8

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

8.1 Stationarity and Unit Root Testing

8.1.1 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 6. 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 8.1.

  As Figure 8.1 shows, although one would have expected the R^2 values for each regression to be close to zero, since the explained and 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 8.1, Figure 8.2 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 8.2 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.

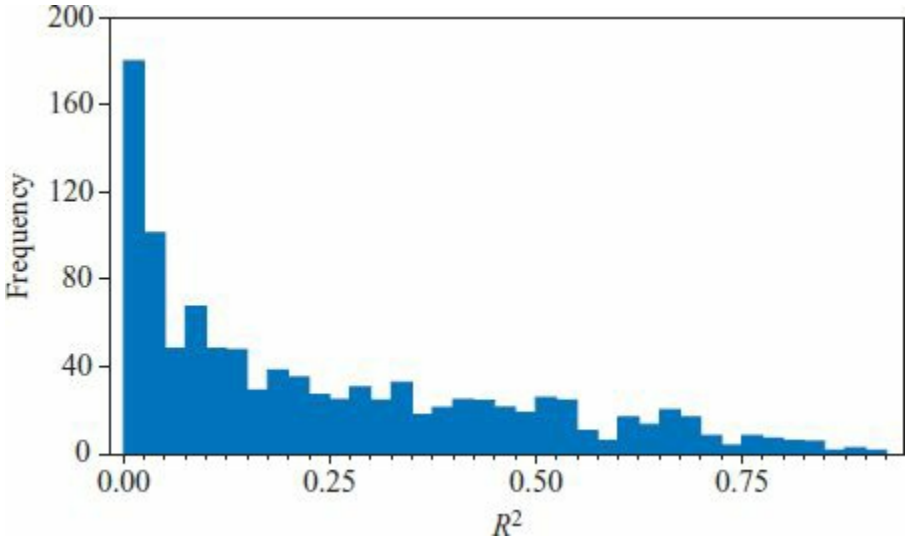

Figure 8.1 Value of R^2 for 1000 sets of regressions of a non-stationary variable on another independent non-stationary variable

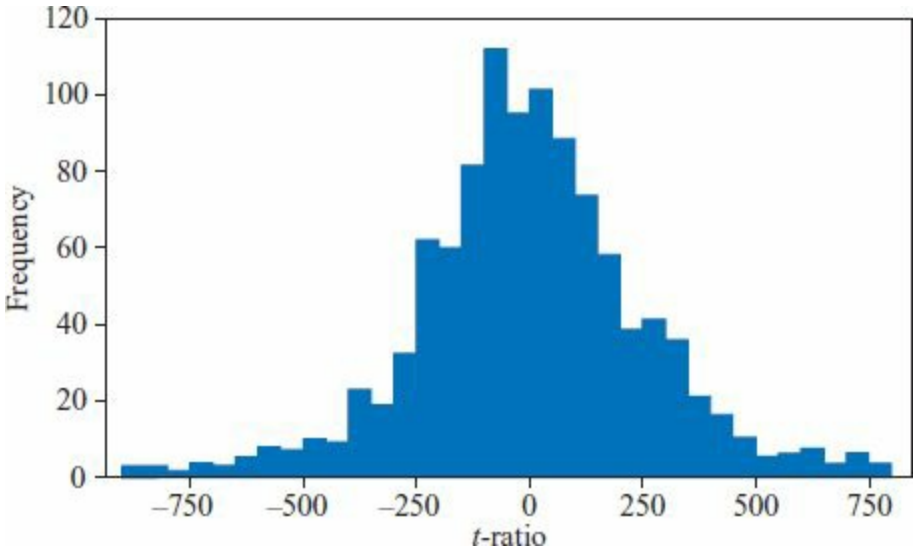

Figure 8.2 Value of *t*-ratio of slope coefficient for 1,000 sets of regressions of a non-stationary variable on another independent non-stationary variable

8.1.2 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*

$$y_t = \mu + y_{t-1} + u_t \tag{8.1}$$

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

$$y_t = \alpha + \beta t + u_t \tag{8.2}$$

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

Note that the model (8.1) could be generalised to the case where y_t is an explosive process

$$y_t = \mu + \phi y_{t-1} + u_t \tag{7.47}$$

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

$$y_t = \phi y_{t-1} + u_t \tag{8.4}$$

Let ϕ take any value for now. Lagging equation (8.4) one and then two periods

$$y_{t-1} = \phi y_{t-2} + u_{t-1} \tag{8.5}$$
$$y_{t-2} = \phi y_{t-3} + u_{t-2} \tag{8.6}$$

Substituting into equation (8.4) from equation (8.5) for y_{t-1} yields

$$y_t = \phi(\phi y_{t-2} + u_{t-1}) + u_t \tag{8.7}$$
$$y_t = \phi^2 y_{t-2} + \phi u_{t-1} + u_t \tag{8.8}$$

Substituting again for y_{t-2} from equation (8.6)

$$y_t = \phi^2(\phi y_{t-3} + u_{t-2}) + \phi u_{t-1} + u_t \tag{8.9}$$
$$y_t = \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \tag{8.10}$$

T successive substitutions of this type lead to

$$y_t = \phi^{T+1} y_{t-(T+1)} + \phi u_{t-1} + \phi^2 u_{t-2} + \phi^3 u_{t-3} + \cdots + \phi^T u_{t-T} + u_t \tag{8.11}$$

There are three possible cases:

(1) $\phi < 1 \Rightarrow \phi^T \to 0$ as $T \to \infty$
So the shocks to the system gradually die away – this is the *stationary case*.

(2) $\phi = 1 \Rightarrow \phi^T = 1 \ \forall \ T$
So shocks persist in the system and never die away. The following is obtained

$$y_t = y_0 + \sum_{t=0}^{\infty} u_t \text{ as } T \to \infty \tag{8.12}$$

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

$$y_t = \mu + y_{t-1} + u_t \tag{8.13}$$

and the trend-stationary process

$$y_t = \alpha + \beta t + u_t \tag{8.14}$$

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 equation (8.14) would be run, and any subsequent estimation would be done on the residuals from equation (8.14), 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 equation (8.13) is taken and y_{t-1} subtracted from both sides

$$y_t - y_{t-1} = \mu + u_t \tag{8.15}$$

$$(1 - L)\, y_t = \mu + u_t \tag{8.16}$$

$$\Delta y_t = \mu + u_t \tag{8.17}$$

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 equation (8.16) 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

$$y_t = \alpha + \beta t + u_t \tag{8.18}$$

This model can be expressed for time $t - 1$, which would be obtained by removing 1 from all of the time subscripts in equation (8.18)

$$y_{t-1} = \alpha + \beta(t-1) + u_{t-1} \tag{8.19}$$

Subtracting equation (8.19) from equation (8.18) gives

$$\Delta y_t = \beta + u_t - u_{t-1} \tag{8.20}$$

Not only is this a moving average in the errors that has been created, it is a noninvertible 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

$$\Delta y_t = \alpha_0 + \alpha_1 t + (\gamma - 1) y_{t-1} + u_t \tag{8.21}$$

Although again, of course the *t*-ratios in equation (8.21) will not follow a *t*-distribution and thus hypotheses about these parameters cannot be tested unless *y* is actually stationary in levels. 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

$$y_t = y_{t-1} + u_t \tag{8.22}$$

or

$$\Delta y_t = u_t \tag{8.23}$$

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 types of processes. Figure 8.3 plots a white noise (pure random) process, while Figures 8.4 and 8.5 plot a random walk versus a random walk with drift and a deterministic trend process, respectively.

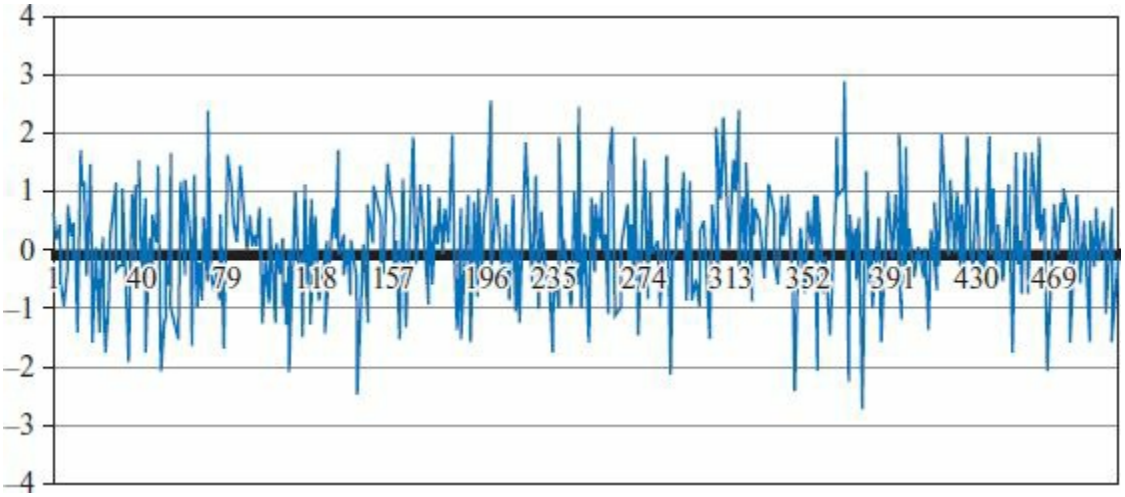

Figure 8.3 Example of a white noise process

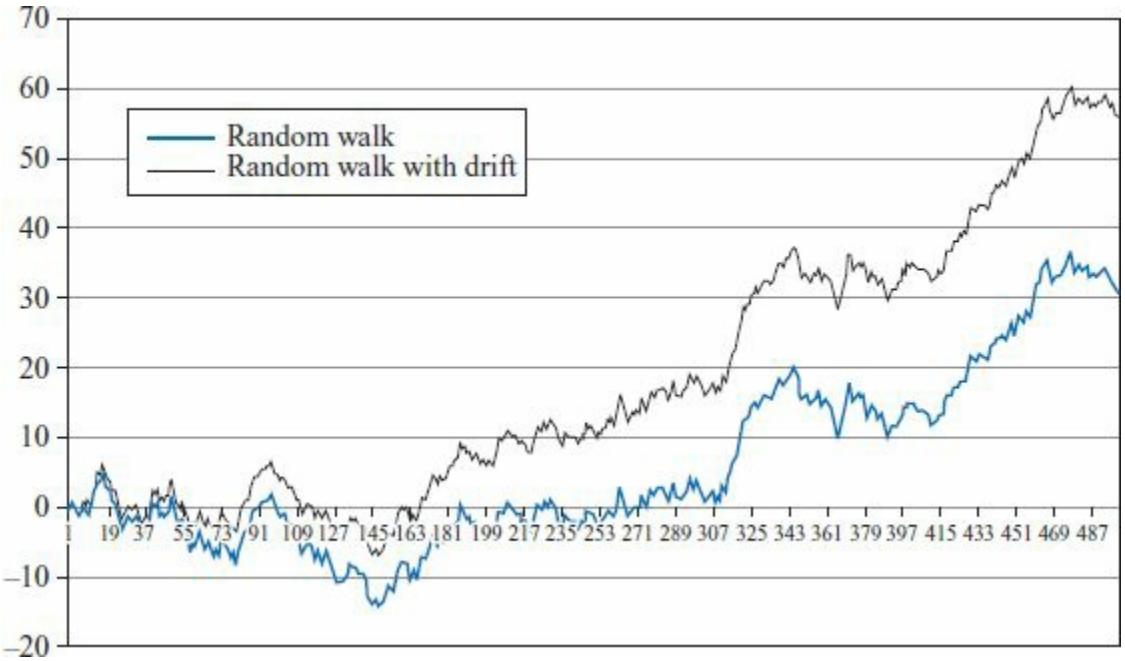

Figure 8.4 Time-series plot of a random walk versus a random walk with

drift

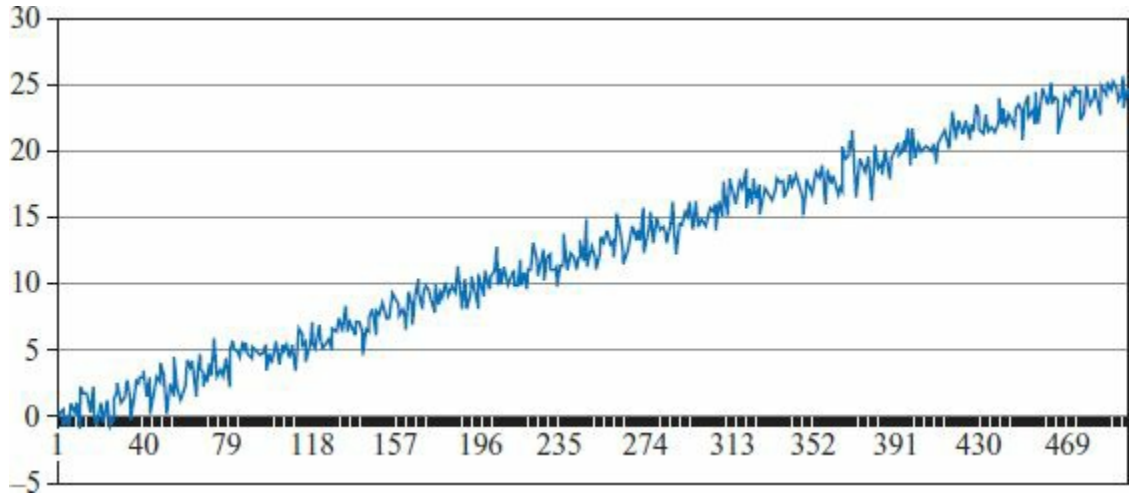

Figure 8.5 Time-series plot of a deterministic trend process

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 8.3, 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 8.4 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 greater the further the two processes are tracked. Finally, the deterministic trend process of Figure 8.5 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 8.3 would result. In this author's opinion, more time series in finance and economics look like Figure 8.4 than either Figure 8.3 or 8.5. Consequently, as stated above, the stochastic trend model will be the focus of the remainder of this chapter.

Finally, Figure 8.6 plots the value of an autoregressive process of order 1 with different values of the autoregressive coefficient as given by equation (8.4). 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.

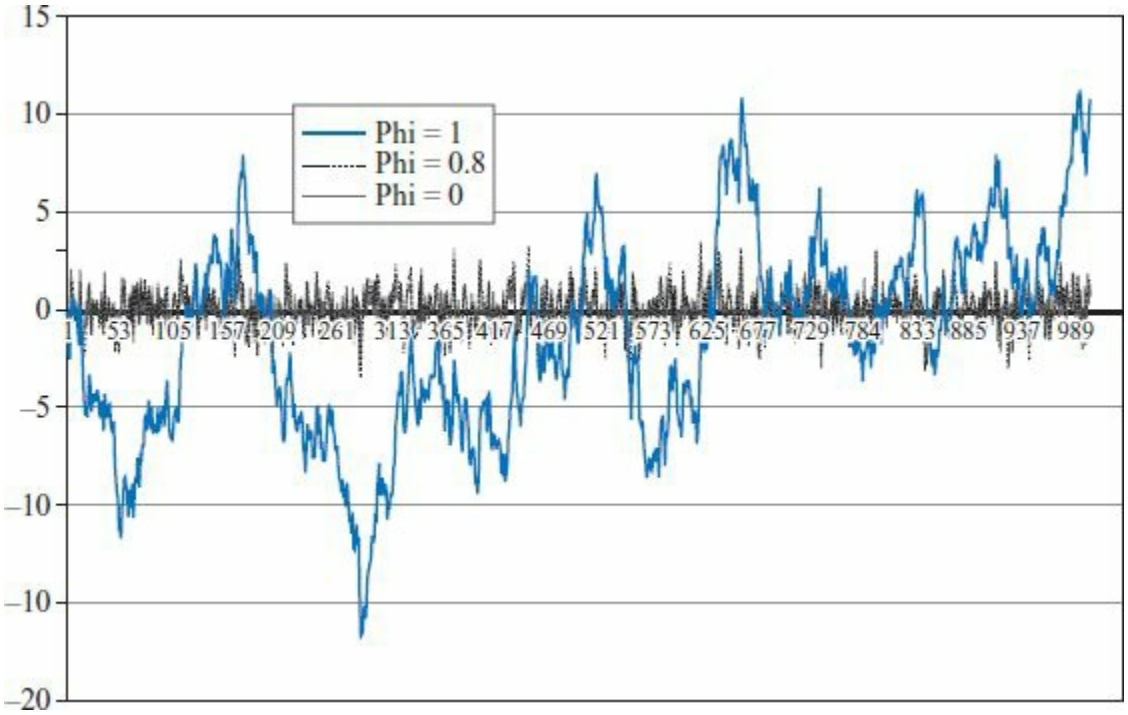

Figure 8.6 Autoregressive processes with differing values of ϕ (0, 0.8, 1)

8.1.3 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 $y_t \sim \mathrm{I}(d)$. So if $y_t \sim \mathrm{I}(d)$ then $\Delta^d y_t \sim \mathrm{I}(0)$. 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

$$y_t = y_{t-1} + u_t \tag{8.24}$$

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

$$y_t = 2y_{t-1} - y_{t-2} + u_t \tag{8.25}$$

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

$$y_t - 2y_{t-1} + y_{t-2} = u_t \tag{8.26}$$

$$(1 - 2L + L^2)y_t = u_t \tag{8.27}$$

$$(1 - L)(1 - L)y_t = u_t \tag{8.28}$$

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 equation (8.25) were differenced only once? Taking first differences of equation (8.25), i.e., subtracting y_{t-1} from both sides

$$y_t - y_{t-1} = y_{t-1} - y_{t-2} + u_t \tag{8.29}$$

$$y_t - y_{t-1} = (y_t - y_{t-1})_{-1} + u_t \tag{8.30}$$

$$\Delta y_t = \Delta y_{t-1} + u_t \tag{8.31}$$

$$(1 - L)\Delta y_t = u_t \tag{8.32}$$

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.

8.1.4 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

$$y_t = \phi y_{t-1} + u_t \tag{8.33}$$

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 equation (8.33), for ease of computation and interpretation

$$\Delta y_t = \psi y_{t-1} + u_t \tag{8.34}$$

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

$$y_t = \phi y_{t-1} + \mu + \lambda t + u_t \tag{8.35}$$

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

$$\Delta y_t = \psi y_{t-1} + \mu + \lambda t + u_t \tag{8.36}$$

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

$$test\ statistic = \frac{\hat{\psi}}{SE(\hat{\psi})} \tag{8.37}$$

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 13 in this book. Relevant examples of the distribution are shown in Table 8.1. A full set of DF critical values is given in the Appendix of Statistical Tables at the end of this book (Appendix 2). A discussion and example of how such critical values (CV) are derived using simulations methods are presented in Chapter 13.

Table 8.1 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

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 8.2), 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) (equation (8.34)) is

now written

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \quad \textbf{(8.38)}$$

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 twelve lags, if the data are quarterly, use four 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 7.

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.

8.1.5 Testing for Higher Orders of Integration

Consider the simple regression

$$\Delta y_t = \psi y_{t-1} + u_t \quad \textbf{(8.39)}$$

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

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

$\Delta^2 y_t (= \Delta y_t - y_{t-1})$ 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.

8.1.6 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.

8.1.7 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

$$y_t = 0.95y_{t-1} + u_t \quad \textbf{(8.40)}$$

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 8.1.

BOX 8.1 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 (Kwaitkowski *et al.*, 1992). The computation of the test statistic is not discussed here but the test is available within standard econometrics software such as EViews. 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*.

8.2 Tests for Unit Roots in the Presence of Structural Breaks

8.2.1 Motivation

The standard Dickey-Fuller-type unit root tests presented above do not perform well if there are one or more structural breaks in the series under investigation, either in the intercept or the slope of the regression. More specifically, the tests have low power in such circumstances and they fail to reject the unit root null hypothesis when it is incorrect as the slope parameter in the regression of y_t on y_{t-1} is biased towards unity by an unparameterised structural break. In general, the larger the break and the smaller the sample, the lower the power of the test. As Leybourne, Mills and Newbold (1998) have shown, unit root tests are also oversized in the presence of structural breaks, so they reject the null hypothesis too frequently when it is correct.[1]

Perron's (1989) work is important since he was able to demonstrate that if we allow for structural breaks in the testing framework, a whole raft of macroeconomic series that Nelson and Plosser (1982) had identified as non-stationary may turn out to be stationary. He argues that most economic time series are best characterised by *broken trend stationary processes*, where the data generating process is a deterministic trend but with a structural break around 1929 that permanently changed the levels (i.e., the intercepts) of the series.

8.2.2 The Perron (1989) Procedure

Recall from above that the flexible framework for unit root testing involves a regression of the form

$$\Delta y_t = \psi y_{t-1} + \mu + \lambda t + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \qquad \textbf{(8.41)}$$

where μ is an intercept and λt captures the time trend, one or both of which could be excluded from the regression if they were thought to be unnecessary.

Perron (1989) proposes three test equations differing dependent on the type of break that was thought to be present. The first he terms a 'crash' model that allows a break in the level (i.e., the intercept) of the series; the

second is a 'changing growth' model that allows for a break in the growth rate (i.e., the slope) of the series; the final model allows for both types of break to occur at the same time, changing both the intercept and the slope of the trend. If we define the break point in the data as T_b, and D_t is a dummy variable defined as

$$D_t = \begin{cases} 0 & \text{if } t < T_b \\ 1 & \text{if } t \geq T_b \end{cases}$$

the general equation for the third type of test (i.e., the most general) is

$$\Delta y_t = \psi y_{t-1} + \mu + \alpha_1 D_t + \alpha_2 (t - T_b) D_t + \lambda t + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \qquad \textbf{(8.42)}$$

For the crash only model, set $\alpha_2 = 0$, while for the changing growth only model, set $\alpha_1 = 0$. In all three cases, there is a unit root with a structural break at T_b under the null hypothesis and a series that is a stationary process with a break under the alternative.

While Perron (1989) commences a new literature on testing for unit roots in the presence of structural breaks, an important limitation of this approach is that it assumes that the break date is known in advance and the test is constructed using this information. It is possible, and perhaps even likely, however, that the date will not be known and must be determined from the data. More seriously, Christiano (1992) has argued that the critical values employed with the test will presume the break date to be chosen exogenously, and yet most researchers will select a break point based on an examination of the data and thus the asymptotic theory assumed will no longer hold.

As a result, Banerjee, Lumsdaine and Stock (1992) and Zivot and Andrews (1992) introduce an approach to testing for unit roots in the presence of structural change that allows the break date to be selected endogenously. Their methods are based on recursive, rolling and sequential tests. For the recursive and rolling tests, Banerjee *et al*. propose four specifications. First, the standard Dickey–Fuller test on the whole sample, which they term $\hat{t}_{DF}$; second, the ADF test is conducted repeatedly on the sub-samples and the minimal DF statistic, $\hat{t}_{DF}^{min}$, is obtained; third, the maximal DF statistic is obtained from the sub-samples, $\hat{t}_{DF}^{max}$; finally, the difference between the maximal and minimal statistics, $\hat{t}_{DF}^{diff} = \hat{t}_{DF}^{max} - \hat{t}_{DF}^{min}$, is taken. For the sequential test, the whole sample is used each time with the following regression being run

$$\Delta y_t = \psi y_{t-1} + \mu + \alpha \tau_t(t_{used}) + \lambda t + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \quad \textbf{(8.43)}$$

where $t_{used} = T_b/T$. The test is run repeatedly for different values of T_b over as much of the data as possible (a 'trimmed sample') that excludes the first few and the last few observations (since it is not possible to reliably detect breaks there). Clearly it is $\tau_t(t_{used})$ that allows for the break, which can either be in the level (where $\tau_t(t_{used}) = 1$ if $t > t_{used}$ and 0 otherwise); or the break can be in the deterministic trend (where $\tau_t(t_{used}) = t - t_{used}$ if $t > t_{used}$ and 0 otherwise). For each specification, a different set of critical values is required, and these can be found in Banerjee *et al.* (1992).

Perron (1997) proposes an extension of the Perron (1989) technique but using a sequential procedure that estimates the test statistic allowing for a break at any point during the sample to be determined by the data. This technique is very similar to that of Zivot and Andrews, except that his is more flexible, and therefore arguably preferable, since it allows for a break under both the null and alternative hypotheses, whereas according to Zivot and Andrews' model it can only arise under the alternative.

A further extension would be to allow for more than one structural break in the series – for example, Lumsdaine and Papell (1997) enhance the Zivot and Andrews (1992) approach to allow for two structural breaks. It is also possible to allow for structural breaks in the cointegrating relationship between series (see Section 8.4 below for a thorough discussion of cointegration) using an extension of the first step in the Engle-Granger approach – see Gregory and Hansen (1996).

8.2.3 An Example: Testing for Unit Roots in EuroSterling Interest Rates

Section 8.11 discusses the expectations hypothesis of the term structure of interest rates based on cointegration between the long and short rates. Clearly, key to this analysis is the question as to whether the interest rates themselves are I(1) or I(0) processes. Perhaps surprisingly, there is not a consensus in the empirical literature on whether this is the case. Brooks and Rew (2002) examine whether EuroSterling interest rates are best viewed as unit root process or not, allowing for the possibility of structural breaks in the series.[2] They argue that failure to account for structural breaks that may be present in the data (caused, for example, by changes in monetary policy or the removal of exchange rate controls) may lead to

incorrect inferences regarding the validity or otherwise of the expectations hypothesis. Their sample covers the period 1 January 1981 to 1 September 1997 to total 4,348 data points.

Brooks and Rew (2002) use the standard Dickey–Fuller test, the recursive and sequential tests of Banerjee *et al.* (1992), and their results are presented in Table 8.2. They also employ the rolling test, the Perron (1997) approach and several other techniques that are not shown here due to space limitations.

Table 8.2 Recursive unit root tests for interest rates allowing for structural breaks

Maturity	t_{DF}	Recursive statistics			Sequential statistics	
		$\hat{t}_{DF}^{max}$	$\hat{t}_{DF}^{min}$	$\hat{t}_{DF}^{diff}$	$\bar{t}_{DF,trend}^{min}$	$\bar{t}_{DF,mean}^{min}$
Short rate	−2.44	−1.33	−3.29	1.96	−2.99	−4.79
7-days	−1.95	−1.33	−3.19	1.86	−2.44	−5.65
1-month	−1.82	−1.07	−2.90	1.83	−2.32	−4.78
3-months	−1.80	−1.02	−2.75	1.73	−2.28	−4.02
6-months	−1.86	−1.00	−2.85	1.85	−2.28	−4.10
1-year	−1.97	−0.74	−2.88	2.14	−2.35	−4.55
Critical values	−3.13	−1.66	−3.88	3.21	−4.11	−4.58

Notes: Source: Brooks and Garrett (2002), taken from Tables 1, 4 and 5. $\bar{t}_{DF,trend}^{min}$ denotes the sequential test statistic allowing for a break in the trend, while $\bar{t}_{DF,mean}^{min}$ is the test statistic allowing for a break in the level. The final row presents the 10% level critical values for each type of test obtained from Banerjee *et al.* (1992, p. 278, Table 2).

The findings for the recursive tests are the same as those for the standard DF test, and show that the unit root null should not be rejected at the 10% level for any of the maturities examined. For the sequential tests, the results are slightly more mixed with the break in trend model still showing no signs of rejecting the null hypothesis, while it is rejected for the short, seven-day and the one-month rates when a structural break is allowed for in the mean.

Brooks and Rew's overall conclusion is that the weight of evidence across all the tests they examine indicates that short term interest rates are best viewed as unit root processes that have a structural break in their level around the time of 'Black Wednesday' (16 September 1992) when the UK dropped out of the European Exchange Rate Mechanism (ERM). The

longer term-rates, on the other hand, are I(1) processes with no breaks.

8.2.4 Seasonal Unit Roots

As we will discuss in detail in Chapter 10, many time series exhibit seasonal patterns. One approach to capturing such characteristics would be to use deterministic dummy variables at the frequency of the data (e.g., monthly dummy variables if the data are monthly). However, if the seasonal characteristics of the data are themselves changing over time so that their mean is not constant, then the use of dummy variables will be inadequate. Instead, we can entertain the possibility that a series may contain seasonal unit roots, so that it requires seasonal differencing to induce stationarity. We would use the notation $I(d, D)$ to denote a series that is integrated of order d, D and requires differencing d times and seasonal differencing D times to obtain a stationary process. Osborn (1990) develops a test for seasonal unit roots based on a natural extension of the Dickey–Fuller approach. Groups of series with seasonal unit roots may also be seasonally cointegrated. However, Osborn also shows that only a small proportion of macroeconomic series exhibit seasonal unit roots; the majority have seasonal patterns that can better be characterised using dummy variables, which may explain why the concept of seasonal unit roots has not been widely adopted.[3]

8.3 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 a set of variables $X_{i,t}$ 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 $i = 1, 2, 3, \ldots, k$ so that there are k variables each integrated of order d_i, and letting

$$z_t = \sum_{i=1}^{k} \alpha_i X_{i,t} \tag{8.44}$$

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 equation (8.44)

$$X_{1,t} = \sum_{i=2}^{k} \beta_i X_{i,t} + z_t' \tag{8.45}$$

where $\beta_i = -\frac{\alpha_i}{\alpha_1}, z_t' = \frac{z_t}{\alpha_1}, i = 2, \ldots, k.$ All that has been done is to take one of the variables, $X_{1,t}$, and to rearrange equation (8.44) to make it the subject. It could also be said that the equation has been normalised on $X_{1,t}$. But viewed another way, equation (8.45) 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)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad \textbf{(8.46)}$$

For the estimated model, the SRF would be written

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{u}_t \quad \textbf{(8.47)}$$

Taking everything except the residuals to the LHS

$$y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t} = \hat{u}_t \quad \textbf{(8.48)}$$

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*.

8.3.1 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

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

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.

8.3.2 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 8.9, 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.

8.4 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

$$\Delta y_t = \beta \Delta x_t + u_t \tag{8.49}$$

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 equation (8.49), i.e., $\Delta y_t = 0$; $\Delta x_t = 0$, and thus

everything in the equation cancels. Model equation (8.49) has no long-run solution and it therefore has nothing to say about whether x and y have an equilibrium relationship (see Chapter 5).

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

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + u_t \quad \textbf{(8.50)}$$

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 equation (8.50). 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 5.

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 equation (8.50) 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