Contents

Chapter 11

Cointegration

11.1 Introduction

Equilibrium relationships in economics are based on the belief that, at least in the long run, certain combinations of economic variables should not diverge from each other. Economic variables may drift apart in the short run or according to seasonal factors, but if they continue to drift too far apart in the long run, then economic forces such as market mechanisms or government intervention will bring them back together again. Examples of such variables include: interest rates on assets of different maturities, prices of commodities in different parts of the country, income and expenditure by local governments and the value of sales and production costs of an industry. Other possible examples may be wages and prices, imports and exports, money supply and prices and spot and future prices of a commodity. In some cases, economic theory involving equilibrium concepts might suggest close relations in the long run. Of course, the theoretical long-term relationship is a question that we might wish to test. The idea behind cointegration is to allow economic models to have long-run components of variables that obey equilibrium constraints and yet also allow a flexible dynamic specification out of equilibrium. Engle and Granger (1991) have collected a set of readings on cointegration that contain many of the papers referred to in this Chapter. Two other sources are Hamilton (1994) and Dhrymes (1998).

Consider a vector, x_t, of economic variables. In an equilibrium relationship, equilibrium is a *stationary point* characterized by forces which tend to push the economy back towards an equilibrium whenever it moves away. The vector, x_t, is defined to be in equilibrium when:

$$\alpha^T x_t = 0 \tag{11.1}$$

where α is some vector of coefficients. At any particular time, x_t will not be in equilibrium and the variable z_t defined by,

$$z_t = \alpha^T x_t \tag{11.2}$$

may be thought of as the *equilibrium error*. If the concept of equilibrium is to have econometric implications, then it will be found in the time series properties of z_t.

11.2 Cointegration: A Definition

Consider two time series y_t and x_t, each of which is I(1), and having no drift or trend in the mean. In general, any linear combination of these two series is also I(1). If, however, there exists a constant, A, such that,

$$z_t = y_t - Ax_t \tag{11.3}$$

where z_t is I(0) then y_t and x_t are said to be *cointegrated*. In the bivariate case with no drift or trend, if A exists, it is unique.

If x_t and y_t are cointegrated, the time series properties of z_t, which is I(0), are quite different from the time series properties of either x_t or y_t (which are both I(1)). Since x_t and y_t are both I(1), they have low-frequency or "long-wave" or "long-memory" components. In contrast, z_t does not have long-wave components, so that y_t and Ax_t must have low-frequency components that *cancel out*.

As an illustration of the canceling effect, consider two time series with seasonal components. In general any linear combination of these series will also have a seasonal component. If, however, the seasonals are identical in shape then there may exist a linear combination which has no seasonals. We can think of other common features of combinations of variables that might cancel.

We consider a long-run or equilibrium relationship which is (hopefully) suggested by some economic theory (this may be a steady state relation) .

$$y_t = Ax_t \tag{11.4}$$

The z_t relationship given in (11.3) may be regarded as a measure of the extent to which the system x_t and y_t is *out of equilibrium*. Following Granger and Engle (1985) we call z_t the "equilibrium error". If x_t and y_t are both I(1) variables, but move together in the long run then z_t must necessarily be I(0) otherwise x_t and y_t would just drift apart. Thus, for a pair of I(1) variables, cointegration is a necessary condition for long-run equilibrium. A constant may be included in (11.3) to ensure z_t has a mean of zero.

The bivariate case can be easily extended to the case where x_t and y_t have trends in their means.

$$x_t = m_x(t) + x_t^* \tag{11.5}$$

$$y_t = m_y(t) + y_t^* \tag{11.6}$$

where x_t^* and y_t^* are both I(1) but without trends in their means. Then.

$$z_t = y_t - Ax_t = m_y(t) - bm_x(t) + y_t^* - ax_t^* \tag{11.7}$$

For z_t to be I(0) with the existence of some long-run equilibrium, then

$$m_y(t) = bm_x(t), \forall t \tag{11.8}$$

and hence x_t^* and y_t^* are cointegrated. If $m_x(t)$ and $m_y(t)$ are different functions of time, say $m_x(t)$ is a cubic and $m_y(t)$ is an exponential, then (11.8) can not hold.

If x_t is I(0) and y_t is I(1) then the regression

$$y_t = \beta x_t + u_t \tag{11.9}$$

makes no sense at all, since the dependent variable and the independent variable have very different time series properties. Theoretically, the only reasonable value is

$$p \lim \hat{\beta} = 0.$$

11.3 Error Correction Models (ECM)

Granger (1983) and Granger and Engle (1985) have shown that if x_t and y_t are both I(1), without trends in mean, and are cointegrated, then there always exists an **error-correcting form** (ECM). This is called the *Granger Representation Theorem* (see Hamilton, 1994). The error correction model is:

$$\begin{aligned}\Delta x_t &= -\phi_1 z_{t-1} + \text{lagged}(\Delta x_t, \Delta y_t) + B_1(L)\epsilon_{1t} \\ \Delta y_t &= -\phi_2 z_{t-1} + \text{lagged}(\Delta x_t, \Delta y_t) + B_2(L)\epsilon_{2t}\end{aligned}$$

where $z_t = y_t - Ax_t$ and $B_i(L)$ $(i = 1, 2)$ are finite polynomials in the lag operator, L with roots outside the unit circle. Note ϕ_1 or ϕ_2 could be 0. The errors ϵ_{1t} and ϵ_{2t} are vector white noise.

Granger and Engle (1985) have shown that not only does a cointegrated system have an error-correcting form, but also every error-correcting model must also be cointegrated. These error correction models are attractive, because they allow long-run components of variables to obey equilibrium constraints while the model has flexible dynamics out of equilibrium. Gregory, Pagan and Smith (1992) consider the estimation of these ECM in linear quadratic cost of adjustment models. At the end of this chapter we develop the linear quadratic model in some detail which is useful for illustrative purposes as it links the econometric theory with a well-known economic model

- Keep in mind, that if z_t were known or estimated and if Δx_{t-j} and Δy_{t-j} $j = 1, \ldots$ were exogenous, then ordinary least squares in (11.10) is appropriate, yielding $\sqrt{T}$ consistent estimators. This is because all variables in (11.10) would be I(0).

11.4 Some Implications of Cointegration

1. If x_t and y_t are cointegrated, so will x_t and $y_{t-k}+w_t$, for any k where $w_t \sim I(0)$.
2. If x_t and y_t are I(1) and cointegrated then there must be **Granger causality** in at least one direction. This is an immediate implication from the Granger Representation Theorem. Thus one variable can always help forecast the other.

11.5 Asymptotics of the Least Squares Regression

11.5.1 Under the Null Hypothesis of no Cointegration

Under the null hypothesis of no cointegration the alternative hypothesis is cointegration. Consider the following linear regression.

$$y_t = \alpha + \beta x_t + u_t, \qquad \mathbf{u}_t \sim \mathbf{I(1)},\ t = 1, 2, \ldots, T \tag{11.10}$$

OLS on (11.10) yields

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t \tag{11.11}$$

Regression asymptotics under the null hypothesis of no cointegration (called the **spurious regression** case) on (11.10) have been studied by Phillips (1986) and Phillips (1998). The following results hold:

1. $\hat{\beta}$ does not converge to the true β (but it does not diverge either)

$$\hat{\beta} = \left(\frac{x^T x}{T}\right)^{-1} \frac{x^T y}{T} \text{ is } O_p(1) \Rightarrow \hat{\beta} \text{ does not diverge}$$

1. $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} \to \infty$ as $T \to \infty \Leftrightarrow \hat{\alpha}$ diverges.
2. Estimated variances diverge

$$s^2 = 1/T \sum_{t=1}^{T} u_t^2 \to \infty \text{ as } T \to \infty \Leftrightarrow s^2 \text{ diverges}$$

3. $s^2_{\beta_i} = \frac{s^2}{T}(x^T x)^{-1}_{ii} \to 0$ as $T \to \infty$.
4. $t_{\beta_i} = \frac{\hat{\beta}}{s_{\beta_i}} \to \infty$ as $T \to \infty$ (everything looks significant)

5. R^2 is well defined.

6. $F = \frac{T-m-1}{m}\frac{R^2}{(1-R^2)} \to \infty$ as $T \to \infty$.

7. $DW = \frac{\sum_{t=1}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2} \to 0$

8. $T \cdot DW$ has a well-defined limiting distribution.

11.5.2 Under Alternative of Cointegration

In this case we have

$$y_t = Ax_t + z_t, \quad \mathbf{z}_t \sim \mathbf{I}(\mathbf{0}),\ t = 1, 2, \dots, T \tag{11.12}$$

- If y_t and x_t are both $I(1)$, x_t is **not weakly exogenous** and $z_t = y_t - Ax_t$ is I(0), then the ordinary least squares estimator of A obtained by regressing y on x, $\hat{A}$, must be normalized by T (see Chapter 10 for motivating this) to have a non-degenerate distribution. Thus

$$T(\hat{A} - A) \sim D(\mu, V) \tag{11.13}$$

  where μ and V are the mean and variance of some non-normal limiting distribution D. This is established in Stock (1987).

- The fact that $\mu \neq 0$ does not imply $\hat{A}$ is inconsistent (it is consistent), since the normalization is T rather than $\sqrt{T}$.

- For I(0) variables, coefficients vary around their probability limits by $O_p(T^{-1/2})$. For cointegrated variables, $\hat{A}$ differs (has a bias) from A by $O_p(T^{-1})$.

- This bias is due to the correlation of the $x_t's$ and the errors from (11.12) (although it is not the usual problem where we need an instruments since consistency already holds).

- Since Ax_t is nonstationary, this correlation results in a nonzero limiting mean rather than in estimator inconsistency. We will return to the discussion of estimating the cointegrating vector with **endogenous** regressors later.

- Note if x is weakly exogenous (meaning it is uncorrelated with the error term in the cointegrating regression) then $T(\hat{A} - A)$ has an asymptotic normal limiting distribution (however, in most economic applications is difficult to sustain the exogeniety of regressors)

11.6 Testing for Cointegration

Consider the following linear regression model

$$y_t = \alpha + \beta x_t + u_t \tag{11.14}$$

If there is no cointegration then u_t is I(1). Thus the null hypothesis of **no cointegration** between the variables in (11.14) can be tested by **testing for a unit root in the residuals** from OLS estimation of (11.14). The procedure is very simple.

$$\begin{aligned} H_0 &: y_t \text{ and } x_t \text{ are not cointegrated } (\textbf{spurious regression}). \\ H_1 &: y_t \text{ and } x_t \text{ are cointegrated } (\alpha \text{ and } \beta \text{ are the cointegrating vectors}) \end{aligned}$$

If α and β are unknown in (11.14), we follow these steps.
Step 1
Run OLS on (11.14) and obtain the residuals, $\hat{u}_t$.
Step 2
Consider the linear regression

-
$$\hat{u}_t = c + \phi \hat{u}_{t-1} + \sum_{i=1}^{p} \gamma_i \bigtriangleup \hat{u}_{t-i} + \text{ errors} \tag{11.15}$$

- We wish to test whether $\phi = 1$. If $\phi = 1$ (unit root) then $\hat{u}_t$ is I(1), and x_t and y_t are not cointegrated.

- If $\phi < 1$ and u_t is I(0), then x_t and y_t are cointegrated. As we saw in Chapter 10 this requires **uncorrelated errors** in (11.15).

- If α and β are known (unlikely in practice) in (11.14) then a unit root in (11.15) can be tested directly using the Dickey-Fuller test or any of the other tests mentioned in Chapter 10. Since the errors in (11.14) may be serially correlated we proceed as before and estimate (augmented Dickey-Fuller, ADF). When α and β are known, then we may use the Table from Fuller (1976), p.373.

- In the more usual situation with unknown α and β, comparisons are against tables in Engle and Yoo (1987) p. 157-158 or MacKinnon (1991).

- In either case, we check to see if the t statistic is negative and statistically significant. A large negative statistic implies a rejection of the unit root and hence a rejection of no cointegration in (11.14). In practice, we would normalize (11.14) and run the test the other way around with x_t serving as the dependent variable.

- The test is **not invariant** to which variable serves as the dependent variable (see Phillips and Ouliaris, 1990).

11.6.1 Important Consideration

- While it is true that much of the early literature on cointegration considered only the bivariate case, the concept of cointegration can be extended to more than two variables.

- The problem is finding the appropriate critical values with which to compare our tests. Again the nuisance parameter problem emerges, since **the distribution of the test statistic depends on the number of regressors.** That is, we must use critical values appropriate to the number of regressors, constant and trends in the regression equation that we wish to test for cointegration.

- Critical values for Dickey and Fuller type tests for cointegrated systems of up to 8 right-hand side variables can be found in Engle and Yoo (1987) and MacKinnon (1991).

- In order to determine the robustness of the cointegrating test, we should always consider the different possible normalizations from the stated model. For example, if we were investigating whether or not 5 variables were cointegrated, we may run five different regressions each with a different dependent variable and test the residuals from each regression for a unit root.

- For these so-called residual based tests, there is a normalization decision where as in the system approach discussed below the normalization is not necessary. Nevertheless as we shall see, this does come with the added difficulty that the estimated coefficients are not identified and therefore inference requires an additional restriction or normalization (the test does not require this normalization)

- Of course, this may lead to conflicts in tests and to nontransitivities. Gregory and Haug (1998) have shown that under the null hypothesis of no cointegration, that many of the more popular tests for cointegration are almost uncorrelated.

11.7 ECM and Two-Step Estimation

Granger and Engle (1987) have shown that not only does a cointegrated system have an error-correction model, but also every error-correction model must be cointegrated. These error-correction models are attractive because they allow the long-run components of variables to obey equilibrium constraints, while the model has a flexible dynamics out of equilibrium. The two-step procedure is straightforward and yields estimated parameters which converge to their population counterparts at faster rates then standard stationary econometric estimates.

11.7.1 An Illustrative Example

As an example of the Granger and Engle two-step procedure, consider two time series variables such as real consumption, c_t, and income y_t.

Testing for I(1) and Form the Cointegrating Equation

- In the bivariate case, which we are discussing there are two possible cointegrating regressions; c_t on y_t and y_t on c_t. We find that for Canadian data, the OLS linear regression of c_t on y_t and a constant yields an R^2 =.99458

  $$c_t = \alpha_1 + \alpha_2 y_t + ec_t,\ ec_t \sim (0, \sigma_c^2)$$

  The ADF tests for cointegration is:

  $$\Delta \hat{e}c_t = \phi_1 \hat{e}c_{t-1} + lagged(\Delta \hat{e}c_t) + errors \tag{11.16}$$

- A linear regression of y_t on c_t and a constant yields an $R^2 = .99426$.

  $$y_t = \beta_1 + \beta_2 c_t + ey_t,\ ey_t \sim (0, \sigma_y^2)$$

  The ADF tests for cointegration is:

  $$\Delta \hat{e}y_t = \phi_2 \hat{e}y_{t-1} + lagged(\Delta \hat{e}y_t) + errors \tag{11.17}$$

  Cointegration is established using the ADF test on the residuals $\hat{e}c_t$ and $\hat{e}y_t$.

- In both (11.16) and (11.17) cointegration is established by noting that $\hat{\phi}_1$ and $\hat{\phi}_2$ both had negative and significant t statistics. Since the regression of c_t on y_t yields the highest R^2 (slightly) we might consider the linear regression of c_t on y_t as a candidate specification of the cointegrating regression (small sample evidence suggests this leads to tests with better properties). It also seems the most sensible from an economic point of view.

Form the Error-Correction Model

We enter the residuals $\hat{e}c_t$ into a standard ECM.

- $$\Delta c_t = \alpha + \beta \hat{e}c_{t-1} + \sum_{j=1}^{4} \delta_j \Delta c_{t-j} + \sum_{j=1}^{4} \gamma_j \Delta y_{t-j} + \text{errors} \tag{11.18}$$

- By looking at the t statistics on the estimated coefficients in (11.18) we can continue simplifying the ECM until the residuals from the ECM are white noise and all the coefficients are individually (jointly) statistically significant. The choice of 4 lags should, of course, be checked. The estimated coefficient $\hat{\beta}$ should be negative and significant reflecting a movement back to equilibrium whenever $(c_{t-1} - y_{t-1})$is positive.

- While normality is asymptotically justified in the estimation of (11.18), Monte Carlo evidence suggest that it will be a poor guide for finite sample inference.

- Monte Carlo evidence in Gregory (1994) and Gregory, Nason, and Watt (1996) indicates that tests for cointegration may be biased towards the null of not rejecting the null of no cointegration when it is false (either due to poor power or the presence of structural breaks in the cointegrating relations). These results suggests a flaw in the strategy of **pretesting** for cointegration and then estimating a cointegrating relation or ECM model **only if** the null hypothesis of cointegration is rejected.

11.8 Linear Quadratic Models and Cointegration (optional)

The following discussion is taken from Gregory (1994), Gregory, Nason and Watt (1996) and Gregory, Pagan and Smith (1992). and is useful for placing cointegration in the context of a familiar and easy economic model.

The goal is to:

1. We review the relationship between the linear quadratic model and cointegration.

2. Outline many of the tests for cointegration currently available in the literature

3. Conclude the chapter with some estimation procedures that lead to χ^2 inference.

The linear quadratic model is a popular and tractable dynamic model in which agents minimize a multi-period quadratic cost function (see Sargent, 1987).

- Agents are assumed to track the long-run target variable y_s^*, as given by a static equilibrium theory and choose the actual y_s so as to minimize the weighted sum of the costs of being away from equilibrium $(y_s - y_s^*)$ and the costs of adjustment $(y_s - y_{s-1})$. The formal problem is:

$$\min_{\{y_s\}} E_t \sum_{s=t}^{\infty} \beta^{s-t} \left[\delta(y_s - y_s^*)^2 + (y_s - y_{s-1})^2\right] \tag{11.19}$$

  for $s \geq t$, where the expectation is taken with respect to information available to the agent at time t $(\mathbb{F}_t)$, $\beta \in (0,1)$ is a discount factor and $\delta > 0$ is a weighting factor.

- The static equilibrium relationship is

$$y_t^* = x_t^T \theta + e_t \tag{11.20}$$

  where e_t is a mean zero, independently and identically distributed error with variance σ_e^2, and x_t is a $(k \times 1)$ vector of forcing variables.

- We assume that e_t is in $\mathbb{F}_t$ but unknown to the investigating econometrician (to avoid an estimation singularity) whose information set is $\mathbb{G}_t \subset \mathbb{F}_t$.

- The forward solution to (11.19) is:

$$y_t = \lambda y_{t-1} + (1-\lambda)(1-\beta\lambda)E_t \sum_{s=t}^{\infty} (\beta\lambda)^{s-t} y_s^*, \tag{11.21}$$

  where $\lambda < 1$ is the stable root of the Euler equation obtained from the first-order conditions.

- The Wiener-Kolmogorov prediction formula can be used to replace the expectations in (11.21) given the law of motion for the forcing variable (see Sargent, 1987).

- We shall be concerned with the case where x_t is a $(k \times 1)$ vector of integrated processes of order 1 denoted $I(1)$:

$$(I-L)R(L)x_t = \varepsilon_t , \tag{11.22}$$

  where $\{\varepsilon_t\}$ is independently and identically distributed with a mean of 0 and variance of Σ and the roots of

$$R(L) = I - R_1 L - \ldots - R_p L^p \tag{11.23}$$

  lie outside the unit circle (only stationary AR models of finite order).

- To simplify the description of the solution of the model we will assume that x_t is scalar $(k = 1)$. Given the stochastic process for x_t in (11.22), equation (11.21) can be solved.

- For instance if $\Delta x_t = \varepsilon_t$ $(k = 1)$ then the error correction model (ECM) can be obtained as:

$$\Delta y_t = (\lambda - 1)(y_{t-1} - \theta x_{t-1}) + (1-\lambda)\theta\Delta x_t + (1-\beta\lambda)(1-\lambda)e_t \tag{11.24}$$

- Alternatively with $\Delta x_t = \rho\Delta x_{t-1} + \varepsilon_t$ and $\rho < 1$, then:

$$\Delta y_t = (\lambda - 1)(y_{t-1} - \theta x_{t-1}) + \frac{(1-\lambda)\theta}{(1-\rho\lambda\beta)}\Delta x_t + (1-\beta\lambda)(1-\lambda)e_t \tag{11.25}$$

- In general, the solution will depend upon the serial correlation properties of Δx_t . However regardless of the exact nature of (11.25), the following relation always holds:

$$y_t = \theta x_t + \eta_t , \qquad t = 1, \ldots , T \tag{11.26}$$

  where η_t is a stationary error.

- Hence y_t and x_t are cointegrated and θ is the cointegrating vector.

- Tests for cointegration using (11.26) are then applied as a weak test of the model (11.19) without having to know the precise form of (11.22) with specific solutions given above.

- The general form for η_t $(k = 1)$ is:

$$\eta_t = [\Psi(L)\lambda/(1-\lambda L)]\varepsilon_t + [\delta\lambda/(1-\lambda L)]e_t , \tag{11.27}$$

  where $\Psi(L)$ depends upon the nature of x_t in (11.22).

- For instance

$$\Delta x_t = \varepsilon_t \Rightarrow \Psi(L) = -\theta \tag{11.28}$$

  and for

$$\Delta x_t = \rho\Delta x_{t-1} + \varepsilon_t \Rightarrow \Psi(L) = -\theta(1-\rho\beta)/[(1-\rho\beta\lambda)(1-\rho L)]. \tag{11.29}$$

- We can make some predictions regarding the tests for cointegration in the linear quadratic model by noting the relationship between the relative cost parameter δ and the stable root λ.

- The stable root $\lambda < 1$ satisfies:

$$\lambda^2\beta + 1 = \lambda + \lambda\beta + \lambda\delta, \tag{11.30}$$

  where $\lambda \to 1$ as $\delta \to 0$.

- That is, as the cost of adjustment gets large (a small δ), the stable root approaches 1 and η_t in (11.26) is nearly integrated. In these circumstances, we might expect that tests for cointegration in linear quadratic models (like augmented Dickey-Fuller, see also Phillips and Ouliaris, 1990 and references therein) would encounter difficulties in detecting a cointegrated relation like (11.26) when the stable root (high cost of adjustment) is near unity.

- Despite the fact that such tests are asymptotically appropriate with serially correlated errors, finite sample evidence in Schwert (1989) for unit root tests suggest a lack of power if η_t is nearly integrated. Unfortunately, applied work has yielded point estimates for the root that have typically been 0.9 or greater

- It is also quite clear that systems approaches like Johansen (1988 and 1990), Phillips and Durlauf (1986), Stock and Watson (1988), and Phillips and Ouliaris (1988 and 1990) may suffer similar problems to the single equation methods when the stable root is near unity.

- The procedures of Johansen (1988 and 1990) and Stock and Watson (1988) examine the system of equations in vector autoregressive form which for independent Δx_t is:

$$\begin{aligned} y_t &= \lambda y_{t-1} + (1-\lambda)\theta x_{t-1} + (1-\lambda)\theta\varepsilon_t + (1-\beta\lambda)(1-\lambda)e_t \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned}$$

$z_t = (y\ , x\ )^T$. The tests for cointegration of Johansen (1988 and 1990) and Stock and Watson (1988) maybe thought of:

$$z_t\ = Rz_{t-1} + \nu_t\ , \tag{11.31}$$

and test whether $R = 1$.

- Clearly from (11.27) as $\lambda \to 1$ this restriction is closer to being true and the tests should do poorly.

- Another multivariate approach due to Phillips and Ouliaris (1988 and 1990) is to examine the long-run covariance matrix of Δz_t (the spectrum of Δz_t at frequency zero) for singularities. With independent Δx_t we have:

$$z_t = \begin{bmatrix} \triangle y_t \\ \triangle x_t \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{1-\lambda L}\right)(1-\lambda)(1-\beta\lambda)(e_t - e_{t-1}) + \left(\frac{1}{1-\lambda L}\right)(1-\lambda)\theta\varepsilon_t \\ \varepsilon_t \end{bmatrix} \tag{11.32}$$

- If $E[e_s\varepsilon_t\ ] = 0$, for all s and t, the long-run covariance matrix of Δz_t denoted by Ω is (with the variance of Δx_t equal to σ_ε^2):

$$\Omega = \begin{bmatrix} \theta^2\sigma_\varepsilon^2 & \theta\sigma_\varepsilon^2 \\ \theta\sigma_\varepsilon^2 & \sigma_\varepsilon^2 \end{bmatrix} \tag{11.33}$$

which is clearly singular. From (11.27) it is also evident that the closer is λ to one, the more difficult it will be to estimate the long-run relation and hence to detect such singularities as (11.33).

11.9 Tests for Cointegration

Critical values for the various tests for cointegration discussed in this section can be found in various places in the literature. James MacKinnon has a link on his home page for computing some of these.

11.9.1 Augmented Dickey Fuller (ADF) Test:

- The most widely used cointegration test is the augmented Dickey-Fuller (ADF) t-ratio test (see Said and Dickey, 1984), recommended by Engle and Granger (1987). Its asymptotic properties have been studied by Phillips and Ouliaris (1990).

- The test is based on the residuals from a cointegrating regression and is constructed to test the null hypothesis of no cointegration by testing the null of a unit root in the residuals against the alternative that the root is less than unity.

- One first estimates the cointegrating regression (11.26) by ordinary least squares (OLS) and tests the null hypothesis of no cointegration using a scalar unit root test $t(\hat{\alpha})$ on the residuals:

$$\Delta\hat{\eta}_t = \hat{\alpha}\hat{\eta}_{t-1} + \sum_{i=1}^{m}\hat{\phi}_i\Delta\hat{\eta}_{t-i} + \hat{\nu}_t \ , \tag{11.34}$$

  where the lag length m is chosen sufficiently large in order for $\hat{\nu}$ to be serially uncorrelated.

- The distribution of $t(\hat{\alpha})$ depends upon the number of regressors in (11.26) with asymptotic critical values provided in Engle and Yoo (1987), MacKinnon (1991) and Phillips and Ouliaris (1990).

- In STATA you can use the $dfuller$ command but be sure to take care not to use the critical values printed out as these are for the unit root test

11.9.2 Phillips's Z_α and Z_t Test

- Closely related to the ADF tests are those suggested by Phillips (1987) and Phillips and Perron (1988). Equation (11.26) is estimated by OLS, the residuals are obtained and the following test regression is run:

$$\hat{\eta}_t = \hat{\alpha}\hat{\eta}_{t-1} + \hat{\zeta}_t \ . \tag{11.35}$$

- Again we test the unit root hypothesis on the residuals. The test statistics are:

$$Z_\alpha = T(\hat{\alpha}-1) - \frac{1}{2}\left[\hat{\omega}_\zeta^2 - \hat{\sigma}_\zeta^2\right]\left(T^{-2}\sum_{t=2}^{T}\hat{\eta}_{t-1}^2\right)^{-1} \tag{11.36}$$

  and

$$Z_t = \left[\frac{(\hat{\alpha}-1)(\sum_{t=2}^{T}\hat{\eta}_{t-1}^2)^{\frac{1}{2}}}{\hat{\omega}_\zeta}\right] - \frac{1}{2}\left[\hat{\omega}_\zeta^2 - \hat{\sigma}_\zeta^2\right]\left(T^{-2}\sum_{t=2}^{T}\hat{\eta}_{t-1}^2\right)^{-\frac{1}{2}} \tag{11.37}$$

  where $\hat{\sigma}_\zeta^2 = T^{-1}\sum\hat{\zeta}_t^2$ and $\hat{\omega}_\zeta^2$ is an estimator of the spectrum of ζ at frequency zero (the long-run variance-see Chapter 10 for a discussion of these tests in the context of testing for unit roots).

- This can be estimated by the Newey-West HAC estimator (see also Andrews, 1991 and Andrews and Monahan, 1990).

- The critical values for the limiting distribution of (11.36) and (11.37) again depend upon the number of regressors k and can be found in Phillips and Ouliaris (1990).

- Z_t has the same limiting distribution as the ADF.

- In STATA you can use the *pperon* command but be sure to take care not to use the critical values printed out as these are for the unit root test

11.9.3 Stock and Watson's Minimum Eigenvalue Test

- The next two multivariate tests (Stock and Watson; 1988 and Johansen; 1988 and 1990) are especially useful in determining the number of cointegrating relations in situations where the researcher does not wish to assume that the x's ($k > 1$) themselves are not cointegrated.

- This generality certainly is an advantage over the other tests where the only possible cointegrating relation is between the y_t and the x_t via (11.26).

- Let $z_t = [y_t \ , x_t^T]^T$ and estimate:

$$z_t = \hat{\Pi} z_{t-1} + \hat{\nu}_t \ . \tag{11.38}$$

- Obtain an estimate of $V = \sum_{i=1}^{\infty} E\left[\nu_t \nu_{t-i}\right]$, say, by adjusting the Newey-West estimator and find the $k+1$ vector of eigenvalues from:

$$\left[T^{-2} \sum_{t=2}^{T} z_t z_{t-1}^T - T^{-1} \hat{V}^T\right] \left[T^{-2} \sum_{t=2}^{T} z_t z_{t-1}^T\right]^{-1}$$

- Let $\hat{\Lambda}_{\min}$ be the minimum real of that vector. Under the null of no cointegration the estimated minimum eigenvalue should be insignificantly different from one. The test statistic suggested by Stock and Watson (1988) is:

$$SW = T(\hat{\Lambda}_{\min} - 1). \tag{11.39}$$

- The critical values for SW depend upon the dimension of z and are in Stock and Watson (1988).

11.9.4 Johansen's Likelihood Ratio Test

- A closely related test to (11.39), and very widely used (programmed extensively in STATA) derived by Johansen (1988 and 1990) is a likelihood ratio test for cointegration.

- This is not to say that the other tests presented could not be viewed as likelihood ratio tests. However, since the Johansen test is so firmly ground in likelihood theory, we have reserved this label for these tests.

- Suppose the data generating process for z_t may be written as a $m^{th}-$ order vector autoregression:

$$z_t = \Pi_1 z_{t-1} + \Pi_2 z_{t-2} + \cdots + \Pi_m z_{t-m} + \psi_t \ , \tag{11.40}$$

  where ψ_t is independent mean zero with a constant covariance.

- We may rewrite (11.40) as:

$$\triangle z_t = \Gamma_1 \triangle z_{t-1} + \Gamma_2 \triangle z_{t-2} + \cdots + \Gamma_{m-1} \triangle z_{t-m+1} + \Gamma_m z_{t-m} + \psi_t$$

  where

$$\begin{aligned} \Gamma_i &= -I + \Pi_1 + \ldots + \Pi_i, \ i = 1, \ldots m \\ \Gamma_m &= I - \Pi_1 - \ldots - \Pi_m. \end{aligned}$$

- The intuition behind the test is simply to test the rank of Γ_m.

- If Γ_m has rank of zero then the null hypothesis of no cointegration is retained. This means all variables are $I(1)$ and none are cointegrated.

- If Γ_m has full rank of $K+1$ (or more generally all variables are stationary in the system)

- The test of Johansen (1988) is a likelihood ratio obtained as (called the trace test):

$$Trace = -T \sum_{i=1}^{k+1} \ln(1 - \hat{\Lambda}_i), \tag{11.41}$$

  where $\hat{\Lambda}_i$ are the eigenvalues from solving (called squared partial canonical correlations or reduced rank regression):

$$\mid \Lambda S_{mm} - S_{m0} S_{00}^{-1} S_{0m} \mid = 0 \tag{11.42}$$

  where we define $z_{0t} = \Delta z_t$, $z_{1t} = [\Delta z_{t-1}, \ldots, \Delta z_{t-m+1}]$, $z_{tm} = z_{t-m}$ and the following moment relations:

$$\begin{aligned} M_{ij} &= T^{-1} \sum_{t=1}^{T} z_{it} z_{jt} \quad i, j = 1, \ldots m \\ S_{ij} &= M_{ij} - M_{i1} M_{11}^{-1} M_{j1} \quad i, j = 0, m \end{aligned}$$

- The critical values depend on the number of regressors and are in Johansen (1988) and Hamilton (1994).

- As mentioned above Johansen's tests may used to test for other possible cointegrating relations in the vector x_t .

- Test (11.41) is, strictly speaking, not directly comparable to the other tests in this study which only allow for one cointegrating relation between y_t and the vector x_t. However Johansen (1990) and Johansen and Juselius (1990) suggest a test which is a special case of (11.41) designed to test whether there is one cointegrating vector in a system of equations.

- It involves only the maximum eigenvalue in the vector $\hat{\Lambda}$ in (11.41):

$$Max = -T \ln(1 - \hat{\Lambda}_{\max}) \tag{11.43}$$

- Even in this case the residual based the researcher chooses a y variable and a set of x variables for the test under the presumption that all x are $I(1)$ and not cointegrated and that the only possible cointegration is between the y and x

- The maximum eigenvalue test selects the relationship that yields the greatest chance of being cointegrated among y and x

- Compare also (11.43) with the Stock-Watson test (11.39).

- In order to be appropriate for the critical values supplied in Johansen and Juselius (1990, Table A2) a constant needs to be added to (**??**), so that z_{1t} would also have a one in it.

- We develop more fully these system tests and estimators at the end of the Chapter (and link it to the STATA estimation, inference procedures)

11.9.5 Park, Ouliaris and Choi's Spurious Regressors Test

- Park, Ouliaris and Choi (1988) have developed a variable addition test in which additional regressors (powers of time trends) are added to a (potentially) cointegrating regression.

- If the variables do indeed define a cointegrating relation then additional variables should have no explanatory power. On the other hand, if the regression is spurious (no cointegration), results from Phillips (1986) indicate that F tests on additional trend terms should diverge.

- Park, Ouliaris and Choi (1988) obtain a limiting distribution by dividing the usual F test by the sample size. Unlike all the other tests of cointegration discussed in this paper, the Park, Ouliaris and Choi (1988) test can be formulated with a null hypothesis of no cointegration or a null of cointegration (see also Hansen, 1991).

- This test has not been widely applied but is again useful to assist our understanding of cointegration and spurious regression

- However, since we wish to link this test with the others in this Chapter we choose the formulation with the null hypothesis of no cointegration.
- Following the motivation as an F test, the unrestricted least squares regression is:

$$y_t = \sum_{i=0}^{q} \hat{\alpha}_i t^i + x^T \hat{\theta} + \hat{\eta}_t \ , \tag{11.44}$$

and denote the residual sum of squares as RSS_q.

- The Park, Ouliaris and Choi test statistic is:

$$J(0, q) = (RSS_0 - RSS_q)/RSS_q, \tag{11.45}$$

where RSS_0 is the (restricted) residual sum of squares from regressing y_t on x_t and a constant.

- Critical values are given in Table 1 of Park, Ouliaris and Choi (1988) and again depend upon the number of regressors.

11.9.6 Hansen's Cochrane-Orcutt Technique (optional)

- All of the tests for cointegration share the feature that the limiting distribution of the test statistic depends on the number of regressors in the cointegrating relation.
- The tests proposed by Hansen (1990) are based upon the Cochrane-Orcutt technique and yield limiting distributions which are invariant to the number of regressors.
- The Hansen test found no favour in the applied literature since it had zero power in some cointegration directions!

- We consider it here again as a pedagogical tool
- Consider equation (11.26) with the following relation for the error structure:

$$\begin{aligned} y_t &= x_t^T \theta + \eta_t \\ \eta_t &= \rho \eta_{t-1} + \xi_t \end{aligned}$$

- Estimate sequentially θ and ρ by OLS. Quasi-difference the data using the estimated $\hat{\rho}$:

$$\begin{aligned} y_t^* &= y_t - \hat{\rho}\ y_{t-1} \\ x_t^* &= x_t - \hat{\rho}\ x_{t-1} \end{aligned} \tag{11.46}$$

and estimate:

$$y_t^* = x_t^{*T}\hat{\theta} + \tilde{\eta}_t \tag{11.47}$$

- We may iterate this procedure or follow some finite sample modifications suggested in Hansen (1990). Tests for unit roots can now be applied to the residual $\tilde{\eta}_t$ from the transformed estimated equation (11.47) (instead of $\hat{\eta}_t$ from (11.46).

- The advantage of these tests is that the limiting distribution does not depend on the number of regressors and the Dickey Fuller t-test and coefficient distribution for the unit root hypothesis respectively.

- The intuition behind this invariance is that the limiting distribution of $\tilde{\theta}$ in (11.47) under the null hypothesis of no cointegration converges to a constant and not a random variable as in the since under $H_0, \hat{\rho} \to 1$ and thus y_t^* and x_t^* are asymptotically first differences.

- Hansen (1990) has observed that the standard residual based tests suffer a considerable loss of power as the number of regressors increases and that tests based upon $\tilde{\eta}_t$ may be more powerful.

- In a Monte Carlo experiment Gregory (1994) considers ADF, Z_α, and Z_t which may be denoted as $HADF$, HZ_α and HZ_t respectively finds that while these Hansen tests retain power as the number of regressors gets large, they are nevertheless not as powerful as their untransformed counterparts.

11.9.7 Phillips and Ouliaris's Trace and Variance Ratio Test (optional)

- Phillips and Ouliaris (1990) suggest examining the long-run covariance matrix of Δz_t , denoted as Ω, for singularities. The tests they propose do not estimate Ω from Δz_t (which would in fact lead to an inconsistent test, see Phillips and Ouliaris, 1990, Theorem 5.3) but instead are based on the residuals from the following first-order vector autoregression:

$$z_t = \hat{\Pi} z_{t-1} + \hat{\nu}_t$$

- Let $\hat{\Omega}$ be the estimated long-run covariance matrix (the estimated spectrum of ν_t at frequency zero). If the variables are cointegrated, then there should be singularities in Ω (Phillips and Ouliaris, 1988). The trace test is:

$$P_z = T\ tr\left[\hat{\Omega}\ M_{zz}^{-1}\right], \quad M_{zz} = T^{-1}\Sigma z_t\ z_t^T, \tag{11.48}$$

and the variance ratio test is:

$$P_u = \frac{T\left[\hat{\Omega}_{yy} - \hat{\Omega}_{xy}^T \hat{\Omega}_{xx}^{-1} \hat{\Omega}_{xy}\right]}{T^{-1}\Sigma\hat{\eta}^2}, \tag{11.49}$$

where $\hat{\eta}_t$ is the residual from the cointegrating relation (11.26).

- The critical values for both (11.48) and (11.49) are in Phillips and Ouliaris (1990) and depend upon the dimensionality of z_t .
- The idea motivating (11.48) is that under the null of no cointegration there is no singularity in Ω and hence the ratio should stabilize asymptotically (the numerator is an estimate of $t\Omega$ and the denominator is the corresponding sample moment). P_u in (11.49) tests whether the conditional variance of y given x is significantly different from zero.

11.9.8 A Warning

- Gregory (1991) (see also Haug,1992) has examined a number of tests for cointegration under a variety of settings for the linear quadratic model. The results indicate sharp differences in the various tests to detect cointegrating relations especially when the cost of adjustment term is large ($\delta \to 0$) and the number of regressors is large.
- Not surprisingly when there is a stationary error that has a lot of persistence it is difficult to detect cointegration and indeed the literature has reflected this failure to reject the null
- The Monte Carlo evidence suggests that the ADF test as well as Phillips's (1987) Z_α test have proper size compared to their asymptotic values and possess the best power.

11.10 Residual Based Testing for Cointegration in Models with Regime Shits

In a paper with Hansen (Gregory and Hansen,1996 Journal of Econometrics pg 99-126) we examined a set of coiintegration tests against an alternative of cointegration with breaks in

1. Intercept
2. Trend
3. Intercept and trend
4. Regime shift

The break point (if they exist) are unknown and a sequential search is conducted.The had power against an alternative of the standard cointegration case so if one rejected the null of unit roots, the conclusion of a break did not necessarily follow. We consider 3 Models (another paper did the shft in intercept and trend)

The regular regression equation can be thought of as

$$y_{1t} = \mu + \alpha^T y_{2t} + e_t \qquad 1...T$$

Interest in the tradional case rests on testing whether e_t has a unit root with the parameters of μ and α time invariant. We show that of there is strutural change the usual test will be biased towrds a unit root and hence no cointegration

Imagine a structual change defined by the following $\tau \in (0,1)$

$$\varphi_{t\tau} = \begin{cases} 0 & if \quad t \leq [T\tau] \\ 1 & if \quad t > [T\tau] \end{cases}$$

The four alternative models are

$$\begin{aligned} y_{1t} &= \mu_1 + \mu_2\varphi_{t\tau} + \alpha^T y_{2t} + e_t \qquad 1...T \qquad \text{Level Shift (C)} \\ y_{1t} &= \mu_1 + \mu_2\varphi_{t\tau} + \beta t + \alpha^T y_{2t} + e_t \qquad 1...T \qquad \text{Level Shift with Trend (C)} \\ y_{1t} &= \mu_1 + \mu_2\varphi_{t\tau} + \alpha_1^T y_{2t} + \alpha_1^T y_{2t}\varphi_{t\tau} + e_t \qquad 1...T \qquad \text{Regime Shift} \end{aligned}$$

In each case you estimate the model as defined sequentially (using a dummy variable that is advanced over the sample). Andrews recommends trimming the data set so the interval for $\tau \in (.15, .85)$.

The test statistic is to take the minimum (the largest negative value) of

$$\begin{aligned} ADF^* &= \underbrace{\inf}_{\tau \in T} ADF(\tau) \\ Z_t^* &= \underbrace{\inf}_{\tau \in T} Z_t(\tau) \end{aligned}$$

Gregory and Hansen provide critical values for this test.

- This procedure is now implemented in Stata as an ado file ($ghansen.ado$)

11.11 Single Equation Approach to Estimation and Inference

- As we have mentioned if the x's are not weakly exogenous (are correlated with the error in the cointegrating error) then OLS estimator of the cointegrating vector leads to estimates that have non-standard distributions.
-
- We illustrate the two single-equation estimation methods:

1. Leading and lagging method
2. Fully Modified estimation (now in Stata download lrcov)
3. Canonical Correlation (is in Stata)

- See cointreg to estimate cointegration regression using fully modified ordinary least squares, dynamic ordinary least squares, and canonical correlation regression methods
- These methods produce χ^2 inference.

11.11.1 Leading and Lagging Approach to Estimating the Cointegrating Vector

- This procedure was developed independently by Saikkonen(1991), Phillips and Loretan (1991) and Stock and Watson (1993).
- There is a program called: coint_int.do which we illustrate this
- Consider the following cointegrated relation

$$y_t = \alpha + \gamma^T x_t + \eta_t \tag{11.50}$$

  with η_t stationary mean zero.

- We assume that there are no other cointegrating relations and that x_t can be written

$$x_t = x_{t-1} + u_t \tag{11.51}$$

  where u_t is stationary and the dimension of x_t is say, k.

- Define η_t^* as the residual from a linear projection of η_t on to p leads and lags of u_t :

$$\eta_t = \sum_{s=-p}^{s=p} \beta_s u_t + \eta_t^* , \tag{11.52}$$

  where η_t^* is uncorrelated (by construction) with u_{t-s} $(s = \pm 1, \ldots \pm p)$.

- We need not have assumed that (11.50) and (11.51) had no trend and drift and trend respectively. Substituting (11.52) into (11.50) forms the lead/ lag relation as:

$$y_t = \alpha + \gamma^T x_t + \sum_{s=-p}^{s=p} \beta_s \triangle x_{t-s} + \eta_t^* \tag{11.53}$$

  where p is chosen under the assumption that no *further* $(\beta_s = 0, \ \forall s > p)$ correlations exist between η_t and u_t.

- How to decide p is not an obvious matter. OLS estimation of (11.53) will yield $\chi^2(k)$ tests.
- Now suppose we wish to test:

$$\begin{aligned} H_0 &: \ \gamma = \gamma_0 \\ H_1 &: \ \gamma \neq \gamma_0 \end{aligned}$$

  then multiply the usual F-test by $\frac{\hat{\lambda}^2}{s^2}$:

$$(\hat{\gamma} - \gamma_o)^T \left(\frac{\hat{\lambda}^2}{s^2} \sum_{t=1}^{T} x_t \, x_t^T \right) (\hat{\gamma} - \gamma_o) \sim \chi^2(k) \tag{11.54}$$

where s^2 = the estimated variance of (11.53)

$$s^2 = \frac{1}{T-k}\hat{\eta}_t^{*2}$$

and regressing the residuals of (11.53) on to its lag residual:

$$\hat{\eta}_t^* = \sum_{i=1}^{p} \phi_i \hat{\eta}_{t-i}^* + e_t \tag{11.55}$$

to calculate

$$\hat{\lambda} = \hat{\sigma}\sum_{i=1}^{p}\hat{\phi}_i \qquad \hat{\sigma} = \sqrt{\frac{1}{T-p}\sum \hat{e}_t^2} \tag{11.56}$$

- Notice also by adding leads and lags of Δy_t we may eliminate some (all) of the serial correlation necessary in correction (11.55) and need not use a HAC estimator.

- That is,

$$y_t = \alpha + \gamma^T x_t + \sum_{s=-p}^{s=p} \beta_s \triangle x_{t-s} + \sum_{r=-d}^{r=d} \theta_r \triangle y_{t-r} + \epsilon_t \tag{11.57}$$

11.11.2 Hansen and Phillips' Fully Modified Approach (FM)

- Hansen and Phillips (1991) suggest an estimator that removes the asymptotic bias and results in a limiting distribution that is χ^2. The estimator is called a fully modified estimator (FM) and as is now in STATA (see *cointreg*).

- Consider again the cointegrating regression (11.26) but we now assume that the first entry in x is a constant (the model is assumed to have an intercept):

$$y_t = \theta x_t + \eta_t \tag{11.58}$$

and

$$x_t = x_{t-1} + \xi_t \ , \tag{11.59}$$

where x_t is a $(k+1)\times 1$ vector ($x_{1t} = 1$) and $\xi_t = R(L)^{-1}\varepsilon_t$.

- Define the vector $u_t = (\eta_t, \xi_t^T)^T$ and the following matrices (the long-run variance matrices):

$$\Omega = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} E\left[u_j u_t^T\right]$$

$$\Lambda = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t} E\left[u_j u_t^T\right]$$

partitioned in conformity with u:

$$\Omega = \begin{bmatrix} \Omega_{\eta\eta} & \Omega_{\eta\xi} \\ \Omega_{\xi\eta} & \Omega_{\xi\xi} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{\eta\eta} & \Lambda_{\eta\xi} \\ \Lambda_{\xi\eta} & \Lambda_{\xi\xi} \end{bmatrix} \tag{11.60}$$

Also define

$$\Omega_{\eta\cdot\xi} = \Omega_{\eta\eta} - \Omega_{\eta\xi}\Omega_{\xi\xi}^{-1}\Omega_{\xi\eta}, \quad \Lambda_{\eta\cdot\xi}^{+} = \Lambda_{\eta\xi} - \Lambda_{\xi\xi}\Omega_{\xi\xi}^{-1}\Omega_{\xi\eta}, \tag{11.61}$$

- To begin, we estimate (11.58) by OLS and obtain the residuals

$$\hat{\eta} = y \; - \hat{\gamma}x \tag{11.62}$$

and define $\hat{u}_t = (\hat{\eta} \; , \Delta x^T)^T$.

- With the $\hat{u}_t$, form estimates of Ω and Λ, denoted by $\hat{\Omega}$ and $\hat{\Lambda}$.
- We have seen how Newey and West (1987) may be used to estimate these quantities.
- Partition $\hat{\Omega}$ and $\hat{\Lambda}$ as Ω and Λ with:

$$\hat{\Omega}_{\eta\cdot\xi} = \hat{\Omega}_{\eta\eta} - \hat{\Omega}_{\eta\xi}\hat{\Omega}_{\xi\xi}^{-1}\hat{\Omega}_{\xi\eta} \quad \hat{\Lambda}_{\eta\cdot\xi}^{+} = \hat{\Lambda}_{\eta\xi} - \hat{\Lambda}_{\xi\xi}\hat{\Omega}_{\xi\xi}^{-1}\hat{\Omega}_{\xi\eta} \tag{11.63}$$

Define the transformed dependent variable:

$$y_t^{+} = y_t \; - \hat{\Omega}_{\eta\xi}\hat{\Omega}_{\xi\xi}^{-1} \triangle x_t \tag{11.64}$$

- The FM estimator of θ is:

$$\hat{\theta}^{+} = \left(\sum_{t=1}^{T} \left[y_t^{+} x_t^T - \left(0 \;\; \hat{\Lambda}_{\eta\cdot\xi}^{+} \right) \right] \right) \left(\sum_{t=1}^{T} x_t x_t^T \right)^{-1} \tag{11.65}$$

- Now suppose we wish to test:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned}$$

then

$$(\hat{\theta}^{+} - \theta_o)^T \left(\frac{1}{\hat{\Omega}_{\eta\cdot\xi}} \sum_{t=1}^{T} x_t \; x_t^T \right) (\hat{\theta}^{+} - \theta_o) \tag{11.66}$$

is distributed as χ^2 with $k+1$ degrees of freedom under H_o.

- This estimator, which has been studied in Monte Carlo experiments by Gregory, Pagan and Smith (1990), Phillips and Loretan (1991) and Stock and Watson (1991), appears to have good finite sample properties in terms of bias and coverage probabilities.

11.12 System Approach to Estimation and Inference

- In this section we develop more fully the system methods associated with Soren Johansen.
- STATA has done a number of programs based on this approach
  - *varsoc* (determining lag length in $VECM$)
  - *vec* ($VECM$ models)
  - *vec* postestimation
  - *vecmar* (LM tests for serial correlation in $VECM$)
  - *vecnorm* (Normality tests on $VECM$ residuals)
  - *vecrank* (rank tests for cointegration as well as restricted $VECM$)
  - *vecstable* (stability conditions for $VECM$)
- Consider an $(m\times1)$ vector time series process, X_t, where each X_t is individually I(1).
- Cointegration exists if some linear combination of the variables in X_t is I(0).
- The central question we want to answer is: How many different linear combinations of X_t are I(0)? Or equivalently: How many cointegrating vectors are there among the variables X_t?
- One advantage of the systems approach over the single-equation methods is that the systems approach estimates all the possible cointegrating (CI) vectors jointly. Thus if there are three CI vectors they are estimated together.
- This approach also aids in performing inference on the CI vectors since the problem of endogenous regressors, as in ingle-equation methods, does not occur.
- The systems approach begins with a VAR representation for X_t in levels:

$$X_t = \Pi_1 X_{t-1} + \ldots + \Pi_p X_{t-p} + \mu + \Phi D_t + \epsilon_t \tag{11.67}$$

  where the Π_i are $(m \times m)$ parameter vectors, D_t are seasonal dummy variables (optional) and Λ is the (mxm) covariance matrix of the errors.
- We can augment (11.67) with trends if it is desired.
- Now as the variables X_t are nonstationary, standard practice in VARs is to estimate the system with the variables entering in first differences. But in doing this we also would apply the first difference operator to ϵ_t, resulting in a loss of information.

- Another way to deal with the nonstationarity is to reparameterize (11.67) as:

$$\Delta X_t = \Pi X_{t-1} + \mu + \Phi D_t + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + \epsilon_t \tag{11.68}$$

  where

$$\Gamma_i = -(I - \Pi_1 - \ldots - \Pi_i) \qquad \forall i = 1, \ldots, p-1$$

  and

$$\Pi = -(I - \Pi_1 - \ldots - \Pi_p).$$

- The matrix Π is sometimes referred to as the **impact matrix**.

- The model (11.68) is very similar to a first difference VAR but for the appearance of the term ΠX_{t-1}, the error correction (EC) term. The presence or absence of CI among the X_t will depend upon this term. Specifically, CI will depend on the rank of the matrix Π. Inference from the rank of Π.

1. $Rank(\Pi) = m$. Since Π is (mxm) this implies that Π has full rank and that each of the X_t are stationary.

2. $Rank(\Pi) = 0$. Π is then a null matrix and there is no cointegration amongst the I(1) variables.

3. $0 < Rank(\Pi) = r < m$. Then there are r vectors of Π that result in stationary linear combinations of X_t. i.e. there are r cointegrating vectors.

- This last condition allows us to decompose Π into 2 $(m \times r)$ matrices α and β such that

$$\Pi = \alpha\beta^T. \tag{11.69}$$

  β will be a matrix containing the cointegrating vectors and α will be a matrix of weights

- In the Johansen procedure, the β can be normalized in a number of different ways (such as normalizing the length to 1 or setting one of the coefficients to 1-Stata only allows the latter). See Johansen and Juselius (1990).

- In fact, this feature is really due to the method of estimating a basis for the space, rather than a structural coefficient. Consequently, individual coefficient estimates themselves suffer some interpretation issues (see Dhrymes, 1998).

- Within this general framework we can test for the number of CI vectors (r) and test hypotheses about α and β.

11.12.1 Some Hypotheses of Interests

1. H_1: This is the model (11.67) and is maintained assumption.
2. H_2: $\Pi = \alpha\beta^T$ which is tested with deciding the order of cointegration (i.e. r).
3. H_3: $\Pi = \alpha\varphi^T H^T$ (or $\beta = H\varphi$).
4. H_4: $\Pi = A\phi\beta^T$ (or $\alpha = A\phi$).
5. H_5: $\Pi = A\phi\varphi^T H^T$ (or $\alpha = A\phi$ and $\beta = H\varphi$).

- The matrices A $(m \times k)$ and H $(m \times s)$ impose linear restrictions on α $(m \times r)$ and β $(m \times r)$ respectively , reducing the parameters to $(k \times r)$ and φ $(s \times r)$, where $r \leq k \leq m, r \leq s \leq m$.
- H_3 and H_4 are nested hypotheses of H_2 and that H_5 is a joint hypothesis of H_3 and H_4.
- Linear restrictions on the CI vectors of β, formulate the same $(m - s)$ restrictions on all the r CI vectors.
- For example, suppose you have a system with $m = 5$ variables that includes, among other variables, the price levels for two countries (p_1 and p_2) and the exchange rate between these two countries (e_{12}).
  - A reasonable hypothesis to test in this system is Purchasing Power Parity (PPP) which implies $p_1 - p_2 = e_{12}$.
  - This restriction implies that the coefficients on the vector (p_1,p_2,e_{12}) are proportional to (1,−1,−1).
  - Thus $s = 3$, the number of restrictions is 2, H will be (5×3) and φ will be $(3 \times r)$, where r is the number of cointegrating vectors.
  - The linear restrictions will then imply that the PPP relation enters each of the r CI vectors where the number of CI vectors is less than or equal to 3.
  - This last result is from the equalities just above, and is easy to understand since under the null hypothesis the two price levels and the exchange rate are effectively collapsed into one variable, leaving only three in the system.
  - Just to summarize this example, with $m = 5$, such that we have

$$X_t = (p_{1t}, p_{2t}, e_{12t}, var_{4t}, var_{5t})^T$$

and $s = 3$ the matrix H will look like:

$$H = \begin{matrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}$$

11.12.2 The Maximum Likelihood Procedure.

Estimating the Unrestricted Equation

- We want to estimate by maximum likelihood the parameters in equation (11.68).
- To do so we first introduce some new notation. Let $Z_{0t} = \Delta X_t$. Then let Z_{1t} be the stacked matrix containing $(\Delta X_{t-1}, \ldots, \Delta X_{t-p+1}, D_t,$ and $, 1)$, and finally let $Z_{pt} = X_{t-1}$.
- Notice that Z_{0t} is the variable on the left-hand side of (11.68), that Z_{1t}, is all the right-hand side variables except for X_{t-p} and X_{t-p} is represented by Z_{pt}.
- With respect to the parameter vectors we can let Γ represent the matrix of parameters for the matrix Z_{1t}, which consists of $(\Gamma_1, \ldots ,\Gamma_{p-1},\Phi,$ and $\mu)$.
- We can now express (11.68) in terms of the Z's:

$$Z_{0t} = \Gamma Z_{1t} + \Pi Z_{pt} + \epsilon_t. \tag{11.70}$$

By regressing (OLS) ΔX_t and X_{t-p} (individually) on the matrix composed of $\Delta X_{t-1}, \ldots, \Delta X_{t-p+1}, D_t,$ and 1 we can define the residuals of these two operations to be:

$$\begin{aligned} R_{0t} &= Z_{0t} - M_{01} M_{11}^{-1} Z_{1t} \\ R_{pt} &= Z_{pt} - M_{p1} M_{11}^{-1} Z_{1t}. \end{aligned}$$

where the matrices M_{ij} are the product moment matrices and are given by:

$$M_{ij} = T^{-1} \sum_{t=1}^{T} Z_{it} Z_{jt}^T \qquad \forall\ i, j = (0, 1, p). \tag{11.71}$$

- Once we have these residuals we can use them to estimate the parameters of model (11.67).
- To do this we introduce just a little more notation. Define

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}^T = M_{ij} - M_{i1} M_{11}^{-1} M_{1j} \qquad \forall\ i, j = (0, p)$$

We can then express the parameter estimates as:

$$\begin{aligned} \hat{\Pi} &= S_{0p} S_{pp}^{-1} \\ \hat{\Lambda} &= S_{00} - S_{0p} S_{pp}^{-1} S_{p0} \\ \hat{\Gamma} &= M_{01} M_{11}^{-1} - \hat{\Pi} M_{k1} M_{11}^{-1}. \end{aligned}$$

Second Set of Hypothesis: $H_2 : \alpha\beta^T$ **Determining Rank**

- The maximum likelihood estimator of β can be found by solving the equation:
$$|\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p}| = 0 \tag{11.72}$$
which will produce an $(m \times m)$ matrix of eigenvalues $\hat{\lambda}$ such that the values on the main diagonal will follow the relation $\hat{\lambda}_1 > \ldots > \hat{\lambda}_m$.
- It will also produce eigenvectors $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_m)$, which can be normalized so that $\hat{V}^T S_{pp} \hat{V} = I$.
- Now β is an $(m \times r)$ matrix, but the procedure above generates m eigenvectors.
- The estimate of $\hat{\beta}$ is the first r eigenvectors of $\hat{V}$. Thus
$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r). \tag{11.73}$$

Third Set of Hypotheses: $H_3 : \alpha\varphi^T H^T$.

- Suppose we want to estimate β under the linear restriction $\beta = H\varphi$.
- The problem is very similar to that just presented. We first must find the maximum likelihood estimate of φ.
- This can be found by solving the following equation:
$$|\lambda H^T S_{pp} H - H^T S_{p0} S_{00}^{-1} S_{0p} H| = 0 \tag{11.74}$$
which will produce an $(m \times m)$ matrix of eigenvalues $\hat{\lambda}_3$ such that the values on the main diagonal will follow the relation $\hat{\lambda}_{3,1} > \ldots > \hat{\lambda}_{3,m}$.
- As well it will produce m eigenvectors $\hat{V}_3$ such that $\hat{V}^T(H^T S_{pp} H)\hat{V} = I$.
- The estimate of φ is given by the first r eigenvectors of $\hat{V}_3$. Thus
$$\hat{\varphi} = (\hat{v}_{3,1}, \ldots, \hat{v}_{3,r}) \tag{11.75}$$
and we can derive the estimate of β as
$$\hat{\beta} = H\hat{\varphi}. \tag{11.76}$$

Fourth Set of Hypothesis: $H_4 : A\phi\beta$.

- The process of estimating α under the linear restriction $\alpha = A\phi$, is essentially no different than that already provided. If the reader wants a detailed explanation they can look to Johansen and Juselius (1990).
- Instead we define only the eigenvalues and eigenvectors that result from the process.
- The eigenvalues from this procedure will be given by the vector $\hat{\lambda}_{4,1} > \ldots > \hat{\lambda}_{4,k} > \hat{\lambda}_{4,k+1} = \ldots = \hat{\lambda}_{4,m} = 0$. We have this result due to the restrictions on α. All eigenvalues greater than the k^{th} are equal to zero.
- The estimate of β will again be the first r eigenvectors that are produced.

11.12.3 Inference on Cointegrating Vectors

- With respect to the model (11.68), suppose we are interested in testing for the number of cointegrating vectors. There are two approaches.

1. Test the hypothesis of **at most r CI vectors**, and the second is to test the hypothesis of r **CI vectors versus $\mathbf{r-1}$ vectors**. Since the derivation of the estimation procedure is maximum likelihood, the resulting test statistics can be referred to as Likelihood Ratio tests. The likelihood ratio test for at most r CI vectors is given by:

$$LR_{(at\ most\ r)} = -T \sum_{i=r+1}^{m} ln(1-\hat{\lambda}_i). \tag{11.77}$$

   This is referred to as a **trace statistic** since it involves the sum of a number of elements of the diagonal of a matrix. This test is not asymptotically distributed as a χ^2. Instead the distributions are multivariate versions of the Dickey-Fuller distribution (see Johansen (1988) for one set of critical values). If the null hypothesis of at most r CI vectors is rejected the researcher proceeds to test for at most $r+1$ CI vectors.

2. The likelihood ration test for r CI vectors versus $r-1$ CI vectors is given by:

$$LR_{(r\ vs\ r-1)} = -Tln(1-\hat{\lambda}_r). \tag{11.78}$$

   This test uses only the r^{th} eigenvalue.

   (a) A special case of this test is the **maximum eigenvalue test**. This is a test for the presence of one CI vector. It is called the maximum eigenvalue given the ordering of the eigenvalues with the first one being the largest. The test is given by:

$$LR_{maximum} = -Tln(1-\hat{\lambda}_1). \tag{11.79}$$

- When we have estimated under linear restrictions to β, the test of the hypothesis of H_3 in H_2 is given by:

$$LR_{H_3 in H_2} = T \sum_{i=1}^{r} ln\left[(1-\hat{\lambda}_{3,i})/(1-\hat{\lambda}_i)\right], \tag{11.80}$$

  where the $\hat{\lambda}$ are from the estimation of model (2) and the $\hat{\lambda}_3$ are from the estimation under the restrictions. This test statistic will be asymptotically χ^2 with $r \times (m-s)$ degrees of freedom.

- When the model is estimated under linear restrictions to α, the test of the hypothesis of H_4 in H_2 is given by:

$$LR_{H_4 in H_2} = T \sum_{i=1}^{r} ln\left[(1-\hat{\lambda}_{4,i})/(1-\hat{\lambda}_i)\right], \tag{11.81}$$

where the $\hat{\lambda}$ are from the estimation of model (2) and the $\hat{\lambda}_4$ are from the estimation under the restrictions. This test statistic will be asymptotically χ^2 with $r \times (m-k)$ degrees of freedom.

11.12.4 Roundup

This has provided some information on CI in systems of equations. It has been fairly dense, since the program that may be used to implement this procedure will assume a working knowledge of the papers by Johansen and Juselius (1990, 1992). These notes should however provide enough of a background to implement the programs available in STATA

11.13 Estimating Restricted Cointegrating Vectors

Elliott (2000, *Journal of Business and Economics Statistics,vol. 18, No. 1* pages 91-99) outlined a method for estimating the restricted cointegrating vector (restricted estimation) using standard minimum distance methods. The procedure also allows for a test of the overidentifying restrictions. The idea is as follows. Let the unrestricted model be written in triangular form as:

$$\begin{aligned} y_{1t} &= d_{1t} + y_{1t-1} + u_{1t} \\ y_{2t} &= d_{2t} + \Gamma y_{1t-1} + u_{1t} \end{aligned}$$

where y_{1t} and y_{2t} are $(n_1 \times 1)$ and $(n_2 \times 1)$ contain unit roots and (11.82) contains the normalization for the cointegrating system. That is we have written the system in terms of (possibly) new variables in a specific though arbitrary cointegration form. Γ has at most $n1 \times n2$ parameters of which we wish to test:

$$\begin{aligned} H_0 &: vec(\Gamma) = g(\theta) \\ H_1 &: vec(\Gamma) \neq g(\theta) \end{aligned} \tag{11.82}$$

where θ is of reduced dimension $q \times 1$ ($q < n1n2$) and g may be nonlinear (vec is a stacking operator). The **minimum distance estimator:**

$$\begin{matrix} Min \\ \theta \end{matrix} \left(\left[vec(\hat{\Gamma}) - g(\theta) \right]^T \hat{V}^{-1} \left[vec(\hat{\Gamma}) - g(\theta) \right] \right) \tag{11.83}$$

where $\hat{V}$ is an estimate of the variance-covariance matrix of the $vec(\hat{\Gamma})$. The test for the restrictions (and consistency of $\hat{\theta}$) as

$$J_T = T^2 \left[vec(\hat{\Gamma}) - g(\theta) \right]^T \hat{V}^{-1} \left[vec(\hat{\Gamma}) - g(\theta) \right] \sim \chi^2_{n_1 n_2 - q} \tag{11.84}$$

asymptotically under H_0. Elliott shows in an appendix how to construct $vec(\hat{\Gamma})$ and $\hat{V}$ for Johansens and Saikkonen estimators.

11.14 Summary

Cointegration and unit roots are important areas of research in time series econometrics (for instance cointegration in fractionally integrated processes done by Katsumi Sitmosui. Given the state of current knowledge there are many unanswered questions. It continues to be one of the more interesting areas of time series development

11.15 Additional References

Andrews, D. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, **Econometrica**, 59, pp. 817-858.

Andrews, D. K. and J. C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, **Econometrica**, 60, pp. 953-966.

Banerjee, A., J. Dolado, D. F. Hendry and G. W. Smith, 1986, Exploring equilibrium relationships in econometrics through static models: some monte carlo evidence, **Oxford Bulletin of Economics and Statistics** 48, 253-277.

Dhrymes, P., 1998, **Time Series, Unit Roots, and Cointegration**, Academic Press.

Engle, R. F. and C. W. J. Granger, 1987, Cointegration and error correction representation, estimation, and testing, **Econometrica** 55, 251-276.

Engle, R. F. and C. W. J. Granger, 1991 **Long-Run Economic Relationships: Readings in Cointegration**, Oxford University Press.

Engle, R. F. and B. S. Yoo, 1987, Forecasting and testing in cointegrated systems, **Journal of Econometrics** 35, 135-159.

Gregory, A. W., 1992, Testing for cointegration in the linear quadratic model, **Journal of Business and Economic Statistics**, ,???

Gregory, A. W. and A. Haug , 1998, Conflict Among Tests for Cointegration, Queen'S Unniversity Discussion Paper No. ???.

Gregory, A. W., J. M. Nason,and David Watt, 1996, Testing for structural breaks in cointegrated relationships, **Journal of Econometrics**

Gregory, A. W., A. R. Pagan, and G. W. Smith, 1991, Estimating linear quadratic models with integrated processes, in **Models, Methods and Applications of Econometrics: Essays in Honour of A.R. Bergstrom**, edited by P.C.B. Phillips, Blackwell Publishers.

Hansen, B., 1990, A powerful, simple test for cointegration using Cochrane-Orcutt, Working paper No. 260, University of Rochester, Rochester.

Hansen, B., 1992, Tests for parameter instability in regressions with I(1) processes, **Journal of Business and Economic Statistics**.

Johansen, S., 1988, Statistical analysis of cointegrating vectors, **Journal of Economic, Dynamics and Control**, 12, 231-254.

Johansen, S. 1989, The power function of the likelihood ratio test for cointegration, Preprint No. 8, Insitute of Mathematical Statistics, University of Copenhagen.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models, **Econometrica** 59, 1551-1580.

Johansen, S. and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration with applications to the demand for money, **Oxford Bulletin of Economics and Statistics**, 52, 169-210.

MacKinnon, J.G., 1991, Critical values for cointegration tests, in **Long-Run Economic Relationships: Readings in Cointegration**, edited by R.F.Engle and C. W. J. Granger, Oxford University Press.

Newey, W.K. and K.D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, **Econometrica**, 55, 703-708.

Park, J. Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: part 1, **Econometric Theory**, 4, 468-498.

Park, J. Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: part 2, **Econometric Theory**, 5, 95-131.

Park, J., S. Ouliaris and B. Choi, 1988, Spurious regressions and tests for cointegration, mimeo, Cornell University.

Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, **Journal of Econometrics**,33, 311-340.

Phillips, P.C.B., 1987, Time series regressions with unit roots, **Econometrica** 55, 271-302.

Phillips, P.C.B., 1991, Optimal inferences in cointegrated systems, **Econometrica** 59, 283-326.

Phillips, P.C.B. and S. N. Durlauf, 1986, Multiple time series with integrated variables, **Review of Economic Studies** 53, 473-396.

Phillips, P.C.B. and B. Hansen, 1990, Statistical inference in instrumental variables regression with I(1) processes, **Review of Economic Studies**, 57, 99-125.

Phillips, P.C.B. and S. Ouliaris, 1988, Testing for cointegration using principal components method, **Journal of Economics Dynamics and Control** 12, 205-130.

Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, **Econometrica** 58, 165-193.

Phillips, P.C.B. and P. Perron, 1988, Testing for a unit root in time series regression, **Biometrika** 75, 335-346.

Said, S. E. and D. A. Dickey , 1984, Testing for unit roots in autoregressive moving average models of unknown orders, **Biometrika** 71, 599-607.

Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regression, **Econometric Theory** 7, 1-21.

Sargent, T. J., 1987, **Macro-economic theory**, Second Edition , Academic Press, Orlando.

Schwert, W. G., 1989, Test for unit roots: a Monte Carlo investigation, **Journal of Business and Economics Statistics** 7, 147-159.

Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, **Econometrica**, 55, 1035-1056.

Stock, J. H. and M. W. Watson, 1988, Testing for common trends, **Journal of the American Statistical Association** 83, 1097-1107.

Stock, J. H. and M. W. Watson, 1993, A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, **Econometrica**, 61, pp. 783-820.