Econometrics Lecture 9: Cointegration in VAR Models

R. G. Pierse

1 Introduction

In last semester's course, the topics of testing for unit roots and cointegration were introduced. This lecture revisits these topics in the context of the *VAR* model and looks in detail at the Johansen (1988, 1991) procedure for testing for cointegration. This introduction briefly reviews the basic concepts. For more details, last semester's lecture notes should be consulted.

1.1 Stationarity

Definition 1.1. *Weak stationarity*
A variable y_t is weakly stationary if its mean and variance are both constant over time.

Most economic variables do not satisfy the conditions of weak stationarity.

1.2 Two Simple Models

1.2.1 Deterministic Trend

$$y_t = \gamma + \beta t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \tag{1.1}$$

This model has a *non-constant* mean, and a *constant* variance. Stationarity is achieved by *detrending*.

1.2.2 Random walk with drift

$$y_t = c + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \tag{1.2}$$

This model has both a *non-constant* mean, and a *non-constant* variance. Stationarity is achieved by first differencing. A series that can be made stationary by differencing is said to be *integrated*, or to possess a *unit root*.

Definition 1.2. *A time series y_t is integrated of order d, denoted $I(d)$, if $\Delta^d y_t$ is stationary. Then the series y_t has d unit roots.*

1.3 Testing for Unit Roots

The two models (1.1) and (1.2) have very different properties. In the deterministic trend model shocks to y_t are merely temporary and have no permanent effect on the future path of the variable. In the unit root model shocks have a permanent effect. We would like to be able to test between these two models. Consider the equation

$$\Delta y_t = \alpha y_{t-1} + bt + c + \varepsilon_t \tag{1.3}$$

A unit root test is a test of the null of

$$H_0 : \alpha = 0 \quad \text{against} \quad H_1 : \alpha < 0$$

in this equation. However, the distribution of the parameter α in (1.3) is non-standard and we cannot use standard t-tests to test the hypothesis that $\alpha = 0$. Instead, new tests with non-standard distributions have to be used. Dickey and Fuller (1979, 1981) propose a test based on the t-ratio $t(\alpha)$ in the *OLS* regression (1.3). The distribution of this statistic is *non-standard* and depends on the presence of the *nuisance parameters*, b and c. Critical values of the statistic are given in Fuller (1976) Table 8.5.2 and in Banerjee *et al.* (1993).

The Dickey-Fuller test is based on the assumption that ε_t is 'white noise' *i.e. serially uncorrelated.* If ε_t is serially correlated then the serial correlation needs to be corrected before the unit root test is performed. If it is assumed that the serial correlation in ε_t can be represented by an $AR(p)$ process, then it can be corrected by adding the p lagged terms $\Delta y_{t-1}, \cdots, \Delta y_{t-p}$ to the regression (1.3) to give

$$\Delta y_t = \alpha y_{t-1} + bt + c + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + \varepsilon_t. \tag{1.4}$$

The distribution of the test statistic is unaffected by the addition of these lagged differences. This procedure is called the *augmented Dickey-Fuller* test or *ADF* test.

1.4 Cointegration

Suppose that y_{1t}, y_{2t}, $\cdots$, y_{kt} are a set of $I(1)$ variables. In general, any linear combination of them such as

$$\sum_{i=1}^{k} w_i y_{it}$$

will also be $I(1)$ for all set of weights $w_i \neq 0$. However, suppose there exists some linear combination such that

$$\sum_{i=1}^{k} w_i^* y_{it} \text{ is } I(0), \quad w_i^* \neq 0 \ .$$

Then we say that the variables Y_{it} are *cointegrated* and that the weights w_i^* form a *cointegrating vector.*

1.5 The Meaning of Cointegration

If $I(1)$ variables are cointegrated, this means that although they are individually non-stationary, they are moving together so that there is some long run relationship between them. Consider a static equation between two $I(1)$ variables which may possibly be cointegrated:

$$y_t = a + bx_t + \eta_t. \tag{1.5}$$

If y_t and x_t are *not* cointegrated then there is no possible value of the parameters a and b such that η_t can be stationary. If they are cointegrated however, then there is a *single* value for the two parameters such that the linear combination $y_t - a - bx_t$ is stationary. This is when the parameters are the weights of a cointegrating vector. For this unique value of the parameters, (1.5) is a valid econometric equation with stationary error term η_t. It represents the long run equilibrium relationship between the two variables and this can only exist when there is cointegration.

Cointegration can thus be seen as the existence of a long run relationship between variables and economic theory leads us to expect that cointegration should exist. Cointegration is a long run property of variables. In the short-run, the variables can be moving in different ways, driven by different dynamic processes. However, cointegration ties the variables together in the long run.

1.6 Testing Cointegration

If a set of variables are cointegrated, then the residuals from a static regression of any one of the variables on all the others will be stationary. If not, then the residuals will be integrated. Thus Dickey-Fuller tests on the *OLS* residuals e_t from a static regression provide a way of testing cointegration. This was proposed by Engle and Granger (1987).

The critical values will be different from those from the standard Dickey-Fuller tests because e_t is based on *estimated* parameters. The null hypothesis in the test is that $e_t \sim I(1)$, i.e. *zero cointegrating vectors*, and the alternative is that $e_t \sim I(0)$, i.e. *one cointegrating vector*. Critical values for the *ADF* tests are given in MacKinnon (1991).

1.7 Error Correction Mechanisms

The cointegrating vector represents the long run relationship between two cointegrated variables. What about the short-run relationship? Granger and Engle (1987) show that this can be represented by an error correction model or *ECM*.

Suppose that two $I(1)$ variables y_{1t} and y_{2t} are cointegrated with cointegrating relationship

$$\eta_t = y_{1t} - a - by_{2t}. \tag{1.6}$$

Then the short run relationship can be represented by

$$\Delta y_{1t} = a_0 + a_1 \Delta y_{1,t-1} + \cdots + a_p \Delta y_{1,t-p} + b_1 \Delta y_{2,t-1} + \cdots + b_p \Delta y_{2,t-p} + \gamma \eta_{t-1} + u_t \tag{1.7}$$

which is an *ECM* representation. Note that all the terms in the representation (1.7) are $I(0)$ so that the coefficients in the equation will all have standard distributions.

The *ECM* can be given an economic interpretation as an adjustment mechanism whereby deviations from the equilibrium relationship in the previous period, as measured by η_{t-1}, lead to adjustments in y_{1t}. This is the reason why it is known as an error correction mechanism.

It can be shown that the *ECM* representation (1.7) is simply a reparameterisation of the general dynamic model

$$y_{1t} = \alpha_0 + \alpha_1 y_{1,t-1} + \cdots + \alpha_{p+1} y_{1,t-p-1} + \beta_1 y_{2,t-1} + \cdots + \beta_{p+1} y_{2,t-p-1} + u_t \tag{1.8}$$

but one that makes explicit the long-run cointegrating relationship (1.6) and which is expressed entirely in terms of stationary variables. Note that this equation is the first equation from a bivariate *VAR* system of order $p + 1$.

2 Cointegration in VAR Models

Let us now consider cointegration in the multivariate context. Let $\mathbf{y}_t$ be an $n \times 1$ set of variables, all of which are $I(1)$. In general, any linear combination

$$\mathbf{a}'\mathbf{y}_t$$

will also be $I(1)$ for arbitrary $\mathbf{a} \neq \mathbf{0}$. However, suppose there exists an $n \times 1$ vector $\boldsymbol{\alpha}_i$ such that

$$\boldsymbol{\alpha}_i' \, \mathbf{y}_t \text{ is } I(0), \quad \boldsymbol{\alpha}_i \neq \mathbf{0} \ .$$

Then we say that the variables $\mathbf{y}_t$ are *cointegrated* and $\boldsymbol{\alpha}_i$ is a *cointegrating vector.*

Note that if $\boldsymbol{\alpha}_i$ is a cointegrating vector, then so is $k\boldsymbol{\alpha}_i$ for any $k \neq 0$ since $k\boldsymbol{\alpha}_i' \mathbf{y}_t \sim I(0)$.

Definition 2.1. *If*

$$\mathbf{y}_t \sim I(d) \quad \textit{and} \quad \boldsymbol{\alpha}_i' \mathbf{y}_t \sim I(d-b), \quad \boldsymbol{\alpha}_i \neq \mathbf{0}$$

then

$$\mathbf{y}_t \sim CI(d,b), \quad d \geq b > 0.$$

There can be r different cointegrating vectors, where $0 \leq r < n$. Note that r must be less than the number of variables n. If a test for r produces the result that $r = n$ then this is *incompatible* with the assumption that $\mathbf{y}_t \sim I(1)$ and suggests some problem in the analysis.

Let

$$\boldsymbol{\alpha} = \left[\begin{array}{ccccc} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_i & \cdots & \boldsymbol{\alpha}_r \end{array} \right]$$

denote the $n \times r$ *matrix* of rank r, comprising all the cointegrating vectors. Then the $r \times 1$ vector

$$\boldsymbol{\alpha}'\mathbf{y}_t \sim I(0)$$

and, for *any nonsingular* $r \times r$ matrix $\mathbf{K}$, it also follows that

$$\boldsymbol{\alpha}'\mathbf{y}_t = \boldsymbol{\alpha}^{*\prime}\mathbf{y}_t \sim I(0).$$

In order to uniquely identify the cointegrating vectors, it is necessary to impose r^2 restrictions to pin down $\mathbf{K}$.

3 *VECM* Representation

Let $\mathbf{y}_t \sim I(1)$ be the pth order *VAR* model

$$\mathbf{y}_t = \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \mathbf{u}_t.$$

If and only if the $\mathbf{y}$'s are cointegrated, with cointegrating vectors $\boldsymbol{\alpha}$, then the reparameterisation

$$\Delta\mathbf{y}_t = \mathbf{A}_1\Delta\mathbf{y}_{t-1} + \mathbf{A}_2\Delta\mathbf{y}_{t-2} + \cdots + \mathbf{A}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\gamma}\boldsymbol{\alpha}'\mathbf{y}_{t-1} + \mathbf{u}_t$$

will consist entirely of $I(0)$ variables. This result is called the *Granger Representation Theorem*, and the parameterisation is known as the *Vector Error Correction Mechanism* or *VECM*. It is a vector generalisation of the simple error correction mechanism considered in the introduction. The terms $\boldsymbol{\alpha}'\mathbf{y}_{t-1}$ are called the error correction terms. There are r of them with coefficients given by the $n \times r$ matrix $\boldsymbol{\gamma}$. Note that

$$\boldsymbol{\gamma}\boldsymbol{\alpha}'\mathbf{y}_{t-1} = \boldsymbol{\gamma}\mathbf{K}^{-1}\mathbf{K}\boldsymbol{\alpha}'\mathbf{y}_{t-1} = \boldsymbol{\gamma}^*\boldsymbol{\alpha}^{*\prime}\mathbf{y}_t$$

so that the coefficients γ are only uniquely identified when $\mathbf{K}$ has been determined.

4 Estimating a Single Cointegrating Vector

Consider estimating a *single* cointegrating vector, $\boldsymbol{\alpha}_1'\mathbf{y}_t \sim I(0)$ in the *VAR* model

$$\boldsymbol{\Phi}(\mathbf{L})\mathbf{y}_t = \mathbf{u}_t$$

4.1 Static Regression

Partition $\mathbf{y}_t$ and $\boldsymbol{\alpha}_1$ conformably as

$$\mathbf{y}_t' = \left[\begin{array}{ccc} y_{1t} & : & \mathbf{y}_{2t}' \end{array} \right]$$

and

$$\boldsymbol{\alpha}_1' = \left[\begin{array}{ccc} 1 & : & -\boldsymbol{\alpha}^{*\prime} \end{array} \right].$$

This is an (*arbitrary*) normalising restriction. Then consider estimating the *static regression*

$$y_{1t} = \boldsymbol{\beta}'\mathbf{y}_{2t} + w_t.$$

From the definition of cointegration we know that for $\boldsymbol{\beta} = \boldsymbol{\alpha}^*$, $w_t \sim I(0)$, but for all other values of $\boldsymbol{\beta}$, then $w_t \sim I(1)$. Since *OLS* estimation minimises the *mean square error*, it is intuitively obvious that

$$\operatorname*{plim}_{T\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\alpha}^*$$

and in fact it can be shown that the rate of convergence is $O(T)$ as opposed to $O(\sqrt{T})$ in conventional models with $I(0)$ variables. This property of *OLS* with $I(1)$ variables is known as *super consistency.*

4.2 Testing Cointegration

Static regression provides a framework for testing cointegration, based on the *OLS* residuals $\widehat{w}_t$. Any of the standard unit root tests can be used, but the critical values will be different because $\widehat{w}_t$ is based on *estimated* parameters. The null hypothesis in the test is that $\widehat{w}_t \sim I(1)$, i.e. *zero cointegrating vectors*, against the alternative that $\widehat{w}_t \sim I(0)$, i.e. *one cointegrating vector.* Critical values for the *ADF* test, based on fitting response surfaces to simulation results, are given in MacKinnon (1991).

4.3 Engle-Granger Two-Step Procedure

Engle and Granger (1987) propose a two-step procedure for estimation.

Step 1: Estimate $\boldsymbol{\alpha}^*$ from the static regression

Step 2: Estimate the dynamics from the *VECM*

$$b_1(L)'\Delta y_{1t} = b_2(L)'\Delta y_{2t} + \gamma\widehat{w}_{t-1} + u_t$$

4.4 Problems with Static Regression

The static regression approach is simple and easy-to-use. However, it has certain drawbacks:

1. *It ignores dynamics*
2. *It ignores simultaneity*
3. *It is based on an artibrary normalisation*
4. If $r > 1$, then it will find a *linear combination* of the r cointegrating vectors.

Although *OLS* estimates of $\boldsymbol{\alpha}^*$ are *super consistent*, they can still be *heavily biased* in finite samples, as has been found in simulation studies. Because of the problems of bias in the static regression, Phillips and Hanson (1990) have suggested a *non-parametric* correction for bias. This corrected *OLS* static regression is called the *Fully-Modified LS* estimator.

5 Estimating Several Cointegrating Vectors

The Johansen (1988, 1991) procedure is based on the maximum likelihood estimation of the *VECM* model

$$\Delta\mathbf{y}_t = \sum_{i=1}^{p-1}\mathbf{A}_i\Delta\mathbf{y}_{t-i} + \boldsymbol{\gamma}\boldsymbol{\alpha}'\mathbf{y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\delta}'\mathbf{x}_t + \mathbf{u}_t \tag{5.1}$$

where the *VAR* model has been generalised to include an intercept term $\boldsymbol{\mu}$ and a set of $I(0)$ exogenous variables $\mathbf{x}_t$. Note that the cointegration term has been redated at $t-p$ rather than $t-1$. (The dating of the cointegration term makes no essential difference to the analysis).

The log-likelihood function of this model, after concentrating out the nuisance parameters $\mathbf{A}_i$, $\boldsymbol{\mu}$, and $\boldsymbol{\delta}$, can be written as

$$\log L(\alpha) = c - \frac{T}{2}\sum_{i=1}^{p}\log(1-\lambda_i) \tag{5.2}$$

where λ_i are *generalised eigenvalues* that are the solution to the problem

$$\left|\lambda \mathbf{S}_{kk} - \mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}\right| = 0$$

where $\mathbf{S}_{ij} = T^{-1}\sum_{t=1}^{T}\mathbf{R}_{it}\mathbf{R}'_{jt}$, $ii, j = 0, k$, and $\mathbf{R}_{0t}$ and $\mathbf{R}_{kt}$ are the vectors of residuals from regressing $\Delta\mathbf{y}_t$ and $\mathbf{y}_{t-p}$ respectively, on $\{\Delta\mathbf{y}_{t-1}, \cdots, \Delta\mathbf{y}_{t-p+1}, \boldsymbol{\mu}, \mathbf{x}_t\}$. The number of cointegrating vectors, r, is equal to the number of *non-zero* eigenvalues, λ_i.

5.1 Tests of the order of r

Let the eigenvalues λ_i, $i = 1, \cdots, n$ be ordered from largest to smallest. Then a test of the null hypothesis of r cointegrating vectors against the alternative of *more than* r can be based on either the *trace statistic*

$$H_0 : -T\sum_{i=r+1}^{n}\log(1-\widehat{\lambda}_i) = 0 \tag{5.3}$$

or the *maximal eigenvalue* statistic

$$H_0 : -T\log(1-\widehat{\lambda}_{r+1}) = 0. \tag{5.4}$$

Both tests are used but the maximal eigenvalue test (5.4) is generally regarded as the more powerful.

In both cases the sequence of testing is as follows: firstly test $r = 0$ against $\dot{r} > 0$. If the null is not rejected then this implies that there is no cointegration. If, on the other hand, the null is rejected, then there is at least one cointegrating vector. In this case, we then test the hypothesis $r = 1$ against $r > 1$. If this null is rejected, then we next test $r = 2$ against $r > 2$ and so on. This procedure continues until a null is not rejected, at which point the the order of r is determined. Note that the last possible test is of the hypothesis $r = n - 1$ against $r > n - 1$. If this hypothesis is rejected, then something has gone wrong with the analysis since we know that $r < n$. The most likely cause of this problem is that at least one of the variables in $\mathbf{y}_t$ is actually $I(0)$ and not $I(1)$.

5.2 The Distribution of the Test Statistics

The distribution of the trace and maximal eigenvalue statistics are *non-standard* and have been tabulated by Johansen (1995) and Osterwald-Lenum (1992). Unfortunately, as with the Dickey-Fuller statistic, the distribution depends on the nuisance parameter $\boldsymbol{\mu}$. Several models can be considered:

1. no intercept: $\boldsymbol{\mu} = 0$

2. restricted intercept (intercept only in error correction term)

$$\Delta \mathbf{y}_t = \sum_{i=1}^{p-1} \mathbf{A}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\gamma}(\boldsymbol{\alpha}' \mathbf{y}_{t-p} + \boldsymbol{\alpha}_0) + \boldsymbol{\delta}' \mathbf{x}_t + \mathbf{u}_t$$

   where $\boldsymbol{\gamma}\boldsymbol{\alpha}_0 = \boldsymbol{\mu}$.

3. unrestricted intercept

It is also possible to consider the case where the data is generated by model 2 but model 3 is estimated. The three models are nested and it is possible to test the restricted models against the less restricted using likelihood ratio tests, *provided that the number of cointegrating vectors is known.*

However, often the model can be chosen on *a priori* grounds. For example, if the variables are trended then an intercept should be included.

5.3 Identification in the Johansen Procedure

In order to identify $\boldsymbol{\alpha}$, r^2 restrictions need to be imposed on the *VECM*. Johansen imposes the statistical restrictions

$$\boldsymbol{\alpha}_i' \mathbf{S}_{kk} \boldsymbol{\alpha}_i = 1 \quad \text{and} \quad \boldsymbol{\alpha}_i' \mathbf{S}_{kk} \boldsymbol{\alpha}_j = 0, \forall i, j \quad j \neq i.$$

These restrictions exactly identify the parameters but do not have any obvious economic interpretation. Since any set of restrictions that exactly identifies the parameters is equally valid, several authors have proposed alternative identification restrictions that are more intuitively appealing.

5.3.1 Phillips Triangular Form

Phillips (1991) proposed the triangular form identification restriction

$$\boldsymbol{\alpha} = \begin{bmatrix} \mathbf{I}_r \\ -\overline{\boldsymbol{\alpha}} \end{bmatrix}$$

where $\overline{\boldsymbol{\alpha}}$ is $(n-r) \times 1$ and is unrestricted. This corresponds to a partitioning of the variables into two sets

$$\mathbf{y}_t' = \begin{bmatrix} \mathbf{y}_{1t}' & \mathbf{y}_{2t}' \end{bmatrix}$$

such that

$$\boldsymbol{\alpha}' \mathbf{y}_t = \mathbf{y}_{1t} - \overline{\boldsymbol{\alpha}}' \mathbf{y}_{2t} \sim I(0)$$

or

$$\mathbf{y}_{1t} = \overline{\boldsymbol{\alpha}}' \mathbf{y}_{2t} + \mathbf{v}_t$$

where the $n-r$ variables, $\mathbf{y}_{2t}$, are not themselves cointegrated.

5.3.2 Pesaran and Shin

Pesaran and Shin (1997) propose imposing the r^2 identifying restrictions on the basis of *a priori* economic theory. This is like imposing structural restrictions on the *VAR*. This procedure is available in *MicroFit Version 4*.

5.4 Hypothesis Testing

It is possible to test *overidentifying* restrictions on the cointegrating vectors. This makes it possible to test the validity of interpreting the vectors as long run equilibrium relationships corresponding to economic theory. The Johansen identification restrictions can make hypothesis testing a little awkward, and it is often easier to perform these tests in one of the alternative parameterisations considered above. However, the principle is straightforward. For example, a set of *homogeneous* restrictions can be tested by

$$H_0 : \mathbf{R}_i' \boldsymbol{\alpha}_i = \mathbf{0}, \quad i = 1, \cdots, r$$

where $\mathbf{R}_i$ is an $n \times s$ matrix of known constants. The test statistic will be *asymptotically* distributed as χ^2 with $r(n-s)$ degrees of freedom.

References

[1] Banerjee, A., J. Dolado, J.W. Galbraith and D.F. Hendry (1993), *Cointegration, Error-Correction, and the Analysis of Non-stationary Data*, Oxford University Press, Oxford.

[2] Dickey, D.A. and W.A. Fuller (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, 427–431.

[3] Dickey, D.A. and W.A. Fuller (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica*, 49, 1057–1072.

[4] Engle, R.F. and C.W. J. Granger (1987), 'Cointegration and error correction: representation, estimation and testing', *Econometrica*, 55, 251–276.

[5] Engle, R.F. and C.W.J. Granger (1991), *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[6] Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.

[7] Johansen, S. (1988), 'Statistical analysis of cointegrating vectors', *Journal of Economic Dynamics and Control*, 12, 231–54.

[8] Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, 59, 1551–80.

[9] Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

[10] MacKinnon, J.G. (1991), 'Critical values of cointegration tests', in R.F. Engle and C.W.J. Granger (eds.) *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[11] Osterwald-Lenum, M. (1992), 'A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: four cases', *Oxford Bulletin of Economics and Statistics*, 54, 461–472.

[12] Pesaran, M.H. and Y. Shin (1997), 'Long-run structural modelling', University of Cambridge, Department of Applied Economics, *mimeo.*

[13] Phillips, P.C.B. (1991), 'Optimal inference in cointegrated systems', *Econometrica*, 59, 282–306.

[14] Phillips, P.C.B. and B.E. Hanson (1990), 'Statistical inference in Instrumental Variables regression with I(1) processes', *Review of Economic Studies*, 57, 99–125.