5 Non-stationarity and cointegration

Cointegration analysis, carefully applied, allows the analysis of long-run economic relationships. In some ways, this work parallels the work on error correction mechanisms which we discussed in Chapter 4 on dynamic modelling. As we shall see, there is a close relationship between cointegration and error correction models.

The basic insight of cointegration analysis is that, although many economic time series may tend to trend up or down over time in a non-stationary fashion, groups of variables may drift together. If there is a tendency for some linear relationships to hold between a set of variables over long periods of time, then cointegration analysis helps us to discover it. If an economic theory is correct we would expect the specific set of variables suggested by the theory to be related to each other (usually with constant parameters). So there should be no tendency for the variables to drift increasingly further away from each other as time goes on. If, however, there is *no* (linear) relationship between the variables they are said *not* to cointegrate and severe doubt must be cast on the usefulness of the underlying theory. This cointegration can be used to test the validity of an economic theory if the latter involves variables which in the data set exhibit strong (stochastic) trends.

5.1 Stationarity

A key concept in the discussion of this chapter is that of stationarity. In general, we shall be concerned with the idea of *weak* stationarity (see Spanos, 1986). A weakly stationary series has a constant mean and constant, finite variance. Thus, a time series (x_t) is stationary if its mean, $E(x_t)$, is independent of t, and its variance, $E[x_t - E(x_t)]^2$ is bounded by some finite number and does not vary systematically with time. Thus it will tend to return to its mean and fluctuations around this mean will have a broadly constant amplitude. A non-stationary series, on the other hand, will have a time-varying mean (or variance) and so we cannot in general refer to it without reference to some particular time period.

The simplest example of a non-stationary process is a random walk (without drift):

$$x_t = x_{t-1} + e_t$$

where e_t is independent and normal, denoted $\sim \mathrm{IN}(0, \delta^2)$ so that, if $x_0 = 0$

$$x_t = \sum_{i=1}^{t} e_i$$

The variance of x_t is $t\delta^2$ and this becomes infinitely large as $t \to \infty$. It is also clear that the concept of a mean value for x_t has no meaning. In fact, if at some point $x_t = c$ then the expected time until x_t again returns to c is infinite.

A stationary series tends to return to its mean and fluctuate around it within a more-or-less constant range. A non-stationary series would have a different mean at different points in time. One of the characteristics of a stationary series then is that it tends to return to, or cross, its mean values repeatedly and this property is the one which is exploited by most stationarity tests. As we discussed in Chapter 3, a stationary series will in general have an ARMA representation.

5.2 Unit roots and orders of integration

If a series must be differenced d times before it becomes stationary, then it is said to be integrated of order d, denoted $I(d)$. Thus, a series x_t is $I(d)$ if x_t is non-stationary but $\Delta^d x_t$ is stationary, where $\Delta x_t = x_t - x_{t-1}$ and $\Delta^2 = \Delta(\Delta x_t)$, etc. An alternative way of stating this is to say that a series is $I(d)$ if it has a stable, invertable non-deterministic ARMA representation after differencing d times – that is, if it is ARIMA (p, d, q) for some p, q. This means that the series can be written as

$$(1 - L)^d \phi(L) x_t = \theta(L) e_t \quad \mathbf{(5.1)}$$

where L is the lag operator $(L^n x_t = x_{t-n})$, $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator and e_t is a stationary process. If x_t is ARIMA (p, d, q), then we would have

$$\phi(L) = \sum_{i=0}^{p} \phi_i L^i \quad \text{and} \quad \theta(L) = \sum_{i=0}^{q} \theta_i L^i$$

Now consider the roots of the polynomial associated with the autoregressive part in (5.1), that is, the solutions to

$$(1 - L)^d \phi(z) = 0 \quad \mathbf{(5.2)}$$

where z is a real variable. Clearly, this has d roots (i.e. solutions) of $z = 1$, or in other words, d *unit roots*. It is for this reason that testing for the order of integration of a series is often referred to as testing for unit roots.

In general, if we take a linear combination of two series; each integrated of a different order, then the resulting series will be integrated at the highest of the two orders of integration. This can be easily demonstrated. Suppose

$$x_t \sim I(d), \qquad y_t \sim I(e) \quad \mathbf{(5.3)}$$

where $e > d$. Now form the linear combination, z_t:

$$z_t = \alpha x_t + \beta y_t \quad \mathbf{(5.4)}$$

If we difference z_t d times, we have:

$$\Delta^d z_t = \alpha \Delta^d x_t + \beta \Delta^d y_t \quad \mathbf{(5.5)}$$

Now the first term on the right-hand side of (5.5) is stationary, since $x_t \sim I(d)$, but the second term is not, since $y_t \sim I(e)$ and $e > d$ – it requires further differencing. As the sum of a stationary series $(\alpha \Delta^d x_t)$ and a non-stationary series $(\beta \Delta^d y_t)$ is non-stationary then $\Delta^d z_t$ is non-stationary. Suppose we now continue differencing up to a total of e times:

$$\Delta^e z_t = \alpha \Delta^e x_t + \beta \Delta^e y_t \quad \mathbf{(5.6)}$$

Now, $\alpha \Delta^e x_t$ is simply $\alpha \Delta^d x_t$ differenced $(e - d)$ times, and differencing a stationary series will always produce another stationary series. Thus, the first term on the right-hand side of (5.6) is stationary. The second term on the right-hand side is stationary since $y_t \sim I(e)$. Thus, $\Delta^e z_t$, as the sum of two stationary series must be stationary. This illustrates the general principle that, given (5.3) and (5.4), any linear combination has an order of integration equal to the highest order of the component series:

$$z_t \sim I[\max(d, e)]$$

Although this discussion has been in terms of two time series, it could easily be generalised to the case of three or more.

There are, however, exceptions to this general rule. Indeed, it is the exceptions of this rule which are of interest in cointegration analysis.

5.3 Cointegration

The important exception to this rule is where the low-frequency (or stochastic trend) components to two or more variables exactly offset each other to give a stationary linear combination. This is the case with a set of cointegrating variables. The basic idea is that if, *in the long run*, two or more series move closely together, even though the series themselves are trended, the difference between them is constant. We may regard these series as defining a long-run equilibrium relationship and, as the difference between them is stationary, the error term in a regression will have well-defined first and second moments. So traditional OLS regression becomes feasible in this case. The term *equilibrium* has many meanings in economics, and its use in the cointegration literature is rather different from most definitions of equilibrium. Within the cointegration literature all that is meant by equilibrium is that it is an *observed relationship* which has, on average, been maintained by a set of variables for a long period. Cointegration may be formally defined as: The components of the vector X_t are said to be cointegrated of order d, b [denoted $x_t \sim CI(d, b)$] if:

(i) all components of X_t are $I(d)$

and

(ii) there exists a vector $\alpha (\neq 0)$ such that $Z_t = \alpha' X_t \sim I(d-b)$, $b > 0$.

Thus if a set of $I(d)$ variables yields a linear combination that has a lower order of integration ($d - b < d$, for $b > 0$) then the vector α is called the cointegrating vector.

An important implication of this definition is that if we have *two* variables which are integrated of *different* orders then these two series cannot possibly be cointegrated. This is an intuitively clear result; it would be very strange to propose a relationship between an $I(0)$ series x_t and and $I(1)$ series y_t. The $I(0)$ series would have a constant mean while the mean of the $I(1)$ series would tend to drift over time. Thus, the 'error' $(y_t - \alpha x_t)$ between them would be expected to become infinitely large over time.

It is, however, possible to have a mixture of different order series when there are *three or more* series under consideration. In this case, a subset of the higher-order series must cointegrate to the order of the lower-order series. For example, suppose $Y_t \sim I(1)$, $X_t \sim I(2)$ and $W_t \sim I(2)$. If X_t and W_t cointegrate then $V_t = aX_t + cW_t$ will be $I(1)$. V_t is now a potential candidate to cointegrate with the remaining $I(1)$ series Y_t. If so, then $Z_t = eV_t + fY_t$, will be $I(0)$. We could summarise this set of circumstances as (i) $X_t, W_t \sim CI(2, 1)$; (ii) $V_t, Y_t \sim CI(1, 1)$ and hence (iii) $Z_t \sim I(0)$.

5.4 The Granger representation theorem

One of the most important results in cointegration analysis is the Granger representation theorem (Granger 1983, Engle and Granger 1987). This theorem states that if a set of variables are cointegrated of order 1, 1 [$CI(1, 1)$], then there exists a valid error-correction representation of the data. Thus, if X_t is an $N \times 1$ vector such that $X_t \sim (1, 1)$ and α is the cointegrating vector [i.e. $\alpha' X_t \sim I(0)$], then the following general error-correction representation may be derived:

$$\Phi(L)(1 - L)X_t = -\alpha' X_{t-1} + \Theta(L)e_t \tag{5.7}$$

where $\Phi(L)$ is a finite order polynomial with $\Phi(0) = I_N$, $\Theta(L)$ is a finite order polynomial, 'L' is the lag operator and at least one element of X is non-zero.

Equation (5.7) is a statistical model containing only stationary variables and so the usual stationary regression theory applies. This supplies a complete theoretical basis for the error-correction model when the 'levels terms' in X_t cointegrate. The Granger representation theorem also demonstrates that if the data generation process is an equation such as (5.7) then X_t must be a cointegrated set of variables. The practical implications of this for dynamic modelling are profound: in order for an error-correction model to be immune from the 'spurious regression problem' it must contain a set of levels terms which cointegrate to give a stationary error term. The danger with dynamic estimation is that the very richness of the dynamic structure may make the residual process appear to be white noise in a small sample when in fact the levels terms do not cointegrate and the true process is non-stationary.

There are a number of other, more minor, implications which follow from a set of variables (X_t, Y_t) being cointegrated. First, if X_t and Y_t are cointegrated then because Y_t and Y_{t-i} will be cointegrated for all i, then X_t and Y_{t-i} will be cointegrated. Second, if X_t and Y_t are cointegrated and individually $I(1)$, then either X_t must

Granger cause Y_t or Y_t must Granger cause X_t. This follows essentially from the existence of the error correction model (5.7) which suggests that, at the very least, the lagged value of one variable must enter the other determining equation.

5.5 Estimating the cointegrating vector

One approach to estimating the cointegrating vector would be to work with (5.7), the error-correction representation of the data. This, however, is not an easy procedure to implement as it must be remembered that (5.7) is a complete system of equations determining *all* of the elements of X_t. Further, there is the cross-equation restriction that the same parameter should occur in the levels parts of all the equations. So it would in principle, need to be estimated as a full system subject to this non-linear constraint. In fact, consistent estimates may be achieved much more easily following a suggestion made by Engle and Granger (1987) which relies on two theorems given in Stock (1987).

We discussed in Chapter 1 the property of *consistency*. A related concept is that of and order of convergence. If $\hat{\beta}$ is the OLS estimator in a regression model which satisfies the classical assumptions, then $\hat{\beta}$ converges in probability to the *true* parameter vector β as the *square root* of the sample size T tends to infinity, denoted $0(T^{1/2})$. Stock (1987) demonstrates that, if a set of variables are cointegrated of order (1, 1) with cointegrating vector α, then if $\hat{\alpha}$ is the OLS estimator of α, then $\hat{\alpha}$ is $0(T)$, i.e. $\hat{\alpha}$ converges in probability to α as T tends to infinity. Since T goes to infinity faster than $T^{1/2}$, this means that OLS estimates of the cointegrating vectors will generally be better, in some sense, than usual. This result is sometimes termed 'super consistency'.

The intuition behind the super consistency result is quite straightforward. Say, for example, we have two variables X_t, $Y_t \sim CI(1, 1)$. Consider the regression model

$$Y_t = \hat{\alpha}\, X_t + Z_t$$

where Z_t is the residual and $\hat{\alpha}$ is the OLS estimator. For the true value of α, $Y_t - \alpha X_t \sim I(0)$. Clearly, for $\hat{\alpha} \neq \alpha$, the OLS residual Z_t will be non-stationary and hence will have a very large variance in any finite sample. For $\hat{\alpha} = \alpha$, however, the estimated variance of Z_t will be much smaller. Since the ordinary least squares estimator essentially chooses $\hat{\alpha}$ to minimise the variance of Z_t, it will be extremely good at 'picking out' an estimate close to α.

However, offsetting this super consistency result is another result, also due to Stock (1987) which shows that there is a small-sample bias present in the OLS estimator of the cointegrating vector and that its limiting distribution is non-normal with a non-zero mean. Banerjee *et al.* (1986) suggest that this small-sample bias may be important in some cases and they show that for certain simple models the bias is related to $1 - R^2$ of the regression, so that a very high R^2 is associated with very little bias.

It is important to note that the proof of the consistency of the OLS estimator of the cointegrating vector does not require the assumption that the regressors are uncorrelated with the error term. In fact, any of the cointegrating variables may be used as the dependent variable in the regression and the estimates remain consistent. This means that problems do not arise when we have endogenous regressors or when these variables are measured with error. The reason for this may be seen quite easily at an intuitive level, the error process in the regression is $I(0)$ while the variables are $I(1)$ (or higher) so the means of the variables are time-dependent and will go to infinity. In effect what happens is that the growth in the means of the variables swamps the error process.

Engle and Granger (1987) demonstrate that once OLS has been used to estimate the cointegrating vector then the other parameters of the error correction model may be consistently estimate by imposing the first-stage estimates of the cointegrating vector in a second-stage regression. This is done simply by including the residuals from the first-stage regression in a general error correction model. This procedure is sometimes referred to as the two-step Granger and Engle estimation procedure. They also demonstrate that the OLS standard errors obtained at the second stage are consistent estimates of the true standard errors.

The advantages of the two-step procedure are that it allows us to make use of the super consistency properties of the first-stage estimates and that at the first stage it is possible to test that the vector of variables properly cointegrates. Thus, we can be sure that the full error correction model is not a spurious regression.

5.6 Testing for cointegration and drawing inference

Testing for cointegration

Suppose that we have an OLS estimate of the cointegrating vector $\hat{\alpha}$ and we may define the OLS residuals from the cointegrating regression as

$$Z_t = \hat{\alpha}' X_t$$

Now suppose Z_t follows an AR(1) process so that

$$Z_t = \rho Z_{t-1} + u_t$$

Then cointegration would imply stationary errors and hence that $\rho < 1$. The latter suggests testing the null hypothesis that $\rho = 1$ (i.e. that the error process is a random walk). The Dickey–Fuller test and the use of the Durbin–Watson statistic proposed by Sargan and Bhargava (1983) can both be used to test this hypothesis. There is, however, a further complication: if α is not known, the problem is much more complex: under the null hypothesis that $\rho = 1$ we cannot estimate α in an unbiased way. Because OLS will seek to produce the minimum squared residuals this will mean that the Dickey–Fuller tables will tend to reject the null too often. So we have to construct tables of critical values for each data generation process *individually* under the null hypothesis. Engle and Granger present some sample calculations of critical values for some simple models. We will discuss three of their proposed test procedures which have been most commonly used, namely the Dickey–Fuller, augmented Dickey–Fuller and cointegrating regression Durbin–Watson tests.

Consider the following autoregressive representation of a variable x_t:

$$x_t = \lambda_0 + \lambda_1 x_{t-1} + \lambda_2 x_{t-2} + \ldots \lambda_{n+1} x_{t-n-1} + u_t \qquad (5.8)$$

where v_t is a white noise, stationary error term.

Now reparameterise (5.8):

$$\Delta x_t = \lambda_0 + \left(\sum_{i=1}^{n+1} \lambda_i - 1\right) x_{t-1} - \sum_{z=1}^{n+1} \left[\left(\sum_{i=z}^{n+1} \lambda_i\right) \Delta x_{t-z}\right] + u_t \qquad (5.9)$$

Consider the regression

$$\Delta x_t = \beta_0 + \beta_1 x_{t-1} + \sum_{i=1}^{n} \alpha_i \Delta x_{t-1} + u_t \qquad (5.10)$$

Comparing (5.8), (5.9) and (5.10), for stationarity we require $\beta_1 < 0$, while if x_t is non-stationary, we would have $\beta_1 = 0$ and the sum of the autoregressive parameters λ_i in (5.8) would be unity (i.e. the series would have a unit root).

Thus, one way of testing for (non) stationarity of x_t would be to estimate a regression of the form (5.10) and to test the null hypothesis $\beta_1 = 0$. Intuitively, this could be done using the ratio of $\hat{\beta}$, to its estimated standard error. This 't-ratio' is termed the augmented Dickey–Fuller statistic (ADF). Unfortunately, under the null hypothesis of non-stationarity, the distribution of the ADF is not Student's t.

Fuller (1976), however, has tabulated approximate critical values of this statistic by Monte Carlo methods. The number of lags of Δx in (5.10) is normally chosen to ensure that the regression residual is approximately white noise. If *no* lags of Δx are required, then the 't-ratio' is termed the (non-augmented) Dickey–Fuller (DF) statistic. The critical values for the DF and ADF statistics, for a *single variable*, are the same and can be found in Fuller (1976). The DF and ADF statistics thus provide a method of testing for the order of integration of a variable.

Suppose, for example, $x_t \sim I(1)$, then in a regression of the form (5.10) we would be *unable* to reject the null hypothesis $\beta_1 = 0$. If we were then to run the regression

$$\Delta^2 x_t = \gamma_0 + \gamma_1 \Delta x_{t-1} + \sum_{i=1}^{n-1} \Psi_i \Delta^2 x_{t-i} + u_t \qquad (5.11)$$

we should be able to reject the hypothesis $\gamma_1 = 0$ against the alternative $\gamma_1 < 0$.

Dickey and Pantula (1988) suggest testing for higher-order unit roots and then 'testing down'. For example, estimate (5.11) first, then (5.10).

If a set of variables is cointegrated of order 1, $1 \sim CI(1, 1)$, then the residual from the cointegrating regression should be $I(0)$. It would therefore seem that one could test for cointegration by subjecting the cointegrating *residuals* to the DF and ADF tests, and this is indeed the case. There is, however, and additional complication in testing cointegrating residuals for non-stationarity using DF or ADF tests, which does not arise when applying these tests to single economic time series. This is because the OLS estimator 'chooses' the residuals in the cointegrating regression to have as small a sample variance as possible, even if the variables are *not* cointegrated, the OLS estimator will make the residuals *look* as stationary as possible. Thus, if we then use the DF or ADF tests on these residuals, we may reject the null hypothesis (non-stationarity) rather more than the nominal significance level would suggest. To correct for this test bias, the critical values have to be raised slightly. Engle and Granger (1987) have tabulated critical values for tests of this kind, generated by Monte Carlo methods.

Another test for the cointegrating residuals to contain a unit root, suggested by Bhargava (1980) and Sargan and Bhargava (1983), is to test the cointegrating regression Durbin–Watson (CRDW) statistic against a value of zero. This provides a useful complement to the two-step DF or ADF test, and the Monte Carlo results reported by Engle and Granger (1987) appear to show that it is quite powerful.

Intuitively, since CRDW $\simeq 2(1-\rho)$, where ρ is the first-order autocorrelation coefficient, CRDW $\simeq 0$ when $\rho = 1$.

Mackinnon (1988) lists the critical values for the CRDW, DF and ADF statistics for a number of cases and degrees of freedom.

5.7 Inference on parameter values

It has been well known for some time that non-stationarity not only presents problems for the consistency of estimation techniques but that the problem of inference is also greatly complicated. The Dickey–Fuller statistics discussed above do not have a standard Student's t distribution even though they are calculated as standard t-tests. Stock (1987), and Engle and Granger (1987) point out that the standard errors produced by OLS when performing a static cointegrating regression are biased and so valid inference about the parameters of the cointegrating vector cannot be carried out in the usual way. This bias arises for two quite separate reasons. First and most simply, a static regression will generally be subject to considerable serial correlation in the error process and for conventional textbook reasons this will give rise to inconsistent estimates of the standard errors of the parameters. The second reason is more important and more complex. The non-stationarity in the data gives rise to 'nuisance' parameters in the asymptotic distribution of the parameter estimates which means that the distribution of the parameter estimates is not generally normal.

More recent work, notably by West (1988), Sims *et al*. (1986) and Park and Phillips (1988, 1989) has shown that the situation is even more complex in that the presence or absence of drift terms in the non-stationary variables can crucially affect the form of the distribution of the parameter estimates.

We will discuss this topic initially within a bivariate framework of only two non-stationary variables Y and X and one stationary variable W, whereby assumption Y and X cointegrate. So the model is

$$Y_t = \alpha X_t + \beta W_t + e_t$$

where Y_t, X_t are $I(1)$ and by assumption W_t, e_t are $I(0)$. Now we assume that X_t is generated by the following univariate process $X_t = X_{t-1} + \mu + u_t$, where μ is the drift term (in the random walk) which may be zero. Now the key point in understanding the way inference lies in noting the way the asymptotic sample moments alter as the drift term alters from zero to a non-zero (positive) value. Some of the key results are summarised below:

Case: $\mu = 0$	*Case:* $\mu \neq 0$
$T^{-1}M_{XX} \rightarrow$ RV	$T^{-2}M_{XX} \rightarrow c$
$M_{Xe} \rightarrow$ RV	$M_{Xe} \rightarrow$ NRV

Distribution of the OLS estimators

$T(\hat{\alpha} - \alpha)$ is NSRV	$T^{3/2}(\hat{\alpha} - \alpha)$ is NRV
$T^{1/2}(\hat{\beta} - \beta)$ is NRV	$T^{1/2}(\hat{\beta} - \beta)$ is NRV

where M_{Xe} is the moment matrix of X and e, etc., RV is a random variable, NSRV is a non-standard random variable and NRV is a normal random variable. Note that with zero drift $\mu = 0$, the distribution of α is non-standard while the presence of non-zero drift causes the distribution to become a normally distributed random variable. Also note that the presence of non-stationary variables does not affect the distribution of the stationary variable W_t, so that inference can proceed in the usual way for the stationary components of a dynamic regression.

The general point here is that a researcher cannot normally use t-statistics to draw inference about the significance of parameters on the non-stationary terms in a regression. One exception which can be made is that if X_t is strictly exogenous then the randomness in the distribution of the OLS estimators comes only from e_t which is, of course, asymptotically normal by virtue of the assumption of cointegration. Then the distribution of the OLS estimators becomes normal and standard t-tests can be used. A further complication is that the above results for the case $\mu \neq 0$ (i.e. presence of drift) apply only to the bivariate case. If there are three $I(1)$ variables with non-zero drift then only some linear combination of the drift terms will be normally distributed and this cannot be assigned uniquely to any of the parameters.

5.8 Exogeneity and cointegration

Engle and Yoo (1989) give a classification of the possible combinations of cointegration and exogeneity assumptions and their effects on the distribution of the OLS estimator of the cointegrating vector. If we again continue the bivariate example, suppose we have the general system

$$Y_t = \alpha X_t + \beta \Delta X_t + u_t \qquad \text{(i)}$$

and

$$X_t = \gamma \Delta Y_{t-1} + \delta(Y_{t-1} - \alpha X_{t-1}) + v_t \qquad \text{(ii)}$$

Note that in this general model the same cointegrating parameter α, appears in both equations. This has an important implication for the weak exogeneity property of both Y and X (the definition of weak exogeneity is given in Chapter 4). The key point is that even though X_t is a function of *lagged* Y (and not current Y_t), it is *not* weakly exogenous in (i) above. This arises because the *parameters* of the equation generating X are not independent of the parameters of the equation generating Y: this is obviously true because they both have α in common. The general properties of the estimators are now given for various restrictions on this general model:

1. No restrictions imposed. The model is equivalent to a general VAR model and the distribution of the estimators are non-standard.
2. $\beta = \gamma = \delta = 0$. X_t is strongly exogenous and so the FIML estimator of α may be obtained from equation (i) alone and the distribution of the parameter is asymptotically normal.
3. $\delta = 0$. X_t is weakly exogenous and again the FIML estimator of α is given by OLS on equation (i) and the distribution of the parameters is asymptotically normal.
4. $\beta = \gamma = 0$. X_t is predetermined but *not* weakly exogenous as α is common to both equations. In this case OLS estimation applied to either the Y or X equations alone will yield non-normal asymptotic distributions and both equations should be estimated using a systems technique (see below for the ML estimator for the unrestricted system).

5.9 Three-step estimation

Engle and Yoo (1989) have proposed a 'third step' to the Engle and Granger two-step estimation technique which is computationally tractable and overcomes two of the disadvantages of the two-step procedure. The full three-step procedure is actually given for an unresticted multivariate system. This general form is not, however, particularly relevant as it has no claim to priority over the maximum likelihood procedure given below. In the special case of a unique cointegrating vector and the assumption of weak exogeneity of the conditioning variables of the dynamic model, the procedure becomes particularly easy to implement and has some claim to being of relevance to practical work. We will discuss this special case.

The two problems of the two-step procedure are:

1. While the static regression gives consistent estimates of the cointegrating vector these estimates are not fully efficient.
2. The distribution of the estimators of the cointegrating vector provided by the static regression is generally non-normal and so inference cannot be drawn about the significance of the parameters.

The third step provides a correction to the parameter estimates of the first stage, static regression which makes them asymptotically equivalent to FIML and provides a set of standard errors which allows the valid calculation of standard 't' tests.

The third stage consists simply of a further regression of the conditioning variables from the static regression multiplied by minus the error correction parameter, regressed on the errors from the second-stage error correction model. The coefficients from this model are the corrections to the parameter estimates while their standard errors are the relevant standard errors for inference.

The three steps are then: first estimate a standard cointegrating regression of the form

$$Y_t = \alpha X_t + Z_t$$

where Z_t is the OLS residual to give first-stage estimates of α, α^1. Then estimate a second-stage dynamic model using the residuals from the cointegrating regression to impose the long run constraint:

$$\Delta Y_t = \Phi(L)\Delta Y_{t-1} + \Omega(L)\Delta X_t + \delta Z_{t-1} + u_t$$

The third stage then consists of the regression

$$u_t = \varepsilon(-\hat{\delta} X_t) + v_t$$

The correction for the first-stage estimates is then simply

$$\alpha^3 = \alpha^1 + \varepsilon$$

and the correct standard errors for α_3 are given by the standard errors for ε in the third-stage regression.

We now turn to a practical example using the cointegration methodology.

5.10 Long-run purchasing power parity in the 1920s

Taylor and McMahon (1988) test for long-run purchasing power parity using cointegration techniques. Purchasing power parity (PPP) requires that the exchange rate between two currencies should be equal to the ratio of their price levels. If this is the case, then, at the going exchange rate, one unit of the domestic currency will have the same purchasing power in both countries. If we write

$$e_t - p_t \equiv u_t$$

where e_t is the (logarithm of the) nominal exchange rate (domestic price of foreign currency), p_t is the (logarithm of the) ratio of domestic to foreign prices, and u_t represents short-run deviations from PPP (logarithm of the real exchange rate), then *long-run* PPP would allow $u_t \neq 0$ in the short run, but would require $u_t = 0$ in the long run. At least a necessary condition for this to be the case is that u_t be a stationary process. If u_t is non-stationary, then it will tend to get larger over time and e_t and p_t will tend to diverge without bound. Thus, if e_t and p_t are $I(1)$, long-run PPP would require that they be cointegrated with a *unit* cointegrating parameter. Taylor (1988) uses simple models of mesurement error and transportation costs, to suggest that, even if long-run PPP holds, the cointegrating parameter may deviate from unity. Taylor and McMahon (1988) (TM) test for cointegration between exchange rates and relative prices for a number of exchange rates during the 1920s (i.e. under floating exchange rates). For illustrative purposes, we consider here only their results for the French franc–UK sterling exchange rate.

TM first test the exchange rate and relative price series for $I(1)$ behaviour. For franc–sterling they obtain ADF statistics of −0.71 and −0.78 for the exchange rate and relative prices respectively. The null hypothesis is that the series in question is $I(1)$. The rejection region is $\{\text{ADF} < c\}$ with $c = -3.58, -2.93$ or -2.60 at a significance level of 1%, 5% or 10% respectively (Fuller 1976). Thus, TM are unable to reject the hypothesis of $I(1)$ behaviour of exchange rates and relative prices.

Regressing the exchange rate on relative prices, they then obtain:

$$e_t = 3.272 + 1.061\ p_t + \omega_t \tag{5.12}$$

where ω_t is the OLS residual. They then use ω_t to construct the ADF and CRDW statistics, and obtain values of −4.62 and 0.662 respectively. The 1% rejection regions for the ADF statistic (applied to the cointegrating residuals) and for the CRDW are:

$$\text{ADF: } \{\text{ADF} < -3.77\}$$

$$\text{CRDW: } \{\text{CRDW} > 0.511\}$$

Thus, TM clearly reject the null hypothesis of $I(1)$ cointegrating residuals (i.e. non-cointegration) and conclude that the exchange rate and relative prices are cointegrated.

Since the slope coefficient in (5.12) is close to unity, TM suspect that the cointegrating parameter may in fact be unity. They thus test the *real* exchange rate, $u_t \equiv e_t - p_t$, for non-stationarity, using the ADF statistic, and obtain a test statistic value of −4.15. Since the cointegrating parameter has been *imposed* rather than estimated, this is compared to the Fuller (1976) critical values and the $I(1)$ null hypothesis is easily rejected. TM thus concluded that exchange rates and relative prices are cointegrated with a unit cointegrating parameter, implying that – at least for the 1920s – long-run PPP held between the franc and sterling. TM then proceed to estimate an error correction model for the franc–sterling exchange rate and report the result:

$$\underset{(0.323)}{\Delta e_t = 0.857} + \underset{(0.135)}{1.727\Delta p_t} - \underset{(0.189)}{0.803\Delta p_{t-1}} - \underset{(0.098)}{0.258(e-p)_{t-1}}$$

$$R^2 = 0.76,\ \text{DW} = 2.05,\ \text{LM}(6, 36) = 0.18$$

which has acceptable diagnostics. (LM is a Lagrange multiplier test statistic for up to sixth-order serial correlation.)

5.11 A maximum likelihood approach to cointegration

In sections 5.5 and 5.6 we outlined methods of testing for cointegration and estimating cointegrating vectors, based on ordinary least squares estimation. A major advantage of the least squares approach is that it is relatively simple and intuitive. It does, however, suffer from a number of disadvantages. One disadvantage is that the distribution of the test statistics discussed in section 5.6 will, in general, be slightly different in any particular application – they are not invariant with respect to the nuisance parameters which characterise any particular situation. Thus, the critical values given in Engle and Granger can be taken only as a rough guide.

A more fundamental problem concerns the *number* of cointegrating combinations which may exist between a set of variables. Consider two variables, each of which is integrated of order one $X_t \sim I(1)$ and $Y_t \sim I(1)$. Now, if (X_t, Y_t) cointegrates with parameter α then:

$$u_t = X_t - \alpha Y_t \sim I(0) \tag{5.13}$$

and α can be shown to be unique. To see this, suppose we had another cointegrating parameter, β:

$$\omega_t = X_t - \beta Y_t \sim I(0) \tag{5.14}$$

Adding and subtracting βY_t in (5.13):

$$u_t = X_t - (\alpha - \beta)\, Y_t - \beta Y_t$$

that is,

$$u_t = \omega_t - (\alpha - \beta) Y_t \qquad (5.15)$$

By assumption, u_t and ω_t are both $I(0)$ while Y_t is $I(1)$. The latter three conditions can hold only if $\alpha = \beta$ that is, α is unique. Unfortunately, once we consider more than two variables, it is no longer possible to demonstrate the uniqueness of the cointegrating vector. Indeed, it turns out that if we have a vector of N variables, each integrated of the same order, then there can be up to $(N-1)$ cointegrating vectors. (In the preceding paragraph, we merely demonstrate this for $N = 2$.)

Thus, if we cannot reject cointegration between a set of three or more variables, based on least squares methods, we have no guarantee that we have an estimate of a *unique* cointegrating vector. In a system with three variables, for example, it is quite possible that there are two statistically significant distinct cointegrating vectors and that our OLS estimate is a linear combination of them.

Johansen (1988) suggests a method for both estimating all the distinct cointegrating relationships which exist within a set of variables and for constructing a range of statistical tests. The method begins by expressing the data generation process of a vector of N variables X as an unrestricted vector autoregression in the levels of the variables:

$$X_t = \Pi_1 X_{t-1} + \ldots + \Pi_k X_{t-k} + e_t \qquad (5.16)$$

where each of the Π_i is an $(N \times N)$ matrix of parameters. The system of equations (5.16) can be reparameterised in ECM form:

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \ldots \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_K X_{t-k} + e_t \qquad (5.17)$$

$$\Gamma_i = -I + \Pi_1 + \ldots \Pi_i, \qquad i = 1, \ldots, k.$$

Thus Γ_k now defines the *long run* 'levels solution' to (5.16).

Now, if X_t is a vector of $I(1)$ variables, we know that the left-hand side and the first $(k-1)$ elements of (5.17) are $I(0)$ but that the last element of (5.17) is a linear combination of $I(1)$ variables. Johansen uses cannonical correlation methods to estimate all the distinct combinations of the levels of X which produce high correlations with the $I(0)$ elements in (5.17); these combinations are, of course, the cointegrating vectors. Johansen's approach is a maximum likelihood method of estimating *all* of the distinct cointegrating vectors which may exist between a set of variables. Johansen also shows how one can test which of these distinct cointegrating vectors are statistically significant, and also how to construct a likelihood ratio test for linear restrictions on the cointegrating parameters.

Consider an N-dimensional vector of variables X_t. Johansen starts by considering a kth order vector autoregression (VAR) for X_t (5.16) where e_t is an independent and identically (normally) distributed vector of disturbances, with zero mean and covariance matrix Δ. All terms on the right-hand side of (5.17) are clearly $I(0)$ except the final term. Thus, the last term on the right-hand side must also be $I(0)$: $\Gamma_K X_{t-k} \sim I(0)$, either X contains a number of cointegrating vectors or Γ_K must be a matrix of zeros.

Now consider an $N \times r$ matrix β such that

$$\beta' X_{t-k} \sim I(0)$$

If all the elements of X_t are $I(1)$, then the columns of β must form cointegrating parameter vectors for X_{t-k} and hence X_t. Since there can only be up to $(N-1)$ cointegrating vectors, β must have r less than N. If, however, X_t is $I(1)$ but the elements are *not* cointegrated, β must be a null matrix. Now define another $(N \times r)$ matrix α such that:

$$-\Gamma_k = \alpha\beta' \qquad (5.18)$$

The Johansen technique is based upon estimating the factorisation (5.18). Suppose, for example, that there was in fact only one cointegrating vector. Then we need consider only the first column of α and β, (5.16) could then be written:

$$\Delta X_t = \Gamma_1 X_{t-1} + \Gamma_2 \Delta X_{t-2}\, (-\alpha_1 Z_{t-k}) + e_t \qquad (5.19)$$

where $Z_t = \beta_1 X_t \sim I(0)$.

The system (5.19) is directly analogous to (5.7). Indeed, it is the error correction representation of the system where the lag length k is assumed high enough to allow one to assume a white noise disturbance vector, e_t, and the error correction term enters with lag k. (It is in fact easy to show, by simply rearranging terms, that the error-correction term can enter at any lag.) Thus, Johansen provides a technique for estimating all possible cointegrating vectors, the β matrix as well as the corresponding set of error-correction coefficients, the α matrix. If the X_t vector does in fact cointegrate – one or more of the β_i vectors are statistically significant – then, by the Granger representation theorem, we know that α_i must contain at least one non-zero element. In general, considering all of the logically possible cointegrating vectors, (5.19) is written

$$\Delta X_t = \Gamma_1 X_{t-1} + \Gamma_2 \Delta X_{t-2} + \ldots (-\alpha\beta')X_{t-k} + e_t \quad (5.20)$$

A fuller discussion of the Johansen technique is given in the appendix. Here, we can consider the following sketch.

The likelihood function for the system (5.20) is proportional to

$$L(\alpha, \beta, \Delta; \Gamma_,, \ldots \Gamma_{k-1}) = |\Omega|^{-T/2} \exp\left\{-1/2 \sum_{t=1}^{T} (e_t \Omega^{-1} e_t)\right\}$$

Where T is the number of observations Ω is the covariance matrix of e.

Rewrite the system (5.20) as

$$\Delta X_t + \alpha\beta' X_{t-k} = \Gamma_1 \Delta X_{t-1} + \ldots \Gamma_{k-1} \Delta X_{t-k+1} + e_t \quad (5.21)$$

If $(\alpha\beta')$ were known, maximum likelihood estimates of the Γ_i could be obtained by ordinary least squares. Consider therefore, correcting for the effect of the *k lags* of ΔX_t on ΔX_t and X_{t-k}. Correcting the right-hand side for ΔX_{t-j} ($j = 1, 2, \ldots$), i.e. taking out their effect, just leaves e_t. We can correct X_t and X_{t-k} for the effects of k lags of X_t by replacing ΔX_t and X_{t-k} with the residuals from regressing them individually on $\{\Delta X_{t-1}, \ldots \Delta X_{t-k}\}$. Note that this will not change their basic properties; X_{t-k} remains $I(1)$ and ΔX_t remains $I(0)$. Thus (5.21) becomes:

$$R_{ot} + \alpha\beta' R_{kt} = e_t \quad (5.22)$$

where R_{ot} is the vector of residuals from regressing ΔX_t on to $\Delta X_{t-1}, \ldots, \Delta X_{t-k}\}$ and R_{kt} is the corresponding residual vector for X_{t-k}. The expression for the likelihood function, (5.20), can now be written:

$$L_1(\alpha, \beta, \Omega)$$

$$= |\Delta|^{-T/2} \exp\left\{-1/2 \sum_{t=1}^{T} (R_{ot} + \alpha\beta' R_{Kt})'\Omega^{-1}(R_{ot} + \alpha\beta' R_{Kt})\right\} \quad (5.23)$$

If β were known, an estimate of α and of Δ could be obtained in the usual way from a regression of R_{ot} on $\beta' R_{Kt}$. Thus, $\hat{\alpha}$ and $\hat{\Delta}$ can be expressed as functions of β.

$$\hat{\alpha}(\beta) = -S_{ok}\beta(\beta' S_{kk}\beta)^{-1} \quad (5.24)$$

$$\hat{\Omega}(\beta) = S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko} \quad (5.25)$$

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it}R'_{jt}, \quad i, j = o, k \quad (5.26)$$

After substituting (5.24) and (5.25) into (5.23), the concentrated likelihood function can be seen to be proportional to

$$L_2(\beta) = |\hat{\Omega}(\beta)|^{-T/2}$$

$$= |S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko}|^{-T/2} \quad (5.27)$$

Thus, maximum likelihood estimation of the full set of possible cointegrating vectors, β, involves choosing β to minimise the function

$$F = |S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko}| \quad (5.28)$$

Johansen shows how this can be done by solving an eigenvalue problem. The matrix $\hat{\beta}$ is thus obtained as a set of eigenvectors together with a corresponding vector of $(N-1)$ eigenvalues $\hat{\lambda}$. The columns of β are significant only if the corresponding eigenvalue is significantly different from zero. Let the elements of $\hat{\lambda}_i$ be ordered such that

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots \hat{\lambda}_{N-1}$$

and let the columns of β also be ordered accordingly (i.e. so that in the column of $\hat{\beta}$, $\hat{\beta}_i$, is the eigenvector corresponding to $\hat{\lambda}_i$). These eigenvalues are defined such that the maximum likelihood estimate of Ω is given by

$$\hat{\Omega}(\beta) = |S_{oo}| \prod_{i=1}^{N} (1 - \hat{\lambda}_i) \quad (5.29)$$

Now suppose we wish to test the null hypothesis that there are at most r cointegrating vectors:

$$H_0: \lambda_i = 0, \ i = r+1, \ldots, N-1$$

where only the first r eigenvalues are non-zero. If these restrictions are imposed, the restricted estimate of Δ denoted $\tilde{\Delta}$ is:

$$\tilde{\Omega}(\beta) + |S_{oo}| \prod_{i=1}^{r} (1 - \lambda_i) \quad (5.30)$$

Since the likelihood function can be expressed in terms of the estimate of Δ, equation (5.27), we can use (5.27), (5.29) and (5.30) to form a likelihood ratio statistic for the null hypothesis of at most r cointegrating vectors.

$$\text{LR}(N-r) = -2\ln(Q) = -T \sum_{i=r+1}^{N} \ln(1 - \hat{\lambda}_i) \quad (5.31)$$

where

$$Q = \frac{\text{restricted maximised likelihood}}{\text{unrestricted maximised likelihood}}$$

LR$(N-r)$ has degrees of freedom equal to the number of restrictions, $(N-r)$. Note that for $\hat{\lambda}_i = 0$, $i = r+1, \ldots N$, LR$(N-r)$ will be zero, and will tend to get large as one or more of the $\hat{\lambda}_i$ approach unity.

The likelihood ratio statistic defined in (5.31) does not, in fact, have a χ^2 distribution, even in large samples. Johansen does, however, find the asymptotic distribution of LR$(N - r)$ by applying some results in Brownian motion theory. This distribution does not vary with the particular model being estimated or other variable factors as in the case of the Dickey–Fuller tests for cointegration; however, it is not invariant to the assumption made about the underlying VAR model. In particular, there are three main assumptions which may be made:

1. The VAR may be as specified in (5.16) without any constant term.
2. The VAR has a restricted constant term which appears only as a part of the cointegrating vectors so that the ECM form (5.17) contains any constants within the term $\Gamma_K X_{t-K}$ only.
3. The VAR has an unrestricted constant. This means that if the ECM form of the VAR (5.17) has some equations which do not contain a cointegrating vector (so that they are purely difference equations) these equations will still contain constants. This is unlike assumption 2, where the constants were associated with the cointegrating vectors. So these variables will behave like generalised random walk variables but with a drift term and the data will contain deterministic trend terms. This assumption is therefore characterised by the presence of deterministic trend in some of the variables.

Johansen (1989) gives the critical values for all three cases for the test outlined in (5.31).

5.12 Testing linear restrictions on the cointegrating parameters

In section 5.11 we gave an outline of a maximum likelihood technique for testing for and estimating the set of unique cointegrating vectors. Johansen (1988) also demonstrates how the technique can be applied to test linear restrictions on the parameters of the cointegrating vectors.

Suppose that, after an initial application of the Johansen technique, we have decided that there are at most r cointegrating vectors among the N-dimensional vector X_t. Let the $(N \times r)$ matrix of cointegrating vectors be β. Johansen considers linear restrictions on β which reduce the number of *independent* cointegrating parameters from N to S, $S \leq N$.

For example, suppose we analysed a vector consisting of the exchange rate (e_t) and domestic and foreign prices, p_t^* and p_t respectively (all in logarithms). As in the example discussed in section 5.10, if $Z_t = (p_t e_t p_t^*)'$ was an $I(1)$ vector, then long-run purchasing power parity would suggest that

$$g_t = e_t - p_t + p_t^* \sim I(0) \tag{5.32}$$

Thus, if we found $r = 1$ statistically significant cointegrating vectors:

$$\beta_{i1} e_t + \beta_{i2} p_t + \beta_{i3} p_t^* = g_t, \quad i = 1, \ldots, r \tag{5.33}$$

Then, the restrictions in (5.32) involve reducing the number of independent cointegrating parameters from three to one. For the full $(N \times r)$ matrix β, they can be written:

$$\beta = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \phi$$

where ϕ is a $(S \times r)$ (in this case 1×1) matrix of parameters.

In general, Johansen considers restrictions which can be written in the form

$$H_o\colon \beta = H\phi \tag{5.34}$$

where H is an $(N \times S)$ matrix of full rank $= S$ and ϕ is an $(S \times r)$ matrix of unknown parameters.

The method of obtaining the restricted estimates is straightforward. Since H is known, simply replace β with $H\phi$ in the procedure discussed in the previous section, to obtain an estimate ϕ^* say. The restricted estimate of β is then given by $\beta^* = H\phi^*$.

Along with the restricted estimates will be produced a set of eigenvalues, $\hat{\lambda}$, corresponding to the set produced in the unrestricted estimation, and similarly ordered such that $\lambda_1^* > \lambda_2^*, \ldots \lambda_r^*$. The relationship between these eigenvalues and the maximised value of the likelihood function, see (5.27), (5.29), (5.30), can then be exploited to yield a test of the hypothesis based on the first r cointegrating vectors:

$$\text{LR}^*[r(N - S)] = -2 \ln Q = T \sum_{i=1}^{r} \ln \{(1 - \lambda_i^*)/(1 - \hat{\lambda}_i)\} \tag{5.35}$$

This will have an asymptotic chi-square distribution with $r(N - S)$ degrees of freedom. As is generally the case with likelihood ratio statistics, the number of degrees of freedom is equal to the number of restrictions $r(N - S)$ since $(N - S)$ fewer parameters are estimated in each of the r cointegrating parameters vectors.

5.13 Example: The demand for broad money during the Gold Standard

Taylor (1991) estimates a 'long-run' demand function for UK broad money for the period 1871–1913. He starts by testing for unit roots in a pre-specified set of variables – broad money (m_t), prices (p_t), real income (y_t), the long bond rate (RL_t) and the prime bill rate (RPB_t), (all data except interest rates in logarithms). The results are listed here as Table 5.1. Taylor uses the DF test and the Johansen statistic (5.31), where '*cointegration in one variable*' simply implies that the variable is $I(0)$. Table 5.1 suggests that all of the 'levels variables' listed above correspond to $I(1)$. Table 5.2(a) then demonstrates that at a nominal significance level of 5% the hypothesis of zero cointegrating vectors is strongly rejected, while the hypothesis of one or more cointegrating vectors is not. Taylor thus concludes that there is a unique statistically significant cointegrating vector relating the variables. This is reported as the unrestricted equation 1 in Table 5.2(b), where the cointegrating parameters have been normalised on m_t.

Taylor then argues that, whilst the short interest rates RD_t and RPB_t may effect 'long run' or 'average' money demand, their long-run effect will be felt only through their constant means. He therefore tests for cointegration amongst, m_t, p_t, y_t and RL_t. He then estimates a VAR for these variables with lag length two (chosen on the basis of standard diagnostics, see Chapter 4) and applies the Johansen procedure. The results are given in Table 5.2.

Table 5.1 Unit root tests for money, prices, income and interest rates

Variable	Dickey–Fuller statistic	Johansen statistic
m_t	1.00	1.04
Δm_t	−4.21	16.80
p_t	−1.60	1.66
Δp_t	−5.03	22.97
y_t	−0.32	0.28
Δy_t	−3.82	21.82
RL_t	−0.68	2.24
ΔRL_t	−3.54	9.87
RD_t	−3.68	14.35
RPB_t	−3.30	11.69

Note: The null hypothesis in each case is that the variable in question is $I(1)$; the 5% rejection region for the Dickey–Fuller statistic is $\{\text{DF}\in R|\text{DF} < -2.93\}$ (Fuller 1976, p. 373); the 5% rejection region for the Johansen statistic is $\{J\in R|J > 9.094\}$ (Johansen 1989).

Table 5.2 Applying the Johansen procedure to money demand during the Gold Standard

(a) Tests for cointegration

Null hypothesis	Likelihood ratio statistic	5% critical value
Number of cointegrating vectors r		
$r \leqslant 3$	0.001	9.094
$r \leqslant 2$	5.12	20.168
$r \leqslant 1$	20.65	35.068
$r = 0$	60.21	53.347

(b) Estimated cointegrating vector (largest eigenvalue only)

1. Unrestricted: $m_t = 1.06p_t + 0.97y_t - 0.097\text{RL}_t$
2. With homogeneity restrictions: $m_t = p_t + y_t - 0.076\text{RL}_t$
   Likelihood ratio statistic: LR(2) = 1.39 (0.50)
3. With exclusion restriction on RL: $m_t = 0.78p_t + 1.01y_t$
   Likelihood ratio statistic: LR(1) = 7.618(0.5E − 2)

Note: Figures in parenthesis are marginal significance levels. The LR(n) statistics are asymptotically $\chi^2(n)$ variates under the null hypothesis.

The likelihood ratio statistics

Given the cointegrating vector (non-normalised) from the Johansen procedure:

$$\beta_{11}m_t + \beta_{12}p_t + \beta_{13}y_t + \beta_{14}\text{RL}_t \sim I(0)$$

then equation 1 in Table 5.2(b) corresponds to

$$m_t = -(\beta_{12}/\beta_{11})p_t - (\beta_{13}/\beta_{11})y_t - (\beta_{14}/\beta_{11})\text{RL}_t$$

This equation clearly looks like a 'textbook' money demand function – it has a negative interest rate semi-elasticity and the coefficients on prices and income are positive and close to unity [Table 5.2(b), equation 1].

Taylor then tests for price and income homogeneity, i.e that the 'long-run' coefficients on prices and income are unity when normalised on money. In terms of (5.34), these restrictions are written:

$$\begin{bmatrix}\beta_{11}\\ \beta_{12}\\ \beta_{13}\\ \beta_{14}\end{bmatrix} = \begin{bmatrix}1 & 0\\ -1 & 0\\ -1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\phi_{11}\\ \phi_{12}\end{bmatrix}$$

so that

$$\beta_{11} = \phi_{11},\ \beta_{12} = -\phi_{11},\ \beta_{13} = -\phi_{11} \quad \text{and} \quad \beta_{14} = \phi_{12}$$

The restricted estimates are given as equation 2 in Table 5.2(b) and the likelihood ratio statistic (5.35) for the restrictions is LR(2) = 4.57. The restrictions are not rejected at the 5% level.

Finally, Taylor tests whether RL_t can be excluded from the cointegrating vector. The likelihood ratio statistic listed alongside the restricted equation 3 in Table 5.2(b) shows that this hypothesis is easily rejected at the 1% level.

5.14 Summary

Cointegration deals with the relationships between variables that have stochastic trends. If cointegration is not rejected then there exists one or more 'long-run' linear relationship between the variables. The economic interpretation of the these relationships requires an *a priori* economic theory. Cointegration implies the existence of a dynamic error correction model, which again must be interpreted with the aid of economic theory. Hypothesis testing on the cointegration parameters, of $I(1)$ variables, is possible although not standard. Cointegration is currently one of the most active research areas in time series econometrics and innovative results are frequently appearing in journals. We have provided an overview of the basic ideas in this area that are likely to be of use to the applied economist.

Appendix: The Johansen procedure

Johansen (1988) sets his analysis within the following framework. Begin by defining a general polynomial distributed lag model of a vector of variables X as

$$X_t = \pi_1 X_{t-1} + \ldots + \pi_k X_{t-k} + \varepsilon_t \qquad t = 1, \ldots, T \qquad \text{(A1)}$$

where X_t is a vector of N variables of interest; π_i are NXN coefficient matrices, and ε_t is an idependently identically distributed N-dimensional vector with zero mean and covariance matrix Ω. Within this framework the long-run, or cointegrating matrix is given by

$$I - \pi_1 - \pi_2 \ldots - \pi_k = \pi \qquad \text{(A2)}$$

where I is the identity matrix.

π will therefore be an NXN matrix. The number, r, of distinct cointegrating vectors which exists between the variables of X, will be given by the rank of π. In general, if X consists of variables which must be differenced once in order to be stationary [integrated of

order one of $I(1)$] then, at most, r must be equal to $N - 1$, so that $r \leq N - 1$. Now we define two matrices α, β both of which are $N \times r$ such that

$$\pi = \alpha\beta'$$

and so the rows of β form the r distinct cointegrating vectors.

Johansen then demonstrates the following theorem.

Theorem: The maximum likelihood estimate of the space spanned by β is the space spanned by the r canonical variates corresponding to the r largest squared canonical correlations between the residuals of X_{t-k} and ΔX_t corrected for the effect of the lagged differences of the X process. The likelihood ratio test statistic for the hypothesis that there are at most r cointegrating vectors is

$$-2 \ln Q = -T \sum_{i=r+1}^{N} \ln(1 - \hat{\lambda}_i) \qquad \text{(A3)}$$

where $\hat{\lambda}_{r+1} \ldots \hat{\lambda}_N$ are the $(N - r)$ smallest squared canonical correlations. Johansen then goes on to demonstrate the properties of the maximum likelihood estimates and, more importantly, he shows that the likelihood ratio test has an asymptotic distribution which is a function of an $(N - r)$ dimensional Brownian motion which is independent of any nuisance parameters. This means that a set of critical values can be tabulated which will be correct for all models. He demonstrates that the space spanned by β is consistently estimated by the space spanned by $\hat{\beta}$.

In order to implement this theorem we begin by reparameterising (A1) into the error correction model:

$$\Delta X_t = \Gamma_i \Delta X_{t-1} + \ldots + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_K X_{t-k} + \varepsilon_t \qquad \text{(A4)}$$

where

$$\Gamma_i = -I + \pi_1 + \ldots \pi_i; \qquad i = 1 \ldots k$$

The equilibrium matrix π is now clearly identified as $-\Gamma_k$.

Johansen's suggested procedure begins by regressing ΔX_t on the lagged differences of ΔX_t which yields a set of residuals R_{ot}. We then regress X_{t-k} on the lagged differences ΔX_{t-j} which yields residuals R_{kt}. The likelihood function, in terms of α, β and Ω is then proportional to

$$L(\alpha, \beta, \Omega) = |\Omega|^{-T/2} \exp[-1/2 \sum_{t=1}^{T} (R_{ot} + \alpha\beta' R_{kt})' \Omega^{-1} (R_{ot} + \alpha\beta' R_{kt})] \qquad \text{(A5)}$$

If β were fixed we could maximise over α and Ω by a regression of R_{ot} on $-\beta' R_{kt}$ which gives

$$\hat{\alpha}(\beta) = - S_{ok}\beta(\beta' S_{kk}\beta)^{-1} \quad \textbf{(A6)}$$

and

$$\hat{\Omega}(\beta) = S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko} \quad \textbf{(A7)}$$

where

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt} \qquad i, j = 0, k$$

and so maximising the likelihood function may be reduced to minimising

$$|S_{oo} - S_{ok}\beta(\beta' S_{kk}\beta)^{-1}\beta' S_{ko}| \quad \textbf{(A8)}$$

It may be shown that (A8) will be minimised when

$$|\beta' S_{kk}\beta - \beta' S_{ko}SS_{oo}{}^{-1}S_{ok}\beta|/|\beta' S_{kk}\beta| \quad \textbf{(A9)}$$

attains a minimum with respect to β.

We now define a diagonal matrix D which consists of the ordered eigenvalues $\lambda_1 > \ldots > \lambda_N$ of $(S_{ko}S_{oo}^{-1}S_{ok})$ with respect to S_{kk}. That is λ_i satisfies

$$|\lambda S_{kk} - S_{ko}S_{oo}{}^{-1}S_{ok}| = 0 \quad \textbf{(A10)}$$

Define E to be the corresponding matrix of eigenvectors so that

$$S_{kk}ED = S_{ko}S_{oo}{}^{-1}S_{ok}E \quad \textbf{(A11)}$$

where we normalise E such that $E' S_{kk}E = I$.

The maximum likelihood estimator of β is now given by the first r rows of E, that is, the first r eigenvectors of $(S_{ko}S_{oo}{}^{-1}S_{ok})$ with respect to S_{kk}. These are the canonical variates and the corresponding eigenvalues are the squared canonical correlations of R_{kt} with respect to R_{ot}. These eigenvalues may then be used in the test proposed in (A3) to test either for the existence of a cointegrating vector $r = 1$ or the number of cointegrating vectors $N > r > 1$.

Johansen (1988) calculates the critical values for the likelihood ratio test for the cases where $m \leqslant 5$, where $m = P - r$, and P is the number of variables in the set under consideration and r is the maximum number of cointegrating vectors being tested for.