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 $X_{i,t} \sim \mathrm{I}(d_i)$ 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 $z_t \sim \mathrm{I}(\max d_i)$. 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)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \tag{8.46}$$

For the estimated model, the SRF would be written

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{u}_t \tag{8.47}$$

Taking everything except the residuals to the LHS

$$y_t - \hat{\beta}_1 - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t} = \hat{u}_t \tag{8.48}$$

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

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

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

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \Delta w_t + \beta_3 (y_{t-1} - \gamma_1 x_{t-1} - \gamma_2 w_{t-1}) + \varepsilon_t \quad \textbf{(8.51)}$$

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).

8.5 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 $k - 1$ xs)

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t \quad \textbf{(8.52)}$$

u_t should be I(0) if the variables y_t, x_{2t}, …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 equation (8.52) 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

$$\Delta \hat{u}_t = \psi \hat{u}_{t-1} + v_t \quad \textbf{(8.53)}$$

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* ≈ 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?

$$\begin{aligned} H_0 &: u_t \sim I(1) \\ H_1 &: u_t \sim I(0). \end{aligned}$$

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 in the following section.

8.6 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.

8.6.1 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.,

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2(\hat{u}_{t-1}) + v_t \tag{8.54}$$

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 equation (8.48) above, the cointegrating vector would be $[1\ -\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3]$. It is now valid to perform 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

$$y_t = \alpha_1 + \beta_1 x_t + u_{1t} \tag{8.55}$$

What if instead the following equation was estimated?

$$x_t = \alpha_2 + \beta_2 y_t + u_{2t} \tag{8.56}$$

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.

(4) There may be more than one cointegrating relationship – see Box 8.2.

BOX 8.2 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}, … 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, 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 8.9.

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 8.8.

8.6.2 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.

8.7 Lead–Lag and Long-Term Relationships Between Spot and Futures Markets

8.7.1 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

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

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

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

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

$$F_t^* = S_t e^{(r-d)(T-t)} \tag{8.57}$$

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 (8.57) gives

$$f_t^* = s_t + (r-d)(T-t) \tag{8.58}$$

where f_t^* is the log of the fair futures price and s_t is the log of the spot price. Equation (8.58) 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, 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 DF test to the logs of the spot and futures prices of the ten-minutely FTSE data are shown in Table 8.3.

Table 8.3 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

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 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 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

$$s_t = \gamma_0 + \gamma_1 f_t + z_t \tag{8.59}$$

are stationary, using a DF test, where z_t is the error term. The coefficient values for the estimated equation (8.59) and the DF test statistic are given in Table 8.4.

Table 8.4 Estimated potentially cointegrating equation and test for cointegration for high frequency FTSE data

Coefficient	**Estimated value**
$\hat{\gamma}_0$	0.1345

$\hat{\gamma}_1$	0.9834
DF test on residuals	**Test statistic**
$\hat{z}_t$	−14.7303

Source: Brooks, Rew, and Ritson (2001).

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 two-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

$$\Delta \ln s_t = \beta_0 + \delta \hat{z}_{t-1} + \beta_1 \Delta \ln s_{t-1} + \alpha_1 \Delta \ln f_{t-1} + v_t \qquad \textbf{(8.60)}$$

where v_t is an error term. The coefficient estimates for this model are presented in Table 8.5.

Table 8.5 Estimated error correction model for high frequency FTSE data

Coefficient	**Estimated value**	***t*-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).

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 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.

8.7.2 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 8.6.

Table 8.6 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).

It can be seen from Table 8.6 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 error correction model with cost of carry (ECM-COC) model yields ten-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 ten 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 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 ten-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 8.7.

Table 8.7 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).

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.

8.7.3 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.

8.8 Testing for and Estimating Cointegration in 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:

$$\underset{g\times 1}{y_t} = \underset{g\times g}{\beta_1}\underset{g\times 1}{y_{t-1}} + \underset{g\times g}{\beta_2}\underset{g\times 1}{y_{t-2}} + \cdots + \underset{g\times g}{\beta_k}\underset{g\times 1}{y_{t-k}} + \underset{g\times 1}{u_t} \tag{8.61}$$

In order to use the Johansen test, the VAR (8.61) above needs to be turned into a vector error correction model (VECM) of the form

$$\Delta y_t = \Pi y_{t-k} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t \tag{8.62}$$

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 7. 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 ys 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 from zero (see Section 1.7.5 for some algebra and examples). The eigenvalues, denoted λ_i are put in descending order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_g$. If the λs are roots, in this context they must be less than one 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 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 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 $\ln(1 - \lambda_i) < 0 \,\forall\, i \geq 1$. That is, for 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

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{g} \ln(1 - \hat{\lambda}_i) \tag{8.63}$$

and

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{8.64}$$

where r is the number of cointegrating vectors under the null hypothesis and $\hat{\lambda}_i$ is the estimated value for the ith 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 an eigenvector. 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. $\lambda_{trace} = 0$ when all the $\lambda_i = 0$, for $i = 1, \ldots, g$.

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 (Appendix 2) 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_{max}) or more than r (for λ_{trace}). The testing is conducted in a sequence and under the null, $r = 0, 1, \ldots, g - 1$ so that the hypotheses for λ_{trace} are

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

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 $1 \leq \text{rank}(\Pi) < g$, 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.,

$$\Pi = \alpha\beta' \quad \textbf{(8.65)}$$

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

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{pmatrix} \tag{8.66}$$

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

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11} \quad \beta_{12} \quad \beta_{13} \quad \beta_{14}) \tag{8.67}$$

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

$$\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \\ \alpha_{13} & \alpha_{23} \\ \alpha_{14} & \alpha_{24} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{pmatrix} \tag{8.68}$$

and so on for r = 3, …

Suppose now that g = 4, and r = 1, as in equation (8.67), 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

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11} \quad \beta_{12} \quad \beta_{13} \quad \beta_{14}) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}_{t-k} \tag{8.69}$$

Equation (8.69) can also be written

$$\Pi = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} (\beta_{11}y_1 + \beta_{12}y_2 + \beta_{13}y_3 + \beta_{14}y_4)_{t-k} \tag{8.70}$$

Given equation (8.70), 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.

8.8.1 Tests for Cointegration with Mixed Orders of Integration

Suppose that we have a set of variables which we believe are related to one another and where there may potentially be a long-term relationship between some of them but where the individual variables are of different orders of integration. In the context of the Engle-Granger single equation approach, the test for cointegration will still be applicable, but the order of integration of the residuals in the potentially cointegrating regression will be the highest of the individual variables if they are not cointegrated and $I(0)$ if they are cointegrated. In practice we will again only be considering variables that are either $I(1)$ or $I(0)$, so suppose we have a set of three variables which are individually $I(1)$, $I(1)$, and $I(0)$. If the variables are cointegrated then the residuals will be $I(0)$ since these residuals will be a stationary linear combination of the two $I(1)$ variables and the variable which was already stationary ($I(0)$), whereas if they are not cointegrated then the residuals will be $I(1)$. Thus the $I(0)$ variable effectively acts like a constant from the perspective of non-stationarity.

Within the Johansen framework, if the number of variables in the system is N, then the cointegrating rank is equal to the sum of the number of linearly independent cointegrating vectors and the number of $I(0)$ variables in the system.

8.8.2 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

$$\text{test statistic} = -T\sum_{i=1}^{r}[\ln(1-\lambda_i) - \ln(1-\lambda_i{}^*)] \sim \chi^2(m) \tag{8.71}$$

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 over-identifying 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 equations (8.66)–(8.68) 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.

8.9 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

$$Q_t = \frac{E_t P_t^*}{P_t} \quad \textbf{(8.72)}$$

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 equation (8.72) and rearranging, another way of stating the PPP relation is obtained

$$e_t - p_t + p_t^* = q_t \quad \textbf{(8.73)}$$

where the lower case letters in equation (8.73) denote logarithmic transforms of the corresponding upper case letters used in equation (8.72). A necessary and sufficient condition for PPP to hold is that the variables on the LHS of equation (8.73) – 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 period April 1973 to December 1990. Pair-wise evaluations of the existence or otherwise of cointegration are examined for all combinations of these countries (ten 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 nonstationary, 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 8.8.

Table 8.8 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 and Francis Ltd (www.tandf.co.uk).

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 8.8. 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.

8.10 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).

8.10.1 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 8.9.

Table 8.9 DF tests for international bond indices

Panel A: test on log-index for country	DF Statistic
Germany	−0.395
Japan	−0.799

UK	−0.884
US	0.174
Panel B: test on log-returns for country	
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.

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 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

$$B_i = \alpha_0 + \alpha_1 B_j + u \tag{8.74}$$

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 8.10 here. They offer findings from applying seven different tests, while we present the results for only the Cointegrating Regression Durbin Watson (CRDW), Dickey–Fuller and Augmented Dickey–Fuller tests (although the lag lengths for the latter are not given in their paper).

Table 8.10 Cointegration tests for pairs of international bond indices

Test	UK–Germany	UK–Japan	UK–US	Germany–Japan	Germany–US	Japa US
CRDW	0.189	0.197	0.097	0.230	0.169	0.13[illegible]
DF	2.970	2.770	2.020	3.180	2.160	2.16[illegible]
ADF	3.160	2.900	1.800	3.360	1.640	1.89[illegible]

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

In this case, the null hypothesis of a unit root in the residuals from regression (8.74) cannot be rejected. The conclusion is therefore that there is no cointegration between any pair of bond indices in this sample.

8.10.2 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 *et al.* (1995), 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).

The Johansen systems procedure is then used to test for cointegration between the series. Unlike Clare *et al.*, Mills and Mills 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

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{g} \ln(1 - \hat{\lambda}_i) \qquad \textbf{(8.75)}$$

where λ_i are the ordered eigenvalues. The results are presented in their Table 2, which is modified slightly here, and presented in Table 8.11.

Table 8.11 Johansen tests for cointegration between international bond yields

r (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.

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 *et al.* – 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

$$\Delta X_t = \sum_{i=1}^{k} \Gamma_i \Delta X_{t-i} + v_t \tag{8.76}$$

where

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

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 → Germany → UK → 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 8.12 and 8.13, respectively.

Table 8.12 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.

Table 8.13 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
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.

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 explained by movements other than their own bond yields. In the German case, however, after twenty 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 twenty days) and UK (6.5% after twenty 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.

8.10.3 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.

8.11 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

$$R_t^{(n)} = \frac{1}{q}\sum_{i=0}^{q-1} E_t R_{t+mi}^{(m)} + c \tag{8.77}$$

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 equation (8.77) is considered, it can be seen that by subtracting $R_t^{(m)}$ from both sides of the relationship we have

$$R_t^{(n)} - R_t^{(m)} = \frac{1}{q}\sum_{i=0}^{q-1}\sum_{j=1}^{j=i} E_t[\Delta^{(m)} R_{t+jm}^{(m)}] + c \tag{8.78}$$

Examination of equation (8.78) 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 equation (8.78) 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 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 one month to twenty-five 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

$$X_t = [R_t \; R_t^{(2)} \; \ldots \; R_t^{(n)}]' \tag{8.79}$$

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 8.14.

Table 8.14 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**	$\lambda_{\max}$	λ_{trace}
1952*M*1–	$X_t = [R_t \; R_t^{(6)}]'$	2	$r = 0$	47.54***	49.82*

1978*M*12					
			$r \leq 1$	2.28	2.28
1952*M*1–1987*M*2	$X_t = [R_t \, R_t^{(120)}]'$	2	$r = 0$	40.66***	43.73*
			$r \leq 1$	3.07	3.07
1952*M*1–1987*M*2	$X_t = [R_t \, R_t^{(60)} \, R_t^{(120)}]'$	2	$r = 0$	40.13***	42.63*
			$r \leq 1$	2.50	2.50
1973*M*5–1987*M*2	$X_t = [R_t \, R_t^{(60)} \, R_t^{(120)}$				
	$R_t^{(180)} \, R_t^{(240)}]'$	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 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.

A Note on Long-memory Models

It is widely believed that (the logs of) asset prices contain a unit root. However, asset return series evidently do not possess a further unit root, although this does not imply that the returns are independent. In particular, it is possible (and indeed, it has been found to be the case with some financial and economic data) that observations from a given series taken some distance apart, show signs of dependence. Such series are argued to possess *long memory*. One way to represent this phenomenon is using a 'fractionally integrated' model. In simple terms, a series is integrated of a given order d if it becomes stationary on differencing a minimum of d times. In the fractionally integrated framework, d is allowed to take on non-integer values. This framework has been applied to the estimation of ARMA models (see, for example, Mills, 2008). Under fractionally integrated models, the corresponding autocorrelation function (ACF) will decline hyperbolically, rather than exponentially to zero. Thus, the ACF for a fractionally integrated model dies away considerably more slowly than that of an ARMA model with $d = 0$. The notion of long memory has also been applied to GARCH models (discussed in Chapter 9), where volatility has been found to exhibit longrange dependence. A new class of models known as fractionally integrated GARCH (FIGARCH) have been proposed to allow for this phenomenon (see Ding, Granger, and Engle, 1993 or Bollerslev and Mikkelsen, 1996).

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

SELF-STUDY 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}	5% 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, …, 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) of the question 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. (a) What issues arise when testing for a unit root if there is a structural break in the series under investigation?

(b) What are the limitations of the Perron (1989) approach for dealing with structural breaks in testing for a unit root?

[1] This material is fairly specialised and thus is not well covered by most of the standard textbooks. But for any readers wishing to see more detail, there is a useful and accessible chapter by Perron in Rao (1994). There is also a chapter on structural change in Maddala and Kim (1999).

[2] EuroSterling interest rates are those at which money is loaned/borrowed in British pounds but outside of the UK.

[3] For further reading on this topic, the book by Harris (1995) provides an extremely clear introduction to unit roots and cointegration, including a section on seasonal unit roots.

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