegrated in levels, they will also be cointegrated in log levels.

7.4 Equilibrium correction or error correction models

When the concept of non-stationarity was first considered in the 1970s, a usual response was to independently take the first differences of each of

the I(1) variables and then to use these first differences in any subsequent modelling process. In the context of univariate modelling (e.g. the construction of ARMA models), this is entirely the correct approach. However, when the relationship between variables is important, such a procedure is inadvisable. While this approach is statistically valid, it does have the problem that pure first difference models have no long-run solution. For example, consider two series, y_t and x_t , that are both I(1). The model that one may consider estimating is

$$\Delta y_t = \beta \Delta x_t + u_t \quad (7.46)$$

One definition of the long run that is employed in econometrics implies that the variables have converged upon some long-term values and are no longer changing, thus $y_t = y_{t-1} = y$; $x_t = x_{t-1} = x$. Hence all the difference terms will be zero in (7.46), i.e. $\Delta y_t = 0$; $\Delta x_t = 0$, and thus everything in the equation cancels. Model (7.46) has no long-run solution and it therefore has nothing to say about whether x and y have an equilibrium relationship (see chapter 4).

Fortunately, there is a class of models that can overcome this problem by using combinations of first differenced and lagged levels of cointegrated variables. For example, consider the following equation

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + u_t \quad (7.47)$$

This model is known as an *error correction model* or an *equilibrium correction model*, and $y_{t-1} - \gamma x_{t-1}$ is known as the *error correction term*. Provided that y_t and x_t are cointegrated with cointegrating coefficient γ , then $(y_{t-1} - \gamma x_{t-1})$ will be I(0) even though the constituents are I(1). It is thus valid to use OLS and standard procedures for statistical inference on (7.47). It is of course possible to have an intercept in either the cointegrating term (e.g. $y_{t-1} - \alpha - \gamma x_{t-1}$) or in the model for Δy_t (e.g. $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + u_t$) or both. Whether a constant is included or not could be determined on the basis of financial theory, considering the arguments on the importance of a constant discussed in chapter 4.

The error correction model is sometimes termed an equilibrium correction model, and the two terms will be used synonymously for the purposes of this book. Error correction models are interpreted as follows. y is purported to change between $t - 1$ and t as a result of changes in the values of the explanatory variable(s), x , between $t - 1$ and t , and also in part to correct for any disequilibrium that existed during the previous period. Note that the error correction term $(y_{t-1} - \gamma x_{t-1})$ appears in (7.47) with a lag. It would be implausible for the term to appear without any lag (i.e. as $y_t - \gamma x_t$), for this would imply that y changes between $t - 1$ and

t in response to a disequilibrium at time t . γ defines the long-run relationship between x and y , while β_1 describes the short-run relationship between changes in x and changes in y . Broadly, β_2 describes the speed of adjustment back to equilibrium, and its strict definition is that it measures the proportion of last period's equilibrium error that is corrected for.

Of course, an error correction model can be estimated for more than two variables. For example, if there were three variables, x_t , w_t , y_t , that were cointegrated, a possible error correction model would be

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \Delta w_t + \beta_3 (y_{t-1} - \gamma_1 x_{t-1} - \gamma_2 w_{t-1}) + u_t \quad (7.48)$$

The *Granger representation theorem* states that if there exists a dynamic linear model with stationary disturbances and the data are $I(1)$, then the variables must be cointegrated of order $(1,1)$.

7.5 Testing for cointegration in regression: a residuals-based approach

The model for the equilibrium correction term can be generalised further to include k variables (y and the $k - 1$ x s)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t \quad (7.49)$$

u_t should be $I(0)$ if the variables $y_t, x_{2t}, \dots, x_{kt}$ are cointegrated, but u_t will still be non-stationary if they are not.

Thus it is necessary to test the residuals of (7.49) to see whether they are non-stationary or stationary. The DF or ADF test can be used on \hat{u}_t , using a regression of the form

$$\Delta \hat{u}_t = \psi \hat{u}_{t-1} + v_t \quad (7.50)$$

with v_t an iid error term.

However, since this is a test on residuals of a model, \hat{u}_t , then the critical values are changed compared to a DF or an ADF test on a series of raw data. Engle and Granger (1987) have tabulated a new set of critical values for this application and hence the test is known as the Engle–Granger (EG) test. The reason that modified critical values are required is that the test is now operating on the residuals of an estimated model rather than on raw data. The residuals have been constructed from a particular set of coefficient estimates, and the sampling estimation error in those coefficients will change the distribution of the test statistic. Engle and Yoo (1987) tabulate a new set of critical values that are larger in absolute

value (i.e. more negative) than the DF critical values, also given at the end of this book. The critical values also become more negative as the number of variables in the potentially cointegrating regression increases.

It is also possible to use the Durbin–Watson (*DW*) test statistic or the Phillips–Perron (*PP*) approach to test for non-stationarity of \hat{u}_t . If the *DW* test is applied to the residuals of the potentially cointegrating regression, it is known as the Cointegrating Regression Durbin Watson (*CRDW*). Under the null hypothesis of a unit root in the errors, $CRDW \approx 0$, so the null of a unit root is rejected if the *CRDW* statistic is larger than the relevant critical value (which is approximately 0.5).

What are the null and alternative hypotheses for any unit root test applied to the residuals of a potentially cointegrating regression?

$$H_0 : \hat{u}_t \sim I(1)$$

$$H_1 : \hat{u}_t \sim I(0).$$

Thus, under the null hypothesis there is a unit root in the potentially cointegrating regression residuals, while under the alternative, the residuals are stationary. Under the null hypothesis, therefore, a stationary linear combination of the non-stationary variables has not been found. Hence, if this null hypothesis is not rejected, there is no cointegration. The appropriate strategy for econometric modelling in this case would be to employ specifications in first differences only. Such models would have no long-run equilibrium solution, but this would not matter since no cointegration implies that there is no long-run relationship anyway.

On the other hand, if the null of a unit root in the potentially cointegrating regression's residuals is rejected, it would be concluded that a stationary linear combination of the non-stationary variables had been found. Therefore, the variables would be classed as cointegrated. The appropriate strategy for econometric modelling in this case would be to form and estimate an error correction model, using a method described below.

Box 7.2 Multiple cointegrating relationships

In the case where there are only two variables in an equation, y_t , and x_t , say, there can be at most only one linear combination of y_t , and x_t that is stationary – i.e. at most one cointegrating relationship. However, suppose that there are k variables in a system (ignoring any constant term), denoted $y_t, x_{2t}, \dots, x_{kt}$. In this case, there may be up to r linearly independent cointegrating relationships (where $r \leq k - 1$). This potentially presents a problem for the OLS regression approach described above, which is capable of finding at most one cointegrating relationship no matter how many variables there are in the system. And if there are multiple cointegrating relationships, how can one know if there are others, or whether the 'best' or strongest cointegrating relationship

has been found? An OLS regression will find the minimum variance stationary linear combination of the variables,¹ but there may be other linear combinations of the variables that have more intuitive appeal. The answer to this problem is to use a systems approach to cointegration, which will allow determination of all r cointegrating relationships. One such approach is Johansen's method – see section 7.8.

7.6 Methods of parameter estimation in cointegrated systems

What should be the modelling strategy if the data at hand are thought to be non-stationary and possibly cointegrated? There are (at least) three methods that could be used: Engle–Granger, Engle–Yoo and Johansen. The first and third of these will be considered in some detail below.

7.6.1 The Engle–Granger 2-step method

This is a single equation technique, which is conducted as follows:

Step 1

Make sure that all the individual variables are $I(1)$. Then estimate the cointegrating regression using OLS. Note that it is not possible to perform any inferences on the coefficient estimates in this regression – all that can be done is to estimate the parameter values. Save the residuals of the cointegrating regression, \hat{u}_t . Test these residuals to ensure that they are $I(0)$. If they are $I(0)$, proceed to Step 2; if they are $I(1)$, estimate a model containing only first differences.

Step 2

Use the step 1 residuals as one variable in the error correction model, e.g.

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2(\hat{u}_{t-1}) + v_t \quad (7.51)$$

where $\hat{u}_{t-1} = y_{t-1} - \hat{\tau}x_{t-1}$. The stationary, linear combination of non-stationary variables is also known as the *cointegrating vector*. In this case, the cointegrating vector would be $[1 - \hat{\tau}]$. Additionally, any linear transformation of the cointegrating vector will also be a cointegrating vector. So, for example, $-10y_{t-1} + 10\hat{\tau}x_{t-1}$ will also be stationary. In (7.45) above, the cointegrating vector would be $[1 - \hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3]$. It is now valid to perform

¹ Readers who are familiar with the literature on hedging with futures will recognise that running an OLS regression will minimise the variance of the hedged portfolio, i.e. it will minimise the regression's residual variance, and the situation here is analogous.

inferences in the second-stage regression, i.e. concerning the parameters β_1 and β_2 (provided that there are no other forms of misspecification, of course), since all variables in this regression are stationary.

The Engle–Granger 2-step method suffers from a number of problems:

- (1) The usual finite sample problem of a *lack of power in unit root and cointegration tests* discussed above.
- (2) There could be a *simultaneous equations bias* if the causality between y and x runs in both directions, but this single equation approach requires the researcher to normalise on one variable (i.e. to specify one variable as the dependent variable and the others as independent variables). The researcher is forced to treat y and x asymmetrically, even though there may have been no theoretical reason for doing so. A further issue is the following. Suppose that the following specification had been estimated as a potential cointegrating regression

$$y_t = \alpha_1 + \beta_1 x_t + u_{1t} \quad (7.52)$$

What if instead the following equation was estimated?

$$x_t = \alpha_2 + \beta_2 y_t + u_{2t} \quad (7.53)$$

If it is found that $u_{1t} \sim I(0)$, does this imply automatically that $u_{2t} \sim I(0)$? The answer in theory is ‘yes’, but in practice different conclusions may be reached in finite samples. Also, if there is an error in the model specification at stage 1, this will be carried through to the cointegration test at stage 2, as a consequence of the sequential nature of the computation of the cointegration test statistic.

- (3) It is not possible to perform any *hypothesis tests* about the actual cointegrating relationship estimated at stage 1.

Problems 1 and 2 are small sample problems that should disappear asymptotically. Problem 3 is addressed by another method due to Engle and Yoo. There is also another alternative technique, which overcomes problems 2 and 3 by adopting a different approach based on estimation of a VAR system – see section 7.8.

7.6.2 The Engle and Yoo 3-step method

The Engle and Yoo (1987) 3-step procedure takes its first two steps from Engle–Granger (EG). Engle and Yoo then add a third step giving updated estimates of the cointegrating vector and its standard errors. The Engle and Yoo (EY) third step is algebraically technical and additionally, EY suffers from all of the remaining problems of the EG approach. There is

arguably a far superior procedure available to remedy the lack of testability of hypotheses concerning the cointegrating relationship – namely, the Johansen (1988) procedure. For these reasons, the Engle–Yoo procedure is rarely employed in empirical applications and is not considered further here.

There now follows an application of the Engle–Granger procedure in the context of spot and futures markets.

7.7 Lead–lag and long-term relationships between spot and futures markets

7.7.1 Background

If the markets are frictionless and functioning efficiently, changes in the (log of the) spot price of a financial asset and its corresponding changes in the (log of the) futures price would be expected to be perfectly contemporaneously correlated and not to be cross-autocorrelated. Mathematically, these notions would be represented as

$$\text{corr}(\Delta \log(f_t), \Delta \ln(s_t)) \approx 1 \quad (\text{a})$$

$$\text{corr}(\Delta \log(f_t), \Delta \ln(s_{t-k})) \approx 0 \quad \forall k > 0 \quad (\text{b})$$

$$\text{corr}(\Delta \log(f_{t-j}), \Delta \ln(s_t)) \approx 0 \quad \forall j > 0 \quad (\text{c})$$

In other words, changes in spot prices and changes in futures prices are expected to occur at the same time (condition (a)). The current change in the futures price is also expected not to be related to previous changes in the spot price (condition (b)), and the current change in the spot price is expected not to be related to previous changes in the futures price (condition (c)). The changes in the log of the spot and futures prices are also of course known as the spot and futures returns.

For the case when the underlying asset is a stock index, the equilibrium relationship between the spot and futures prices is known as the *cost of carry model*, given by

$$F_t^* = S_t e^{(r-d)(T-t)} \quad (7.54)$$

where F_t^* is the fair futures price, S_t is the spot price, r is a continuously compounded risk-free rate of interest, d is the continuously compounded yield in terms of dividends derived from the stock index until the futures contract matures, and $(T - t)$ is the time to maturity of the futures contract. Taking logarithms of both sides of (7.54) gives

$$f_t^* = s_t + (r - d)(T - t) \quad (7.55)$$

Table 7.2 DF tests on log-prices and returns for high frequency FTSE data

	Futures	Spot
Dickey-Fuller statistics for log-price data	-0.1329	-0.7335
Dickey-Fuller statistics for returns data	-84.9968	-114.1803

where f_t^* is the log of the fair futures price and s_t is the log of the spot price. Equation (7.55) suggests that the long-term relationship between the logs of the spot and futures prices should be one to one. Thus the basis, defined as the difference between the futures and spot prices (and if necessary adjusted for the cost of carry) should be stationary, for if it could wander without bound, arbitrage opportunities would arise, which would be assumed to be quickly acted upon by traders such that the relationship between spot and futures prices will be brought back to equilibrium.

The notion that there should not be any lead-lag relationships between the spot and futures prices and that there should be a long-term one to one relationship between the logs of spot and futures prices can be tested using simple linear regressions and cointegration analysis. This book will now examine the results of two related papers – Tse (1995), who employs daily data on the Nikkei Stock Average (NSA) and its futures contract, and Brooks, Rew and Ritson (2001), who examine high-frequency data from the FTSE 100 stock index and index futures contract.

The data employed by Tse (1995) consists of 1,055 daily observations on NSA stock index and stock index futures values from December 1988 to April 1993. The data employed by Brooks *et al.* comprises 13,035 ten-minutely observations for all trading days in the period June 1996–May 1997, provided by FTSE International. In order to form a statistically adequate model, the variables should first be checked as to whether they can be considered stationary. The results of applying a Dickey-Fuller (DF) test to the logs of the spot and futures prices of the 10-minutely FTSE data are shown in table 7.2.

As one might anticipate, both studies conclude that the two log-price series contain a unit root, while the returns are stationary. Of course, it may be necessary to augment the tests by adding lags of the dependent variable to allow for autocorrelation in the errors (i.e. an Augmented Dickey-Fuller or ADF test). Results for such tests are not presented, since the conclusions are not altered. A statistically valid model would therefore be one in the returns. However, a formulation containing only first differences has no

Table 7.3 Estimated potentially cointegrating equation and test for cointegration for high frequency FTSE data

<i>Coefficient</i>	<i>Estimated value</i>
$\hat{\gamma}_0$	0.1345
$\hat{\gamma}_1$	0.9834
<i>DF test on residuals</i>	<i>Test statistic</i>
\hat{z}_t	-14.7303

Source: Brooks, Rew and Ritson (2001).

long-run equilibrium solution. Additionally, theory suggests that the two series should have a long-run relationship. The solution is therefore to see whether there exists a cointegrating relationship between f_t and s_t which would mean that it is valid to include levels terms along with returns in this framework. This is tested by examining whether the residuals, \hat{z}_t , of a regression of the form

$$s_t = \gamma_0 + \gamma_1 f_t + z_t \quad (7.56)$$

are stationary, using a Dickey–Fuller test, where z_t is the error term. The coefficient values for the estimated (7.56) and the DF test statistic are given in table 7.3.

Clearly, the residuals from the cointegrating regression can be considered stationary. Note also that the estimated slope coefficient in the cointegrating regression takes on a value close to unity, as predicted from the theory. It is not possible to formally test whether the true population coefficient could be one, however, since there is no way in this framework to test hypotheses about the cointegrating relationship.

The final stage in building an error correction model using the Engle–Granger 2-step approach is to use a lag of the first-stage residuals, \hat{z}_t , as the equilibrium correction term in the general equation. The overall model is

$$\Delta \log s_t = \beta_0 + \delta \hat{z}_{t-1} + \beta_1 \Delta \ln s_{t-1} + \alpha_1 \Delta \ln f_{t-1} + v_t \quad (7.57)$$

where v_t is an error term. The coefficient estimates for this model are presented in table 7.4.

Consider first the signs and significances of the coefficients (these can now be interpreted validly since all variables used in this model are stationary). $\hat{\alpha}_1$ is positive and highly significant, indicating that the futures market does indeed lead the spot market, since lagged changes in futures prices lead to a positive change in the subsequent spot price. $\hat{\beta}_1$ is positive

Table 7.4 Estimated error correction model for high frequency FTSE data

Coefficient	Estimated value	<i>t</i> -ratio
$\hat{\beta}_0$	9.6713E-06	1.6083
$\hat{\delta}$	-0.8388	-5.1298
$\hat{\beta}_1$	0.1799	19.2886
$\hat{\alpha}_1$	0.1312	20.4946

Source: Brooks, Rew and Ritson (2001).

Table 7.5 Comparison of out-of-sample forecasting accuracy

	ECM	ECM-COC	ARIMA	VAR
RMSE	0.0004382	0.0004350	0.0004531	0.0004510
MAE	0.4259	0.4255	0.4382	0.4378
% Correct direction	67.69%	68.75%	64.36%	66.80%

Source: Brooks, Rew and Ritson (2001).

and highly significant, indicating on average a positive autocorrelation in spot returns. $\hat{\delta}$, the coefficient on the error correction term, is negative and significant, indicating that if the difference between the logs of the spot and futures prices is positive in one period, the spot price will fall during the next period to restore equilibrium, and vice versa.

7.7.2 Forecasting spot returns

Both Brooks, Rew and Ritson (2001) and Tse (1995) show that it is possible to use an error correction formulation to model changes in the log of a stock index. An obvious related question to ask is whether such a model can be used to forecast the future value of the spot series for a holdout sample of data not used previously for model estimation. Both sets of researchers employ forecasts from three other models for comparison with the forecasts of the error correction model. These are an error correction model with an additional term that allows for the cost of carry, an ARMA model (with lag length chosen using an information criterion) and an unrestricted VAR model (with lag length chosen using a multivariate information criterion).

The results are evaluated by comparing their root-mean squared errors, mean absolute errors and percentage of correct direction predictions. The forecasting results from the Brooks, Rew and Ritson paper are given in table 7.5.

It can be seen from table 7.5 that the error correction models have both the lowest mean squared and mean absolute errors, and the highest proportion of correct direction predictions. There is, however, little to choose between the models, and all four have over 60% of the signs of the next returns predicted correctly.

It is clear that on statistical grounds the out-of-sample forecasting performances of the error correction models are better than those of their competitors, but this does not necessarily mean that such forecasts have any practical use. Many studies have questioned the usefulness of statistical measures of forecast accuracy as indicators of the profitability of using these forecasts in a practical trading setting (see, for example, Leitch and Tanner, 1991). Brooks, Rew and Ritson (2001) investigate this proposition directly by developing a set of trading rules based on the forecasts of the error correction model with the cost of carry term, the best statistical forecasting model. The trading period is an out-of-sample data series not used in model estimation, running from 1 May–30 May 1997. The ECM-COC model yields 10-minutely one-step-ahead forecasts. The trading strategy involves analysing the forecast for the spot return, and incorporating the decision dictated by the trading rules described below. It is assumed that the original investment is £1,000, and if the holding in the stock index is zero, the investment earns the risk-free rate. Five trading strategies are employed, and their profitabilities are compared with that obtained by passively buying and holding the index. There are of course an infinite number of strategies that could be adopted for a given set of spot return forecasts, but Brooks, Rew and Ritson use the following:

- *Liquid trading strategy* This trading strategy involves making a round-trip trade (i.e. a purchase and sale of the FTSE 100 stocks) every 10 minutes that the return is predicted to be positive by the model. If the return is predicted to be negative by the model, no trade is executed and the investment earns the risk-free rate.
- *Buy-and-hold while forecast positive strategy* This strategy allows the trader to continue holding the index if the return at the next predicted investment period is positive, rather than making a round-trip transaction for each period.
- *Filter strategy: better predicted return than average* This strategy involves purchasing the index only if the predicted returns are greater than the average positive return (there is no trade for negative returns therefore the average is only taken of the positive returns).
- *Filter strategy: better predicted return than first decile* This strategy is similar to the previous one, but rather than utilising the average as

Table 7.6 Trading profitability of the error correction model with cost of carry

Trading strategy	Terminal wealth (£)	Return(%) annualised	Terminal wealth (£) with slippage	Return(%) annualised with slippage	Number of trades
Passive investment	1040.92	4.09 {49.08}	1040.92	4.09 {49.08}	1
Liquid trading	1156.21	15.62 {187.44}	1056.38	5.64 {67.68}	583
Buy-and-Hold while forecast positive	1156.21	15.62 {187.44}	1055.77	5.58 {66.96}	383
Filter I	1144.51	14.45 {173.40}	1123.57	12.36 {148.32}	135
Filter II	1100.01	10.00 {120.00}	1046.17	4.62 {55.44}	65
Filter III	1019.82	1.98 {23.76}	1003.23	0.32 {3.84}	8

Source: Brooks, Rew and Ritson (2001).

previously, only the returns predicted to be in the top 10% of all returns are traded on.

- *Filter strategy: high arbitrary cutoff* An arbitrary filter of 0.0075% is imposed, which will result in trades only for returns that are predicted to be extremely large for a 10-minute interval.

The results from employing each of the strategies using the forecasts for the spot returns obtained from the ECM-COC model are presented in table 7.6.

The test month of May 1997 was a particularly bullish one, with a pure buy-and-hold-the-index strategy netting a return of 4%, or almost 50% on an annualised basis. Ideally, the forecasting exercise would be conducted over a much longer period than one month, and preferably over different market conditions. However, this was simply impossible due to the lack of availability of very high frequency data over a long time period. Clearly, the forecasts have some market timing ability in the sense that they seem to ensure trades that, on average, would have invested in the index when it rose, but be out of the market when it fell. The most profitable trading strategies in gross terms are those that trade on the basis of every positive spot return forecast, and all rules except the strictest filter make more money than a passive investment. The strict filter appears not to work well since it is out of the index for too long during a period when the market is rising strongly.

However, the picture of immense profitability painted thus far is somewhat misleading for two reasons: slippage time and transactions costs. First, it is unreasonable to assume that trades can be executed in the market the minute they are requested, since it may take some time to find counterparties for all the trades required to 'buy the index'. (Note, of course, that in practice, a similar returns profile to the index can be achieved with a very much smaller number of stocks.) Brooks, Rew and Ritson therefore allow for ten minutes of 'slippage time', which assumes that it takes ten minutes from when the trade order is placed to when it is executed. Second, it is unrealistic to consider gross profitability, since transactions costs in the spot market are non-negligible and the strategies examined suggested a lot of trades. Sutcliffe (1997, p. 47) suggests that total round-trip transactions costs for FTSE stocks are of the order of 1.7% of the investment.

The effect of slippage time is to make the forecasts less useful than they would otherwise have been. For example, if the spot price is forecast to rise, and it does, it may have already risen and then stopped rising by the time that the order is executed, so that the forecasts lose their market timing ability. Terminal wealth appears to fall substantially when slippage time is allowed for, with the monthly return falling by between 1.5% and 10%, depending on the trading rule.

Finally, if transactions costs are allowed for, none of the trading rules can outperform the passive investment strategy, and all in fact make substantial losses.

7.7.3 Conclusions

If the markets are frictionless and functioning efficiently, changes in the spot price of a financial asset and its corresponding futures price would be expected to be perfectly contemporaneously correlated and not to be cross-autocorrelated. Many academic studies, however, have documented that the futures market systematically 'leads' the spot market, reflecting news more quickly as a result of the fact that the stock index is not a single entity. The latter implies that:

- Some components of the index are infrequently traded, implying that the observed index value contains 'stale' component prices
- It is more expensive to transact in the spot market and hence the spot market reacts more slowly to news
- Stock market indices are recalculated only every minute so that new information takes longer to be reflected in the index.

Clearly, such spot market impediments cannot explain the inter-daily lead-lag relationships documented by Tse (1995). In any case, however, since it appears impossible to profit from these relationships, their existence is entirely consistent with the absence of arbitrage opportunities and is in accordance with modern definitions of the efficient markets hypothesis.

7.8 Testing for and estimating cointegrating systems using the Johansen technique based on VARs

Suppose that a set of g variables ($g \geq 2$) are under consideration that are $I(1)$ and which are thought may be cointegrated. A VAR with k lags containing these variables could be set up:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + u_t \quad (7.58)$$

$g \times 1 \quad g \times g \times 1 \quad g \times g \times 1 \quad g \times g \times 1 \quad g \times 1$

In order to use the Johansen test, the VAR (7.58) above needs to be turned into a vector error correction model (VECM) of the form

$$\Delta y_t = \Pi y_{t-k} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t \quad (7.59)$$

where $\Pi = (\sum_{i=1}^k \beta_i) - I_g$ and $\Gamma_i = (\sum_{j=1}^i \beta_j) - I_g$

This VAR contains g variables in first differenced form on the LHS, and $k - 1$ lags of the dependent variables (differences) on the RHS, each with a Γ coefficient matrix attached to it. In fact, the Johansen test can be affected by the lag length employed in the VECM, and so it is useful to attempt to select the lag length optimally, as outlined in chapter 6. The Johansen test centres around an examination of the Π matrix. Π can be interpreted as a long-run coefficient matrix, since in equilibrium, all the Δy_{t-i} will be zero, and setting the error terms, u_t , to their expected value of zero will leave $\Pi y_{t-k} = 0$. Notice the comparability between this set of equations and the testing equation for an ADF test, which has a first differenced term as the dependent variable, together with a lagged levels term and lagged differences on the RHS.

The test for cointegration between the y s is calculated by looking at the rank of the Π matrix via its eigenvalues.² The rank of a matrix is equal to the number of its characteristic roots (eigenvalues) that are different

² Strictly, the eigenvalues used in the test statistics are taken from rank-restricted product moment matrices and not of Π itself.

from zero (see the appendix at the end of this book for some algebra and examples). The eigenvalues, denoted λ_i are put in ascending order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_g$. If the λ s are roots, in this context they must be less than 1 in absolute value and positive, and λ_1 will be the largest (i.e. the closest to one), while λ_g will be the smallest (i.e. the closest to zero). If the variables are not cointegrated, the rank of Π will not be significantly different from zero, so $\lambda_i \approx 0 \forall i$. The test statistics actually incorporate $\ln(1 - \lambda_i)$, rather than the λ_i themselves, but still, when $\lambda_i = 0$, $\ln(1 - \lambda_i) = 0$.

Suppose now that $\text{rank}(\Pi) = 1$, then $\ln(1 - \lambda_1)$ will be negative and $\ln(1 - \lambda_i) = 0 \forall i > 1$. If the eigenvalue i is non-zero, then $\ln(1 - \lambda_i) < 0 \forall i > 1$. That is, for Π to have a rank of 1, the largest eigenvalue must be significantly non-zero, while others will not be significantly different from zero.

There are two test statistics for cointegration under the Johansen approach, which are formulated as

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^g \ln(1 - \hat{\lambda}_i) \quad (7.60)$$

and

$$\lambda_{\text{max}}(r, r + 1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (7.61)$$

where r is the number of cointegrating vectors under the null hypothesis and $\hat{\lambda}_i$ is the estimated value for the i th ordered eigenvalue from the Π matrix. Intuitively, the larger is $\hat{\lambda}_i$, the more large and negative will be $\ln(1 - \hat{\lambda}_i)$ and hence the larger will be the test statistic. Each eigenvalue will have associated with it a different cointegrating vector, which will be eigenvectors. A significantly non-zero eigenvalue indicates a significant cointegrating vector.

λ_{trace} is a joint test where the null is that the number of cointegrating vectors is less than or equal to r against an unspecified or general alternative that there are more than r . It starts with p eigenvalues, and then successively the largest is removed. $\lambda_{\text{trace}} = 0$ when all the $\lambda_i = 0$, for $i = 1, \dots, g$.

λ_{max} conducts separate tests on each eigenvalue, and has as its null hypothesis that the number of cointegrating vectors is r against an alternative of $r + 1$.

Johansen and Juselius (1990) provide critical values for the two statistics. The distribution of the test statistics is non-standard, and the critical

values depend on the value of $g - r$, the number of non-stationary components and whether constants are included in each of the equations. Intercepts can be included either in the cointegrating vectors themselves or as additional terms in the VAR. The latter is equivalent to including a trend in the data generating processes for the levels of the series. Osterwald-Lenum (1992) provides a more complete set of critical values for the Johansen test, some of which are also given in the appendix of statistical tables at the end of this book.

If the test statistic is greater than the critical value from Johansen's tables, reject the null hypothesis that there are r cointegrating vectors in favour of the alternative that there are $r + 1$ (for λ_{trace}) or more than r (for λ_{max}). The testing is conducted in a sequence and under the null, $r = 0, 1, \dots, g - 1$ so that the hypotheses for λ_{max} are

$$\begin{array}{lll} H_0 : r = 0 & \text{versus} & H_1 : 0 < r \leq g \\ H_0 : r = 1 & \text{versus} & H_1 : 1 < r \leq g \\ H_0 : r = 2 & \text{versus} & H_1 : 2 < r \leq g \\ \vdots & & \vdots \\ H_0 : r = g - 1 & \text{versus} & H_1 : r = g \end{array}$$

The first test involves a null hypothesis of no cointegrating vectors (corresponding to Π having zero rank). If this null is not rejected, it would be concluded that there are no cointegrating vectors and the testing would be completed. However, if $H_0 : r = 0$ is rejected, the null that there is one cointegrating vector (i.e. $H_0 : r = 1$) would be tested and so on. Thus the value of r is continually increased until the null is no longer rejected.

But how does this correspond to a test of the rank of the Π matrix? r is the rank of Π . Π cannot be of full rank (g) since this would correspond to the original y_t being stationary. If Π has zero rank, then by analogy to the univariate case, Δy_t depends only on Δy_{t-j} and not on y_{t-1} , so that there is no long-run relationship between the elements of y_{t-1} . Hence there is no cointegration. For $1 < \text{rank}(\Pi) < g$, there are r cointegrating vectors. Π is then defined as the product of two matrices, α and β' , of dimension $(g \times r)$ and $(r \times g)$, respectively, i.e.

$$\Pi = \alpha\beta' \tag{7.62}$$

The matrix β gives the cointegrating vectors, while α gives the amount of each cointegrating vector entering each equation of the VECM, also known as the 'adjustment parameters'.

For example, suppose that $g = 4$, so that the system contains four variables. The elements of the Π matrix would be written

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{pmatrix} \quad (7.63)$$

If $r = 1$, so that there is one cointegrating vector, then α and β will be (4×1)

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11} \quad \beta_{12} \quad \beta_{13} \quad \beta_{14}) \quad (7.64)$$

If $r = 2$, so that there are two cointegrating vectors, then α and β will be (4×2)

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \\ \alpha_{13} & \alpha_{23} \\ \alpha_{14} & \alpha_{24} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{pmatrix} \quad (7.65)$$

and so on for $r = 3, \dots$

Suppose now that $g = 4$, and $r = 1$, as in (7.64) above, so that there are four variables in the system, y_1 , y_2 , y_3 , and y_4 , that exhibit one cointegrating vector. Then Πy_{t-k} will be given by

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11} \quad \beta_{12} \quad \beta_{13} \quad \beta_{14}) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}_{t-k} \quad (7.66)$$

Equation (7.66) can also be written

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11}y_1 + \beta_{12}y_2 + \beta_{13}y_3 + \beta_{14}y_4)_{t-k} \quad (7.67)$$

Given (7.67), it is possible to write out the separate equations for each variable Δy_t . It is also common to 'normalise' on a particular variable, so that the coefficient on that variable in the cointegrating vector is one. For example, normalising on y_1 would make the cointegrating term in

the equation for Δy_1

$$\alpha_{11} \left(y_1 + \frac{\beta_{12}}{\beta_{11}} y_2 + \frac{\beta_{13}}{\beta_{11}} y_3 + \frac{\beta_{14}}{\beta_{11}} y_4 \right)_{t-k}, \text{ etc.}$$

Finally, it must be noted that the above description is not exactly how the Johansen procedure works, but is an intuitive approximation to it.

7.8.1 Hypothesis testing using Johansen

Engle–Granger did not permit the testing of hypotheses on the cointegrating relationships themselves, but the Johansen setup does permit the testing of hypotheses about the equilibrium relationships between the variables. Johansen allows a researcher to test a hypothesis about one or more coefficients in the cointegrating relationship by viewing the hypothesis as a restriction on the Π matrix. If there exist r cointegrating vectors, only these linear combinations or linear transformations of them, or combinations of the cointegrating vectors, will be stationary. In fact, the matrix of cointegrating vectors β can be multiplied by any non-singular conformable matrix to obtain a new set of cointegrating vectors.

A set of required long-run coefficient values or relationships between the coefficients does not necessarily imply that the cointegrating vectors have to be restricted. This is because any combination of cointegrating vectors is also a cointegrating vector. So it may be possible to combine the cointegrating vectors thus far obtained to provide a new one or, in general, a new set, having the required properties. The simpler and fewer are the required properties, the more likely that this recombination process (called *renormalisation*) will automatically yield cointegrating vectors with the required properties. However, as the restrictions become more numerous or involve more of the coefficients of the vectors, it will eventually become impossible to satisfy all of them by renormalisation. After this point, all other linear combinations of the variables will be non-stationary. If the restriction does not affect the model much, i.e. if the restriction is not binding, then the eigenvectors should not change much following imposition of the restriction. A test statistic to test this hypothesis is given by

$$\text{test statistic} = -T \sum_{i=1}^r [\ln(1 - \lambda_i) - \ln(1 - \lambda_i^*)] \sim \chi^2(m) \quad (7.68)$$

where λ_i^* are the characteristic roots of the restricted model, λ_i are the characteristic roots of the unrestricted model, r is the number of non-zero characteristic roots in the unrestricted model and m is the number of restrictions.

Restrictions are actually imposed by substituting them into the relevant α or β matrices as appropriate, so that tests can be conducted on either the cointegrating vectors or their loadings in each equation in the system (or both). For example, considering (7.63)–(7.65) above, it may be that theory suggests that the coefficients on the loadings of the cointegrating vector(s) in each equation should take on certain values, in which case it would be relevant to test restrictions on the elements of α (e.g. $\alpha_{11} = 1$, $\alpha_{23} = -1$, etc.). Equally, it may be of interest to examine whether only a sub-set of the variables in y_t is actually required to obtain a stationary linear combination. In that case, it would be appropriate to test restrictions of elements of β . For example, to test the hypothesis that y_4 is not necessary to form a long-run relationship, set $\beta_{14} = 0$, $\beta_{24} = 0$, etc.).

For an excellent detailed treatment of cointegration in the context of both single equation and multiple equation models, see Harris (1995). Several applications of tests for cointegration and modelling cointegrated systems in finance will now be given.

7.9 Purchasing power parity

Purchasing power parity (PPP) states that the equilibrium or long-run exchange rate between two countries is equal to the ratio of their relative price levels. Purchasing power parity implies that the real exchange rate, Q_t , is stationary. The real exchange rate can be defined as

$$Q_t = \frac{E_t P_t^*}{P_t} \quad (7.69)$$

where E_t is the nominal exchange rate in domestic currency per unit of foreign currency, P_t is the domestic price level and P_t^* is the foreign price level. Taking logarithms of (7.69) and rearranging, another way of stating the PPP relation is obtained

$$e_t - p_t + p_t^* = q_t \quad (7.70)$$

where the lower case letters in (7.70) denote logarithmic transforms of the corresponding upper case letters used in (7.69). A necessary and sufficient condition for PPP to hold is that the variables on the LHS of (7.70) – that is the log of the exchange rate between countries A and B , and the logs of the price levels in countries A and B be cointegrated with cointegrating vector $[1 \ -1 \ 1]$.

A test of this form is conducted by Chen (1995) using monthly data from Belgium, France, Germany, Italy and the Netherlands over the

Table 7.7 Cointegration tests of PPP with European data

Tests for cointegration between	$r = 0$	$r \leq 1$	$r \leq 2$	α_1	α_2
FRF-DEM	34.63*	17.10	6.26	1.33	-2.50
FRF-ITL	52.69*	15.81	5.43	2.65	-2.52
FRF-NLG	68.10*	16.37	6.42	0.58	-0.80
FRF-BEF	52.54*	26.09*	3.63	0.78	-1.15
DEM-ITL	42.59*	20.76*	4.79	5.80	-2.25
DEM-NLG	50.25*	17.79	3.28	0.12	-0.25
DEM-BEF	69.13*	27.13*	4.52	0.87	-0.52
ITL-NLG	37.51*	14.22	5.05	0.55	-0.71
ITL-BEF	69.24*	32.16*	7.15	0.73	-1.28
NLG-BEF	64.52*	21.97*	3.88	1.69	-2.17
Critical values	31.52	17.95	8.18	-	-

Notes: FRF – French franc; DEM – German mark; NLG – Dutch guilder; ITL – Italian lira; BEF – Belgian franc.

Source: Chen (1995). Reprinted with the permission of Taylor & Francis Ltd <www.tandf.co.uk>.

period April 1973 to December 1990. Pair-wise evaluations of the existence or otherwise of cointegration are examined for all combinations of these countries (10 country pairs). Since there are three variables in the system (the log exchange rate and the two log nominal price series) in each case, and that the variables in their log-levels forms are non-stationary, there can be at most two linearly independent cointegrating relationships for each country pair. The results of applying Johansen's trace test are presented in Chen's table 1, adapted and presented here as table 7.7.

As can be seen from the results, the null hypothesis of no cointegrating vectors is rejected for all country pairs, and the null of one or fewer cointegrating vectors is rejected for France-Belgium, Germany-Italy, Germany-Belgium, Italy-Belgium, Netherlands-Belgium. In no cases is the null of two or less cointegrating vectors rejected. It is therefore concluded that the PPP hypothesis is upheld and that there are either one or two cointegrating relationships between the series depending on the country pair. Estimates of α_1 and α_2 are given in the last two columns of table 7.7. PPP suggests that the estimated values of these coefficients should be 1 and -1, respectively. In most cases, the coefficient estimates are a long way from these expected values. Of course, it would be possible to impose this restriction and to test it in the Johansen framework as discussed above, but Chen does not conduct this analysis.

7.10 Cointegration between international bond markets

Often, investors will hold bonds from more than one national market in the expectation of achieving a reduction in risk via the resulting diversification. If international bond markets are very strongly correlated in the long run, diversification will be less effective than if the bond markets operated independently of one another. An important indication of the degree to which long-run diversification is available to international bond market investors is given by determining whether the markets are cointegrated. This book will now study two examples from the academic literature that consider this issue: Clare, Maras and Thomas (1995), and Mills and Mills (1991).

7.10.1 Cointegration between international bond markets: a univariate approach

Clare, Maras and Thomas (1995) use the Dickey–Fuller and Engle–Granger single-equation method to test for cointegration using a pair-wise analysis of four countries' bond market indices: US, UK, Germany and Japan. Monthly Salomon Brothers' total return government bond index data from January 1978 to April 1990 are employed. An application of the Dickey–Fuller test to the log of the indices reveals the following results (adapted from their table 1), given in table 7.8.

Neither the critical values, nor a statement of whether a constant or trend are included in the test regressions, are offered in the paper. Nevertheless, the results are clear. Recall that the null hypothesis of a unit root is rejected if the test statistic is smaller (more negative) than the critical value. For samples of the size given here, the 5% critical value would

Table 7.8 DF tests for international bond indices

<i>Panel A: test on log-index for country</i>	DF Statistic
Germany	−0.395
Japan	−0.799
UK	−0.884
US	0.174
<i>Panel B: test on log-returns for country</i>	
Germany	−10.37
Japan	−10.11
UK	−10.56
US	−10.64

Source: Clare, Maras and Thomas (1995). Reprinted with the permission of Blackwell Publishers.

Table 7.9 Cointegration tests for pairs of international bond indices

Test	UK– Germany	UK– Japan	UK–US	Germany– Japan	Germany– US	Japan– US	5% Critical value
CRDW	0.189	0.197	0.097	0.230	0.169	0.139	0.386
DF	2.970	2.770	2.020	3.180	2.160	2.160	3.370
ADF	3.160	2.900	1.800	3.360	1.640	1.890	3.170

Source: Clare, Maras and Thomas (1995). Reprinted with the permission of Blackwell Publishers.

be somewhere between -1.95 and -3.50 . It is thus demonstrated quite conclusively that the logarithms of the indices are non-stationary, while taking the first difference of the logs (that is, constructing the returns) induces stationarity.

Given that all logs of the indices in all four cases are shown to be $I(1)$, the next stage in the analysis is to test for cointegration by forming a potentially cointegrating regression and testing its residuals for non-stationarity. Clare, Maras and Thomas use regressions of the form

$$B_i = \alpha_0 + \alpha_1 B_j + u \quad (7.71)$$

with time subscripts suppressed and where B_i and B_j represent the log-bond indices for any two countries i and j . The results are presented in their tables 3 and 4, which are combined into table 7.9 here. They offer results from applying 7 different tests, while we present results only for the Cointegrating Regression Durbin Watson (CRDW), Dickey–Fuller and Augmented Dickey–Fuller tests (although the lag lengths for the latter are not given) are presented here.

In this case, the null hypothesis of a unit root in the residuals from regression (7.71) cannot be rejected. The conclusion is therefore that there is no cointegration between any pair of bond indices in this sample.

7.10.2 Cointegration between international bond markets: a multivariate approach

Mills and Mills (1991) also consider the issue of cointegration or non-cointegration between the same four international bond markets. However, unlike Clare, Maras and Thomas, who use bond price indices, Mills and Mills employ daily closing observations on the redemption yields. The latter's sample period runs from 1 April 1986 to 29 December 1989, giving 960 observations. They employ a Dickey–Fuller-type regression procedure to test the individual series for non-stationarity and conclude that all four yields series are $I(1)$.

Table 7.10 Johansen tests for cointegration between international bond yields

<i>r</i> (number of cointegrating vectors under the null hypothesis)	Test statistic	Critical values	
		10%	5%
0	22.06	35.6	38.6
1	10.58	21.2	23.8
2	2.52	10.3	12.0
3	0.12	2.9	4.2

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.

The Johansen systems procedure is then used to test for cointegration between the series. Unlike the Clare, Maras and Thomas paper, Mills and Mills (1991) consider all four indices together rather than investigating them in a pair-wise fashion. Therefore, since there are four variables in the system (the redemption yield for each country), i.e. $g = 4$, there can be at most three linearly independent cointegrating vectors, i.e., $r \leq 3$. The trace statistic is employed, and it takes the form

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^g \ln(1 - \hat{\lambda}_i) \tag{7.72}$$

where λ_i are the ordered eigenvalues. The results are presented in their table 2, which is modified slightly here, and presented in table 7.10.

Looking at the first row under the heading, it can be seen that the test statistic is smaller than the critical value, so the null hypothesis that $r = 0$ cannot be rejected, even at the 10% level. It is thus not necessary to look at the remaining rows of the table. Hence, reassuringly, the conclusion from this analysis is the same as that of Clare, Maras and Thomas - i.e. that there are no cointegrating vectors.

Given that there are no linear combinations of the yields that are stationary, and therefore that there is no error correction representation, Mills and Mills then continue to estimate a VAR for the first differences of the yields. The VAR is of the form

$$\Delta X_t = \sum_{i=1}^k \Gamma_i \Delta X_{t-i} + v_t \tag{7.73}$$

where:

$$X_t = \begin{bmatrix} X(US)_t \\ X(UK)_t \\ X(WG)_t \\ X(JAP)_t \end{bmatrix}, \Gamma_i = \begin{bmatrix} \Gamma_{11i} & \Gamma_{12i} & \Gamma_{13i} & \Gamma_{14i} \\ \Gamma_{21i} & \Gamma_{22i} & \Gamma_{23i} & \Gamma_{24i} \\ \Gamma_{31i} & \Gamma_{32i} & \Gamma_{33i} & \Gamma_{34i} \\ \Gamma_{41i} & \Gamma_{42i} & \Gamma_{43i} & \Gamma_{44i} \end{bmatrix}, v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{bmatrix}$$

Table 7.11 Variance decompositions for VAR of international bond yields

Explaining movements in	Days ahead	Explained by movements in			
		US	UK	Germany	Japan
US	1	95.6	2.4	1.7	0.3
	5	94.2	2.8	2.3	0.7
	10	92.9	3.1	2.9	1.1
	20	92.8	3.2	2.9	1.1
UK	1	0.0	98.3	0.0	1.7
	5	1.7	96.2	0.2	1.9
	10	2.2	94.6	0.9	2.3
	20	2.2	94.6	0.9	2.3
Germany	1	0.0	3.4	94.6	2.0
	5	6.6	6.6	84.8	3.0
	10	8.3	6.5	82.9	3.6
	20	8.4	6.5	82.7	3.7
Japan	1	0.0	0.0	1.4	100.0
	5	1.3	1.4	1.1	96.2
	10	1.5	2.1	1.8	94.6
	20	1.6	2.2	1.9	94.2

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.

They set k , the number of lags of each change in the yield in each regression, to 8, arguing that likelihood ratio tests rejected the possibility of smaller numbers of lags. Unfortunately, and as one may anticipate for a regression of daily yield changes, the R^2 values for the VAR equations are low, ranging from 0.04 for the US to 0.17 for Germany. Variance decompositions and impulse responses are calculated for the estimated VAR. Two orderings of the variables are employed: one based on a previous study and one based on the chronology of the opening (and closing) of the financial markets considered: Japan \rightarrow Germany \rightarrow UK \rightarrow US. Only results for the latter, adapted from tables 4 and 5 of Mills and Mills (1991), are presented here. The variance decompositions and impulse responses for the VARs are given in tables 7.11 and 7.12, respectively.

As one may expect from the low R^2 of the VAR equations, and the lack of cointegration, the bond markets seem very independent of one another. The variance decompositions, which show the proportion of the movements in the dependent variables that are due to their 'own' shocks, versus shocks to the other variables, seem to suggest that the US, UK and Japanese markets are to a certain extent exogenous in this system. That is, little of the movement of the US, UK or Japanese series can be

Table 7.12 Impulse responses for VAR of international bond yields

Days after shock	Response of US to innovations in			
	US	UK	Germany	Japan
0	0.98	0.00	0.00	0.00
1	0.06	0.01	-0.10	0.05
2	-0.02	0.02	-0.14	0.07
3	0.09	-0.04	0.09	0.08
4	-0.02	-0.03	0.02	0.09
10	-0.03	-0.01	-0.02	-0.01
20	0.00	0.00	-0.10	-0.01

Days after shock	Response of UK to innovations in			
	US	UK	Germany	Japan
0	0.19	0.97	0.00	0.00
1	0.16	0.07	0.01	-0.06
2	-0.01	-0.01	-0.05	0.09
3	0.06	0.04	0.06	0.05
4	0.05	-0.01	0.02	0.07
10	0.01	0.01	-0.04	-0.01
20	0.00	0.00	-0.01	0.00

Days after shock	Response of Germany to innovations in			
	US	UK	Germany	Japan
0	0.07	0.06	0.95	0.00
1	0.13	0.05	0.11	0.02
2	0.04	0.03	0.00	0.00
3	0.02	0.00	0.00	0.01
4	0.01	0.00	0.00	0.09
10	0.01	0.01	-0.01	0.02
20	0.00	0.00	0.00	0.00

Days after shock	Response of Japan to innovations in			
	US	UK	Germany	Japan
0	0.03	0.05	0.12	0.97
1	0.06	0.02	0.07	0.04
2	0.02	0.02	0.00	0.21
3	0.01	0.02	0.06	0.07
4	0.02	0.03	0.07	0.06
10	0.01	0.01	0.01	0.04
20	0.00	0.00	0.00	0.01

Source: Mills and Mills (1991). Reprinted with the permission of Blackwell Publishers.

explained by movements other than their own bond yields. In the German case, however, after 20 days, only 83% of movements in the German yield are explained by German shocks. The German yield seems particularly influenced by US (8.4% after 20 days) and UK (6.5% after 20 days) shocks. It also seems that Japanese shocks have the least influence on the bond yields of other markets.

A similar pattern emerges from the impulse response functions, which show the effect of a unit shock applied separately to the error of each equation of the VAR. The markets appear relatively independent of one another, and also informationally efficient in the sense that shocks work through the system very quickly. There is never a response of more than 10% to shocks in any series three days after they have happened; in most cases, the shocks have worked through the system in two days. Such a result implies that the possibility of making excess returns by trading in one market on the basis of 'old news' from another appears very unlikely.

7.10.3 *Cointegration in international bond markets: conclusions*

A single set of conclusions can be drawn from both of these papers. Both approaches have suggested that international bond markets are not cointegrated. This implies that investors can gain substantial diversification benefits. This is in contrast to results reported for other markets, such as foreign exchange (Baillie and Bollerslev, 1989), commodities (Baillie, 1989), and equities (Taylor and Tonks, 1989). Clare, Maras and Thomas (1995) suggest that the lack of long-term integration between the markets may be due to 'institutional idiosyncrasies', such as heterogeneous maturity and taxation structures, and differing investment cultures, issuance patterns and macroeconomic policies between countries, which imply that the markets operate largely independently of one another.

7.11 **Testing the expectations hypothesis of the term structure of interest rates**

The following notation replicates that employed by Campbell and Shiller (1991) in their seminal paper. The single, linear expectations theory of the term structure used to represent the expectations hypothesis (hereafter EH), defines a relationship between an n -period interest rate or yield, denoted $R_t^{(n)}$, and an m -period interest rate, denoted $R_t^{(m)}$, where $n > m$. Hence $R_t^{(n)}$ is the interest rate or yield on a longer-term instrument relative to a shorter-term interest rate or yield, $R_t^{(m)}$. More precisely, the EH states

that the expected return from investing in an n -period rate will equal the expected return from investing in m -period rates up to $n - m$ periods in the future plus a constant risk-premium, c , which can be expressed as

$$R_t^{(n)} = \frac{1}{q} \sum_{i=0}^{q-1} E_t R_{t+mi}^{(m)} + c \quad (7.74)$$

where $q = n/m$. Consequently, the longer-term interest rate, $R_t^{(n)}$, can be expressed as a weighted-average of current and expected shorter-term interest rates, $R_t^{(m)}$, plus a constant risk premium, c . If (7.74) is considered, it can be seen that by subtracting $R_t^{(m)}$ from both sides of the relationship we have

$$R_t^{(n)} - R_t^{(m)} = \frac{1}{q} \sum_{i=0}^{q-1} \sum_{j=1}^{j=i} E_t [\Delta^{(m)} R_{t+jm}^{(m)}] + c \quad (7.75)$$

Examination of (7.75) generates some interesting restrictions. If the interest rates under analysis, say $R_t^{(n)}$ and $R_t^{(m)}$, are I(1) series, then, by definition, $\Delta R_t^{(n)}$ and $\Delta R_t^{(m)}$ will be stationary series. There is a general acceptance that interest rates, Treasury Bill yields, etc. are well described as I(1) processes and this can be seen in Campbell and Shiller (1988) and Stock and Watson (1988). Further, since c is a constant then it is by definition a stationary series. Consequently, if the EH is to hold, given that c and $\Delta R_t^{(m)}$ are I(0) implying that the RHS of (7.75) is stationary, then $R_t^{(n)} - R_t^{(m)}$ must by definition be stationary, otherwise we will have an inconsistency in the order of integration between the RHS and LHS of the relationship. $R_t^{(n)} - R_t^{(m)}$ is commonly known as the *spread* between the n -period and m -period rates, denoted $S_t^{(n,m)}$, which in turn gives an indication of the slope of the term structure. Consequently, it follows that if the EH is to hold, then the spread will be found to be stationary and therefore $R_t^{(n)}$ and $R_t^{(m)}$ will cointegrate with a cointegrating vector $(1, -1)$ for $[R_t^{(n)}, R_t^{(m)}]$. Therefore, the integrated process driving each of the two rates is common to both and hence it can be said that the rates have a common stochastic trend. As a result, since the EH predicts that each interest rate series will cointegrate with the one-period interest rate, it must be true that the stochastic process driving all the rates is the same as that driving the one-period rate, i.e. any combination of rates formed to create a spread should be found to cointegrate with a cointegrating vector $(1, -1)$.

Many examinations of the expectations hypothesis of the term structure have been conducted in the literature, and still no overall consensus appears to have emerged concerning its validity. One such study that tested

Table 7.13 Tests of the expectations hypothesis using the US zero coupon yield curve with monthly data

Sample period	Interest rates included	Lag length of VAR	Hypothesis is	λ_{max}	λ_{trace}
1952M1–1978M12	$X_t = [R_t R_t^{(6)}]'$	2	$r = 0$	47.54***	49.82***
			$r \leq 1$	2.28	2.28
1952M1–1987M2	$X_t = [R_t R_t^{(120)}]'$	2	$r = 0$	40.66***	43.73***
			$r \leq 1$	3.07	3.07
1952M1–1987M2	$X_t = [R_t R_t^{(60)} R_t^{(120)}]'$	2	$r = 0$	40.13***	42.63***
1973M5–1987M2	$X_t = [R_t R_t^{(60)} R_t^{(120)} R_t^{(180)} R_t^{(240)}]'$	7	$r = 0$	34.78***	75.50***
			$r \leq 1$	23.31*	40.72
			$r \leq 2$	11.94	17.41
			$r \leq 3$	3.80	5.47
			$r \leq 4$	1.66	1.66

Notes: *, ** and *** denote significance at the 20%, 10% and 5% levels, respectively; r is the number of cointegrating vectors under the null hypothesis.

Source: Shea (1992). Reprinted with the permission of American Statistical Association. All rights reserved.

the expectations hypothesis using a standard data-set due to McCulloch (1987) was conducted by Shea (1992). The data comprises a zero coupon term structure for various maturities from 1 month to 25 years, covering the period January 1952–February 1987. Various techniques are employed in Shea's paper, while only his application of the Johansen technique is discussed here. A vector X_t containing the interest rate at each of the maturities is constructed

$$X_t = [R_t R_t^{(2)} \dots R_t^{(n)}]' \quad (7.76)$$

where R_t denotes the spot interest rate. It is argued that each of the elements of this vector is non-stationary, and hence the Johansen approach is used to model the system of interest rates and to test for cointegration between the rates. Both the λ_{max} and λ_{trace} statistics are employed, corresponding to the use of the maximum eigenvalue and the cumulated eigenvalues, respectively. Shea tests for cointegration between various combinations of the interest rates, measured as returns to maturity. A selection of Shea's results is presented in table 7.13.

The results below, together with the other results presented by Shea, seem to suggest that the interest rates at different maturities are typically cointegrated, usually with one cointegrating vector. As one may have

expected, the cointegration becomes weaker in the cases where the analysis involves rates a long way apart on the maturity spectrum. However, cointegration between the rates is a necessary but not sufficient condition for the expectations hypothesis of the term structure to be vindicated by the data. Validity of the expectations hypothesis also requires that any combination of rates formed to create a spread should be found to cointegrate with a cointegrating vector $(1, -1)$. When comparable restrictions are placed on the β estimates associated with the cointegrating vectors, they are typically rejected, suggesting only limited support for the expectations hypothesis.

7.12 Testing for cointegration and modelling cointegrated systems using EViews

The S&P500 spot and futures series that were discussed in chapters 2 and 3 will now be examined for cointegration using EViews. If the two series are cointegrated, this means that the spot and futures prices have a long-term relationship, which prevents them from wandering apart without bound. To test for cointegration using the Engle–Granger approach, the residuals of a regression of the spot price on the futures price are examined.³ **Create two new variables**, for the log of the spot series and the log of the futures series, and call them 'lspot' and 'lfutures' respectively. Then generate a new equation object and run the regression:

LSPOT C LFUTURES

Note again that it is not valid to examine anything other than the coefficient values in this regression. The residuals of this regression are found in the object called RESID. First, if you click on the **Resids** tab, you will see a plot of the levels of the residuals (blue line), which looks much more like a stationary series than the original spot series (the red line corresponding to the actual values of y) looks. The plot should appear as in screenshot 7.2.

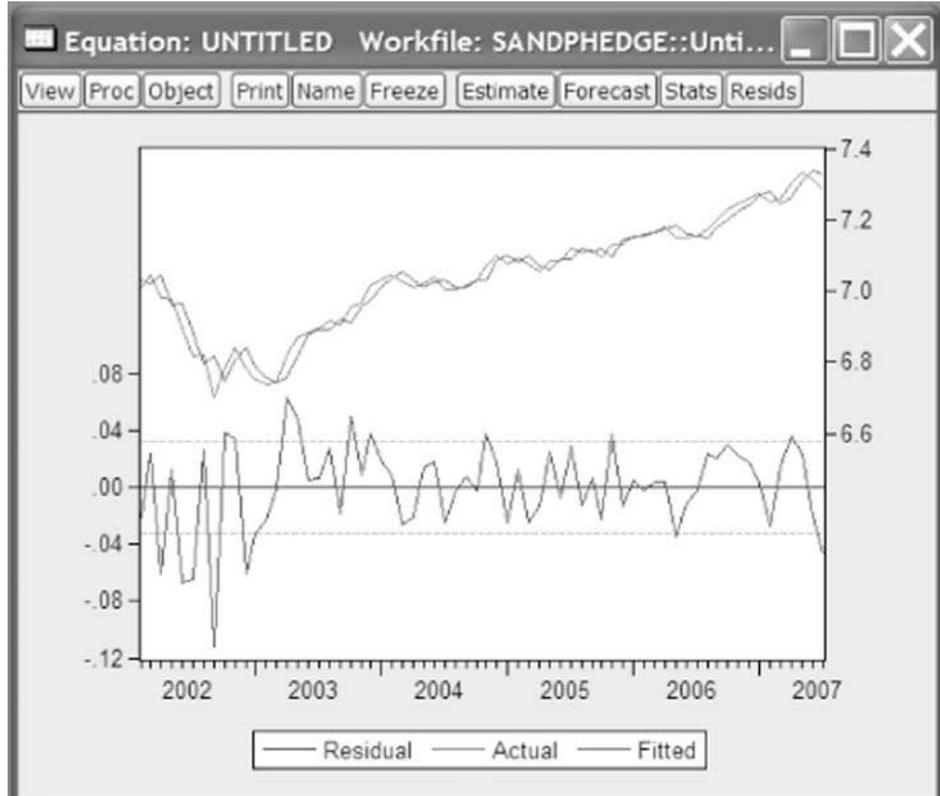
Generate a new series that will keep these residuals in an object for later use:

STATRESIDS = RESID

³ Note that it is common to run a regression of the log of the spot price on the log of the futures rather than a regression in levels; the main reason for using logarithms is that the differences of the logs are returns, whereas this is not true for the levels.

Screenshot 7.2

Actual, Fitted and Residual plot to check for stationarity



This is required since every time a regression is run, the RESID object is updated (overwritten) to contain the residuals of the most recently conducted regression. **Perform the ADF Test** on the residual series STATRESIDS. Assuming again that up to 12 lags are permitted, and that a constant but not a trend are employed in a regression on the levels of the series, the results are:

Null Hypothesis: STATRESIDS has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.050542	0.0000
Test critical values:		
1% level	-3.534868	
5% level	-2.906923	
10% level	-2.591006	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(STATRESIDS)

Method: Least Squares

Date: 09/06/07 Time: 10:55

Sample (adjusted): 2002M03 2007M07

Included observations: 65 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
STATRESIDS(-1)	-1.027830	0.127672	-8.050542	0.000000
C	0.000352	0.003976	0.088500	0.929800
R-squared	0.507086	Mean dependent var		-0.000387
Adjusted R-squared	0.499262	S.D. dependent var		0.045283
S.E. of regression	0.032044	Akaike info criterion		-4.013146
Sum squared resid	0.064688	Schwarz criterion		-3.946241
Log likelihood	132.4272	Hannan-Quinn criter.		-3.986748
F-statistic	64.81123	Durbin-Watson stat		1.935995
Prob(F-statistic)	0.000000			

Since the test statistic (-8.05) is more negative than the critical values, even at the 1% level, the null hypothesis of a unit root in the test regression residuals is strongly rejected. We would thus conclude that the two series are cointegrated. This means that an error correction model (ECM) can be estimated, as there is a linear combination of the spot and futures prices that would be stationary. The ECM would be the appropriate model rather than a model in pure first difference form because it would enable us to capture the long-run relationship between the series as well as the short-run one. We could now estimate an error correction model by running the regression⁴

```
rspot c rfutures statresids(-1)
```

Although the Engle-Granger approach is evidently very easy to use, as outlined above, one of its major drawbacks is that it can estimate only up to one cointegrating relationship between the variables. In the spot-futures example, there can be at most one cointegrating relationship since there are only two variables in the system. But in other situations, if there are more variables, there could potentially be more than one linearly independent cointegrating relationship. Thus, it is appropriate instead to examine the issue of cointegration within the Johansen VAR framework.

⁴ If you run this regression, you will see that the estimated ECM results from this example are not entirely plausible but may have resulted from the relatively short sample period employed!

The application we will now examine centres on whether the yields on treasury bills of different maturities are cointegrated. **Re-open the 'macro.wf1' workfile** that was used in chapter 3. There are six interest rate series corresponding to three and six months, and one, three, five and ten years. Each series has a name in the file starting with the letters 'ustb'. The first step in any cointegration analysis is to ensure that the variables are all non-stationary in their levels form, so **confirm that this is the case** for each of the six series, by running a unit root test on each one.

Next, to run the cointegration test, **highlight the six series** and then click **Quick/Group Statistics/Cointegration Test**. A box should then appear with the names of the six series in it. Click OK, and then the following list of options will appear (screenshot 7.3).

Screenshot 7.3

Johansen
cointegration test

Johansen Cointegration Test

Cointegration Test Specification

Deterministic trend assumption of test

Assume no deterministic trend in data:

1) No intercept or trend in CE or test VAR

2) Intercept (no trend) in CE - no intercept in VAR

Allow for linear deterministic trend in data:

3) Intercept (no trend) in CE and test VAR

4) Intercept and trend in CE - no trend in VAR

Allow for quadratic deterministic trend in

5) Intercept and trend in CE - linear trend in VAR

Summary:

6) Summarize all 5 sets of assumptions

* Critical values may not be valid with exogenous variables; do not include C or Trend.

Exog variables**

Lag intervals

14

Lag spec for differenced endogenous

Critical Values

MHM

Size 0.05

Osterwald-Lenur

OK Cancel

The differences between models 1 to 6 centre on whether an intercept or a trend or both are included in the potentially cointegrating relationship and/or the VAR. It is usually a good idea to examine the sensitivity of the result to the type of specification used, so select **Option 6** which will do this and click OK. The results appear as in the following table

Date: 09/06/07 Time: 11:43
 Sample: 1986M03 2007M04
 Included observations: 249
 Series: USTB10Y USTB1Y USTB3M USTB3Y USTB5Y USTB6M
 Lags interval: 1 to 4

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	4	3	4	4	6
Max-Eig	3	2	2	1	1

*Critical values based on MacKinnon-Haug-Michelis (1999)

Information Criteria by Rank and Model

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or No. of CEs	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend

Log Likelihood by Rank (rows) and Model (columns)

0	1667.058	1667.058	1667.807	1667.807	1668.036
1	1690.466	1691.363	1691.975	1692.170	1692.369
2	1707.508	1709.254	1709.789	1710.177	1710.363
3	1719.820	1722.473	1722.932	1726.801	1726.981
4	1728.513	1731.269	1731.728	1738.760	1738.905
5	1733.904	1737.304	1737.588	1746.100	1746.238
6	1734.344	1738.096	1738.096	1751.143	1751.143

Akaike Information Criteria by Rank (rows) and Model (columns)

0	-12.23340	-12.23340	-12.19122	-12.19122	-12.14487
1	-12.32503	-12.32420	-12.28896	-12.28249	-12.24393
2	-12.36552	-12.36349	-12.33566	-12.32271	-12.29208
3	-12.36803*	-12.36524	-12.34484	-12.35182	-12.32916
4	-12.34147	-12.33148	-12.31910	-12.34345	-12.32856
5	-12.28838	-12.27553	-12.26979	-12.29799	-12.29107
6	-12.19553	-12.17748	-12.17748	-12.23408	-12.23408

Schwarz Criteria by Rank (rows) and Model (columns)

0	-10.19921*	-10.19921*	-10.07227	-10.07227	-9.941161
1	-10.12132	-10.10637	-10.00049	-9.979903	-9.870707
2	-9.992303	-9.962013	-9.877676	-9.836474	-9.749338
3	-9.825294	-9.780129	-9.717344	-9.681945	-9.616911
4	-9.629218	-9.562721	-9.522087	-9.489935	-9.446787
5	-9.406616	-9.323131	-9.303259	-9.260836	-9.239781
6	-9.144249	-9.041435	-9.041435	-9.013282	-9.013282

The results across the six types of model and the type of test (the ‘trace’ or ‘max’ statistics) are a little mixed concerning the number of cointegrating vectors (the top panel) but they do at least all suggest that the series are cointegrated – in other words, all specifications suggest that there is at least one cointegrating vector. The following three panels all provide information that could be used to determine the appropriate lag length for the VAR. The values of the log-likelihood function could be used to run tests of whether a VAR of a given order could be restricted to a VAR of lower order; AIC and SBIC values are provided in the final two panels. Fortunately, which ever model is used concerning whether intercepts and/or trends are incorporated, AIC selects a VAR with 3 lags and SBIC a VAR with 0 lags. Note that the difference in optimal model order could be attributed to the relatively small sample size available with this monthly sample compared with the number of observations that would have been available were daily data used, implying that the penalty term in SBIC is more severe on extra parameters in this case.

So, in order to see the estimated models, click **View/Cointegration Test** and select **Option 3** (Intercept (no trend) in CE and test VAR), changing the ‘Lag Intervals’ to 1 3, and clicking **OK**. EViews produces a very large quantity of output, as shown in the following table.⁵

Date: 09/06/07 Time: 13:20
 Sample (adjusted): 1986M07 2007M04
 Included observations: 250 after adjustments
 Trend assumption: Linear deterministic trend
 Series: USTB10Y USTB1Y USTB3M USTB3Y USTB5Y USTB6M
 Lags interval (in first differences): 1 to 3

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace 0.05		
		Statistic	Critical Value	Prob.**
None*	0.185263	158.6048	95.75366	0.0000
At most 1*	0.140313	107.3823	69.81889	0.0000
At most 2*	0.136686	69.58558	47.85613	0.0001
At most 3*	0.082784	32.84123	29.79707	0.0216
At most 4	0.039342	11.23816	15.49471	0.1973
At most 5	0.004804	1.203994	3.841466	0.2725

Trace test indicates 4 cointegrating eqn(s) at the 0.05 level

*denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

⁵ Estimated cointegrating vectors and loadings are provided by EViews for 2–5 cointegrating vectors as well, but these are not shown to preserve space.

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None*	0.185263	51.22249	40.07757	0.0019
At most 1*	0.140313	37.79673	33.87687	0.0161
At most 2*	0.136686	36.74434	27.58434	0.0025
At most 3*	0.082784	21.60308	21.13162	0.0429
At most 4	0.039342	10.03416	14.26460	0.2097
At most 5	0.004804	1.203994	3.841466	0.2725

Max-eigenvalue test indicates 4 cointegrating eqn(s) at the 0.05 level

*denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by $b'S11*b = I$):

USTB10Y	USTB1Y	USTB3M	USTB3Y	USTB5Y	USTB6M
2.775295	-6.449084	-14.79360	1.880919	-4.947415	21.32095
2.879835	0.532476	-0.398215	-7.247578	0.964089	3.797348
6.676821	-15.83409	1.422340	21.39804	-20.73661	6.834275
-7.351465	-9.144157	-3.832074	-6.082384	15.06649	11.51678
1.301354	0.034196	3.251778	8.469627	-8.131063	-4.915350
-2.919091	1.146874	0.663058	-1.465376	3.350202	-1.422377

Unrestricted Adjustment Coefficients (alpha):

D(USTB10Y)	0.030774	0.009498	0.038434	-0.042215	0.004975	0.012630
D(USTB1Y)	0.047301	-0.013791	0.037992	-0.050510	-0.012189	0.004599
D(USTB3M)	0.063889	-0.028097	0.004484	-0.031763	-0.003831	0.001249
D(USTB3Y)	0.042465	0.014245	0.035935	-0.062930	-0.006964	0.010137
D(USTB5Y)	0.039796	0.018413	0.041033	-0.058324	0.001649	0.010563
D(USTB6M)	0.042840	-0.029492	0.018767	-0.046406	-0.006399	0.002473

1 Cointegrating Equation(s): Log likelihood 1656.437

Normalized cointegrating coefficients (standard error in parentheses)

USTB10Y	USTB1Y	USTB3M	USTB3Y	USTB5Y	USTB6M
1.000000	-2.323747	-5.330461	0.677737	-1.782662	7.682407
	(0.93269)	(0.78256)	(0.92410)	(0.56663)	(1.28762)

 Adjustment coefficients (standard error in parentheses)

D(USTB10Y)	0.085407 (0.04875)
D(USTB1Y)	0.131273 (0.04510)
D(USTB3M)	0.177312 (0.03501)
D(USTB3Y)	0.117854 (0.05468)
D(USTB5Y)	0.110446 (0.05369)
D(USTB6M)	0.118894 (0.03889)

 2 Cointegrating Equation(s): Log likelihood 1675.335

Normalized cointegrating coefficients (standard error in parentheses)

USTB10Y	USTB1Y	USTB3M	USTB3Y	USTB5Y	USTB6M
1.000000	0.000000	-0.520964 (0.76929)	-2.281223 (0.77005)	0.178708 (0.53441)	1.787640 (0.97474)
0.000000	1.000000	2.069717 (0.43972)	-1.273357 (0.44016)	0.844055 (0.30546)	-2.536751 (0.55716)

Adjustment coefficients (standard error in parentheses)

D(USTB10Y)	0.112760 (0.07021)	-0.193408 (0.11360)
D(USTB1Y)	0.091558 (0.06490)	-0.312389 (0.10500)
D(USTB3M)	0.096396 (0.04991)	-0.426988 (0.08076)
D(USTB3Y)	0.158877 (0.07871)	-0.266278 (0.12735)
D(USTB5Y)	0.163472 (0.07722)	-0.246844 (0.12494)
D(USTB6M)	0.033962 (0.05551)	-0.291983 (0.08981)

Note: Table truncated.

The first two panels of the table show the results for the λ_{trace} and λ_{max} statistics respectively. The second column in each case presents the ordered eigenvalues, the third column the test statistic, the fourth column the critical value and the final column the p -value. Examining the trace test, if we look at the first row after the headers, the statistic of 158.6048 considerably exceeds the critical value (of 95) and so the null of no cointegrating vectors is rejected. If we then move to the next row, the test statistic (107.3823) again exceeds the critical value so that the null of at most one cointegrating vector is also rejected. This continues, until we do not reject the null hypothesis of at most four cointegrating vectors at the 5% level, and this is the conclusion. The *max* test, shown in the second panel, confirms this result.

The unrestricted coefficient values are the estimated values of coefficients in the cointegrating vector, and these are presented in the third panel. However, it is sometimes useful to normalise the coefficient values to set the coefficient value on one of them to unity, as would be the case in the cointegrating regression under the Engle–Granger approach. The normalisation will be done by EViews with respect to the first variable given in the variable list (i.e. which ever variable you listed first in the system will by default be given a coefficient of 1 in the normalised cointegrating vector). Panel 6 of the table presents the estimates if there were only one cointegrating vector, which has been normalised so that the coefficient on the ten-year bond yield is unity. The adjustment coefficients, or loadings in each regression (i.e. the ‘amount of the cointegrating vector’ in each equation), are also given in this panel. In the next panel, the same format is used (i.e. the normalised cointegrating vectors are presented and then the adjustment parameters) but under the assumption that there are two cointegrating vectors, and this proceeds until the situation where there are five cointegrating vectors, the maximum number possible for a system containing six variables.

In order to see the whole VECM model, select **Proc/Make Vector Autoregression...** Starting on the default ‘Basics’ tab, in ‘VAR type’, select **Vector Error Correction**, and in the ‘Lag Intervals for D(Endogenous):’ box, type 1 3. Then click on the **cointegration** tab and leave the default as 1 cointegrating vector for simplicity in the ‘Rank’ box and option 3 to have an intercept but no trend in the cointegrating equation and the VAR. When OK is clicked, the output for the entire VECM will be seen.

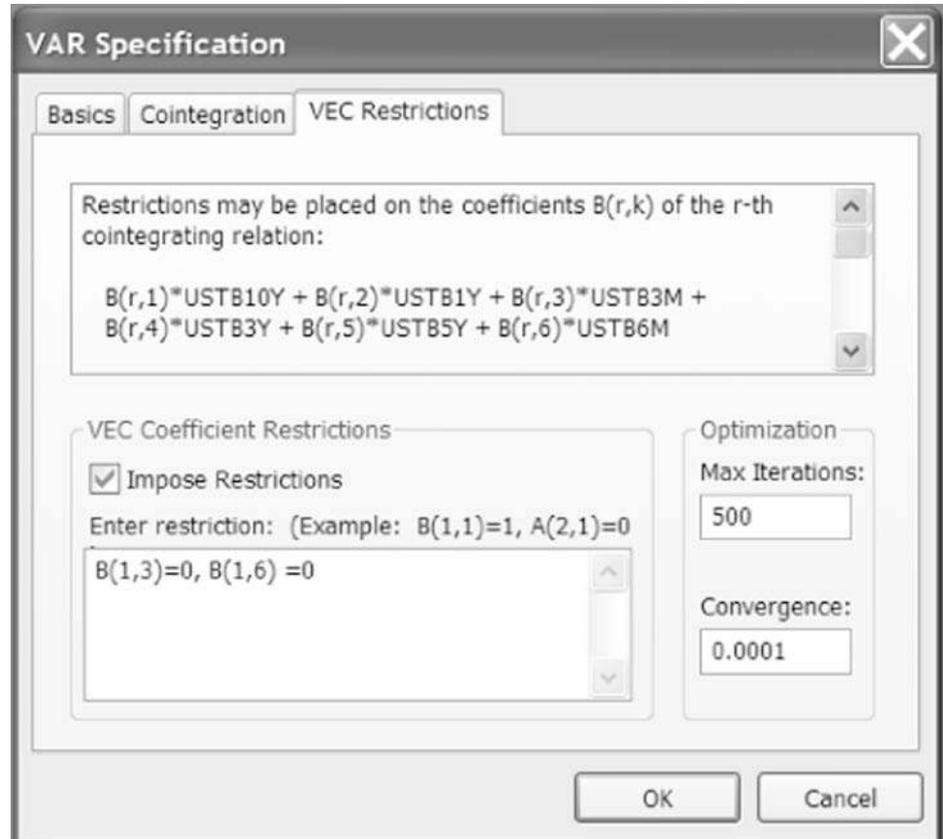
It is sometimes of interest to test hypotheses about either the parameters in the cointegrating vector or their loadings in the VECM. To do this

from the 'Vector Error Correction Estimates' screen, click the **Estimate** button and click on the **VEC Restrictions** tab.

In EViews, restrictions concerning the cointegrating relationships embodied in β are denoted by $B(i,j)$, where $B(i,j)$ represents the j th coefficient in the i th cointegrating relationship (screenshot 7.4).

Screenshot 7.4

VAR specification for Johansen tests



In this case, we are allowing for only one cointegrating relationship, so suppose that we want to test the hypothesis that the three-month and six-month yields do not appear in the cointegrating equation. We could test this by specifying the restriction that their parameters are zero, which in EViews terminology would be achieved by writing $B(1,3) = 0, B(1,6) = 0$ in the 'VEC Coefficient Restrictions' box and clicking OK. EViews will then show the value of the test statistic, followed by the restricted cointegrating vector and the VECM. To preserve space, only the test statistic and restricted cointegrating vector are shown in the following table.

In this case, there are two restrictions, so that the test statistic follows a χ^2 distribution with 2 degrees of freedom. In this case, the p -value for the test is 0.001, and so the restrictions are not supported by the data and

Vector Error Correction Estimates

Date: 09/06/07 Time: 14:04

Sample (adjusted): 1986M07 2007M04

Included observations: 250 after adjustments

Standard errors in () & t-statistics in []

Cointegration Restrictions:	
B(1,3) = 0, B(1,6) = 0	
Convergence achieved after 38 iterations.	
Not all cointegrating vectors are identified	
LR test for binding restrictions (rank = 1):	
Chi-square(2)	13.50308
Probability	0.001169
Cointegrating Eq:	CointEq1
USTB10Y(-1)	-0.088263
USTB1Y(-1)	-2.365941
USTB3M(-1)	0.000000
USTB3Y(-1)	5.381347
USTB5Y(-1)	-3.149580
USTB6M(-1)	0.000000
C	0.923034

Note: Table truncated

we would conclude that the cointegrating relationship must also include the short end of the yield curve.

When performing hypothesis tests concerning the adjustment coefficients (i.e. the loadings in each equation), the restrictions are denoted by $A(i, j)$, which is the coefficient on the cointegrating vector for the i th variable in the j th cointegrating relation. For example, $A(2, 1) = 0$ would test the null that the equation for the second variable in the order that they were listed in the original specification (USTB1Y in this case) does not include the first cointegrating vector, and so on. Examining some restrictions of this type is left as an exercise.

Key concepts

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

- non-stationary
- unit root
- augmented Dickey-Fuller test
- error correction model
- Johansen technique
- eigenvalues
- explosive process
- spurious regression
- cointegration
- Engle-Granger 2-step approach
- vector error correction model

Review questions

1. (a) What kinds of variables are likely to be non-stationary? How can such variables be made stationary?
- (b) Why is it in general important to test for non-stationarity in time series data before attempting to build an empirical model?
- (c) Define the following terms and describe the processes that they represent
 - (i) Weak stationarity
 - (ii) Strict stationarity
 - (iii) Deterministic trend
 - (iv) Stochastic trend.
2. A researcher wants to test the order of integration of some time series data. He decides to use the DF test. He estimates a regression of the form

$$\Delta y_t = \mu + \psi y_{t-1} + u_t$$

and obtains the estimate $\hat{\psi} = -0.02$ with standard error = 0.31.

- (a) What are the null and alternative hypotheses for this test?
- (b) Given the data, and a critical value of -2.88 , perform the test.
- (c) What is the conclusion from this test and what should be the next step?
- (d) Why is it not valid to compare the estimated test statistic with the corresponding critical value from a t -distribution, even though the test statistic takes the form of the usual t -ratio?
3. Using the same regression as for question 2, but on a different set of data, the researcher now obtains the estimate $\hat{\psi} = -0.52$ with standard error = 0.16.
 - (a) Perform the test.
 - (b) What is the conclusion, and what should be the next step?
 - (c) Another researcher suggests that there may be a problem with this methodology since it assumes that the disturbances (u_t) are white noise. Suggest a possible source of difficulty and how the researcher might in practice get around it.
4. (a) Consider a series of values for the spot and futures prices of a given commodity. In the context of these series, explain the concept of cointegration. Discuss how a researcher might test for cointegration between the variables using the Engle–Granger approach. Explain also the steps involved in the formulation of an error correction model.

- (b) Give a further example from finance where cointegration between a set of variables may be expected. Explain, by reference to the implication of non-cointegration, why cointegration between the series might be expected.
5. (a) Briefly outline Johansen's methodology for testing for cointegration between a set of variables in the context of a VAR.
- (b) A researcher uses the Johansen procedure and obtains the following test statistics (and critical values):

r	λ_{max}	95% critical value
0	38.962	33.178
1	29.148	27.169
2	16.304	20.278
3	8.861	14.036
4	1.994	3.962

Determine the number of cointegrating vectors.

- (c) 'If two series are cointegrated, it is not possible to make inferences regarding the cointegrating relationship using the Engle–Granger technique since the residuals from the cointegrating regression are likely to be autocorrelated.' How does Johansen circumvent this problem to test hypotheses about the cointegrating relationship?
- (d) Give one or more examples from the academic finance literature of where the Johansen systems technique has been employed. What were the main results and conclusions of this research?
- (e) Compare the Johansen maximal eigenvalue test with the test based on the trace statistic. State clearly the null and alternative hypotheses in each case.
6. (a) Suppose that a researcher has a set of three variables, $y_t (t = 1, \dots, T)$, i.e. y_t denotes a p -variate, or $p \times 1$ vector, that she wishes to test for the existence of cointegrating relationships using the Johansen procedure.
- What is the implication of finding that the rank of the appropriate matrix takes on a value of
(i) 0 (ii) 1 (iii) 2 (iv) 3?
- (b) The researcher obtains results for the Johansen test using the variables outlined in part (a) as follows:

r	λ_{max}	5% critical value
0	38.65	30.26
1	26.91	23.84
2	10.67	17.72
3	8.55	10.71

Determine the number of cointegrating vectors, explaining your answer.

7. Compare and contrast the Engle–Granger and Johansen methodologies for testing for cointegration and modelling cointegrated systems. Which, in your view, represents the superior approach and why?
8. In EViews, open the ‘currencies.wf1’ file that will be discussed in detail in the following chapter. Determine whether the exchange rate series (in their raw levels forms) are non-stationary. If that is the case, test for cointegration between them using both the Engle–Granger and Johansen approaches. Would you have expected the series to cointegrate? Why or why not?