E 4101/5101
Lecture 13: Cointegration, estimation and testing: Part 2

Ragnar Nymoen

Department of Economics, University of Oslo

30 April 2014

Introduction I

For the vector $\boldsymbol{y}_t$ consisting of $n \times 1$ variables, we have the Gaussian $VAR(p)$:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}(L)\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1}$$

We assume that if there are unit-roots in the associated characteristic equation, they are located at the zero frequency. By using the transformed equation

$$\Delta\boldsymbol{y}_t = \boldsymbol{\Phi}^*(L)\Delta\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{2}$$

We write the levels coefficient matrix $\boldsymbol{\Pi}$ as the product of two matrices $\boldsymbol{\alpha}_{n\times r}$ and $\boldsymbol{\beta}'_{r\times n}$ where $r \equiv rank(\boldsymbol{\Pi})$:

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' \tag{3}$$

Introduction II

We are interested in both the cointegrating case

$$0 < rank(\mathbf{\Pi}) < n$$

and the case with no cointegration

$$rank(\mathbf{\Pi}) = 0$$

- Since $rank(\mathbf{\Pi})$ is given by the number of non-zero eigenvalues of $\mathbf{\Pi}$, one approach to testing is to find the number of eigenvalues of $\hat{\mathbf{\Pi}}$ that are significantly different from zero.
- Fortunately, this problem has a solution because the eigenvalues has an interpretation as a special kind of squared correlation coefficients.

Introduction III

- This method has become known as the Johansen approach. It is “likelihood based”, see Johansen (1995) and the underlying assumption is a VAR with normal, or Gaussian, disturbances.
- Hamilton presentation of this approach is in Ch 20
- Davidson and MacKinnon in Ch 14.6, Bårdsen and Nymoen (2014, Ch. 11)
- The Johansen approach is covered in other textbooks as well, Juselius (2006) is perhaps the most comprehensive and accessible.

Concentrated likelihood functions I

We first consider the n dimensional $VAR(1)$:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T. \tag{4}$$

$$\boldsymbol{\varepsilon}_t \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{5}$$

The conditional log-likelihood function is

$$L(\boldsymbol{\Pi}, \boldsymbol{\Omega}) = -\frac{Tn}{2} ln(2\pi) - \frac{T}{2} ln \left|\boldsymbol{\Omega}\right| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}_t. \tag{6}$$

which we write without any explicit notation for the conditioning on $\boldsymbol{y}_0$, i.e., as in Hamilton Ch. 11.1

Concentrated likelihood functions II

The ML-estimator for $\mathbf{\Omega}$ is the usual one, namely:

$$\widehat{\mathbf{\Omega}}=\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t' \tag{7}$$

where $\hat{\varepsilon}_t$ is the vector of OLS residuals
Given (7) we proceed to use the concentrated likelihood-function

$$L^c(\mathbf{\Pi}) = -\frac{T}{2}ln\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'\right|. \tag{8}$$

Concentrated likelihood functions III

where we have dropped a constant, and maximize this function "under the restriction" that $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, which gives

$$L^c(\boldsymbol{\alpha},\boldsymbol{\beta}) = -\frac{T}{2} ln \left| \frac{1}{T} \sum_{t=1}^{T} [(\Delta \boldsymbol{y}_{t-1} - \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1})(\Delta \boldsymbol{y}_t - \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1})' \right| \tag{9}$$

The expresssion inside the determinant can be written as

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} \Delta \boldsymbol{y}_t \Delta \boldsymbol{y}_t^{'}}_{\boldsymbol{S}_{00}} - \boldsymbol{\alpha}\boldsymbol{\beta}' \underbrace{\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}^{'}}_{\boldsymbol{S}_{11}} \boldsymbol{\beta}\boldsymbol{\alpha}' -$$

$$-\boldsymbol{\alpha}\boldsymbol{\beta}' \underbrace{\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{'}}_{\boldsymbol{S}_{10}} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} \Delta \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}^{'}}_{\boldsymbol{S}_{01}} \boldsymbol{\beta}\boldsymbol{\alpha}'$$

Concentrated likelihood functions IV

$$L^c(\boldsymbol{\alpha},\boldsymbol{\beta}) = -\frac{T}{2} ln \left| \boldsymbol{S}_{00} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{11} - \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{S}_{10} - \boldsymbol{S}_{01}\boldsymbol{\beta}\boldsymbol{\alpha}' \right|$$

If we first consider $\boldsymbol{\beta}$ as given, we obtain $\boldsymbol{\alpha}(\boldsymbol{\beta})$ from

$$\frac{\partial L^c(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \boldsymbol{\alpha}} = 0 \Rightarrow$$

$$\boldsymbol{\alpha}(\boldsymbol{\beta}) = \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1} \tag{10}$$

Insertion in L^c gives a new concentrated likelihood:

$$L^{cc}(\boldsymbol{\beta}) = -\frac{T}{2} ln \left| \boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10} \right|$$

If we define

$$\boldsymbol{\Theta}(\boldsymbol{\beta}) = \left| \boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10} \right| \tag{11}$$

we now want to find the $\boldsymbol{\beta}$ that minimizes the function $\boldsymbol{\Lambda}(\boldsymbol{\beta})$.

A trick, and a result from multivariate analysis I

The determinant of a 2×2 matrix can be written as

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = \begin{cases} d(a - bd^{-1}c) \\ a(d - ba^{-1}c) \end{cases}$$

The same holds for partitioned matrices.
With suitable definitions of a, b, c and d we can formulate a determinant

$$\begin{vmatrix} \boldsymbol{S}_{00} & \boldsymbol{S}_{01}\boldsymbol{\beta} \\ \boldsymbol{\beta}'\boldsymbol{S}_{10} & \boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta} \end{vmatrix} = \begin{cases} \boldsymbol{\Theta}(\boldsymbol{\beta}) \cdot \left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right| \\ \left|\boldsymbol{S}_{00}\right| \left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\mathbf{fi}'\boldsymbol{S}_{01}\boldsymbol{\beta}\right| \end{cases} \tag{12}$$

We have a case of $c = b'$, meaning that

$$\boldsymbol{S}_{01}\boldsymbol{\beta}\boldsymbol{S}_{00}^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10} = \boldsymbol{\beta}'\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{\beta}$$

A trick, and a result from multivariate analysis II

and the two ways of writing the determinant in (12) give

$$\frac{\Theta(\boldsymbol{\beta})}{|\boldsymbol{S}_{00}|} = \left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right|^{-1}\left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}-\boldsymbol{\beta}'\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{\beta}\right| \qquad (13)$$

Since $\boldsymbol{S}_{00}$ does not depend on $\boldsymbol{\beta}$, minimizing $\Theta(\boldsymbol{\beta})$ is equivalent to minimize:

$$\frac{\left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}-\boldsymbol{\beta}'\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{\beta}\right|}{\left|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}\right|} \qquad (14)$$

This problem has a standard solution in multivariate analysis, Anderson (1951):

A trick, and a result from multivariate analysis III

The columns of $\boldsymbol{\beta}$ are estimated by the eigenvectors corresponding to the r largest solutions of the generalized eigenvalue problem:

$$\rho \boldsymbol{S}_{11} - \boldsymbol{S}_{10} \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{01} = 0 \tag{15}$$

The idea is now to determine the r *largest eigenvalues* $\rho_1 \geq \rho_2 \geq \ldots \rho_r \ldots \geq \rho_n \geq 0$.
The columns of $\boldsymbol{\beta}$ are found as the corresponding r eigenvectors.
The full set of eigenvectors is

$$\left[\rho_i \boldsymbol{S}_{11} - \boldsymbol{S}_{10} \boldsymbol{S}_{00}^{-1} \boldsymbol{S}_{01}\right] \boldsymbol{v}_i = \boldsymbol{0}, \quad i = 1, 2, \ldots, n \tag{16}$$

Below, it is shown that the matrix $\mathbf{V}$, with the v_i vectors as columns, satisfies the normalization

$$\boldsymbol{V}^{'} \boldsymbol{S}_{11} \boldsymbol{V} = \boldsymbol{I}_{n \times n} \tag{17}$$

A trick, and a result from multivariate analysis IV

by virtue of being so called *canonical variates*.
We finally define a selection matrix $\boldsymbol{P}^{'} = \left[\boldsymbol{I}_{r\times r}, \boldsymbol{0}^{'}\right]$. This gives

$$\hat{\beta} = \boldsymbol{VP} \tag{18}$$

and

$$\hat{\beta}^{'}\boldsymbol{S}_{11}\hat{\beta} = \boldsymbol{I}_{r\times r} \tag{19}$$

while the ML estimator of $\boldsymbol{\alpha}$ is obtained by substitution in (10). In the light of (19), this gives simply

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{S}_{01}\hat{\beta}. \tag{20}$$

Identification of the cointegration space

- It us common to say that the r vectors $\hat{\boldsymbol{\beta}}$ given by (18) *span the (cointegrating) space* containing the true $\boldsymbol{\beta}$ asymptotically.
- In that *special* sense they are consistent estimates of $\boldsymbol{\beta}$. In view of (19):

$$\hat{\boldsymbol{\beta}}' \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}' \hat{\boldsymbol{\beta}} = \boldsymbol{I}_{r \times r}$$

$$\frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\beta}}' \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}' \hat{\boldsymbol{\beta}} = \boldsymbol{I}_{r \times r}$$

meaning that the $I(0)$ disequilibrium terms have unit variance and are orthogonal.

- However this an arbitrary normalization
- As we learned in a previous lecture, the cointegration *vectors* are unidentified in general (if $r = 1$, up to a constant)

Canonical variates and correlations I

We have T observations of $\Delta \boldsymbol{y}_t$ and $\boldsymbol{y}_{t-1}$. Collect in a $2n \times T$ matrix $\mathbf{X}$:

$$\boldsymbol{X}_{2n\times T} = \begin{bmatrix} \boldsymbol{X}_0 \\ \boldsymbol{X}_1 \end{bmatrix}. \quad (21)$$

Assume that we want to reduce the dimensionality of the problem by finding the two linear combinations of $\Delta \mathbf{y}_t$ and $\mathbf{y}_{t-1}$ that have the highest correlation. We define two new variables

$$\boldsymbol{u}{=}\boldsymbol{a}^{'}\boldsymbol{X}_0 \quad (22)$$

and

$$\boldsymbol{z}{=}\boldsymbol{b}^{'}\mathbf{X}_1 \quad (23)$$

Canonical variates and correlations II

where $\boldsymbol{a}$ and $\boldsymbol{b}$ is $1 \times n$. From the definitions of variance and covariance of linear combinations:

$$\text{Cov}(\boldsymbol{u},\boldsymbol{z}) = \boldsymbol{a}^{'}\mathbf{S}_{01}\boldsymbol{b} = \text{Cov}(\boldsymbol{u},\boldsymbol{z}) = \boldsymbol{b}^{'}\boldsymbol{S}_{10}\boldsymbol{a}, \tag{24}$$

where the $\boldsymbol{S}_{ij}$ matrices are defined above:

$$\text{Var}(\boldsymbol{u}) = \text{E}(\boldsymbol{u}^{'}\boldsymbol{u}) = \boldsymbol{a}^{'}\boldsymbol{S}_{00}\boldsymbol{a} \tag{25}$$

and

$$\text{Var}(\boldsymbol{z}) = \text{E}(\boldsymbol{z}^{'}\boldsymbol{z}) = \boldsymbol{b}^{'}\boldsymbol{S}_{11}\boldsymbol{b} \tag{26}$$

We *define* the first pair of canonical variates as the pair $\hat{\boldsymbol{u}}_1$, $\hat{\boldsymbol{z}}_1$ with variance equal to 1 and which maximizes $\text{Corr}(\boldsymbol{u},\boldsymbol{z})$.

Canonical variates and correlations III

Formally, choose the $\boldsymbol{a}$ and $\boldsymbol{b}$ that maximize

$$\mathsf{L} = \boldsymbol{a}'\boldsymbol{S}_{01}\boldsymbol{b} - \lambda_1(\boldsymbol{a}'\boldsymbol{S}_{00}\boldsymbol{a} - 1) - \lambda_2(\boldsymbol{b}'\boldsymbol{S}_{11}\boldsymbol{b} - 1). \quad (27)$$

1oc:

$$\boldsymbol{S}_{01}\boldsymbol{b} - 2\lambda_1\boldsymbol{S}_{00}\boldsymbol{a} = \boldsymbol{0} \quad (28)$$

$$\boldsymbol{S}_{10}\boldsymbol{a} - 2\lambda_2\boldsymbol{S}_{11}\boldsymbol{b} = \boldsymbol{0} \quad (29)$$

Pre-multiply in (28) by $\boldsymbol{a}'$, and in (29) by $\boldsymbol{b}'$:

$$\boldsymbol{a}'\boldsymbol{S}_{01}\boldsymbol{b} = 2\lambda_1 \quad (30)$$

$$\boldsymbol{b}'\boldsymbol{S}_{10}\boldsymbol{a} = 2\lambda_2. \quad (31)$$

Canonical variates and correlations IV

But note that

$$\boldsymbol{a}^{'}\boldsymbol{S}_{01}\boldsymbol{b} = \boldsymbol{b}^{'}\boldsymbol{S}_{10}\boldsymbol{a} = \text{Corr}(\boldsymbol{u},\boldsymbol{z}) \equiv R \tag{32}$$

since the variances are 1. Hence we have

$$\boldsymbol{a}^{'}\boldsymbol{S}_{01}\boldsymbol{b} = R = 2\lambda_1 = 2\lambda_2. \tag{33}$$

and (28) and (29) can be re-expressed as

$$\boldsymbol{S}_{01}\boldsymbol{b} = R\boldsymbol{S}_{00}\boldsymbol{a} \tag{34}$$

$$\boldsymbol{S}_{10}\boldsymbol{a} = R\boldsymbol{S}_{11}\boldsymbol{b}. \tag{35}$$

From (34):

$$\boldsymbol{a} = \frac{1}{R}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{b} \tag{36}$$

Canonical variates and correlations V

substitution in (35) gives

$$\left[\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01} - R^2\boldsymbol{S}_{11}\right]\boldsymbol{b} = \mathbf{0}. \tag{37}$$

Similarly:

$$\boldsymbol{b} = \frac{1}{R}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}\boldsymbol{a} \tag{38}$$

and

$$\left[\boldsymbol{S}_{01}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10} - R^2\boldsymbol{S}_{00}\right]\mathbf{a} = \mathbf{0}. \tag{39}$$

R^2 is found from

$$\left|\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01} - R^2\boldsymbol{S}_{11}\right| = \left|\boldsymbol{S}_{11}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01} - R^2\boldsymbol{I}_{nxn}\right| = 0 \tag{40}$$

meaning that R^2 is the eigenvalue to the matrix $\boldsymbol{S}_{11}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ (or $\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}$).

Canonical variates and correlations VI

- The eigenvector **b** is then determined from (37).
- The matrix $\boldsymbol{S}_{11}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ has n eigenvalues, but $R = \boldsymbol{a}^{'}\boldsymbol{S}_{01}\boldsymbol{b}$ is the number we wish to maximize, therefore the solution is to choose the largest eigenvalue, R_1^2, and the associated eigenvectors $\boldsymbol{b}_1$ and $\boldsymbol{a}_1$.
- The two first canonical variates must therefore be

$$\boldsymbol{u}_1 = \boldsymbol{a}_1^{'}\boldsymbol{X}_0 \tag{41}$$

$$\boldsymbol{z}_1 = \boldsymbol{b}_1^{'}\boldsymbol{X}_1, \tag{42}$$

$Corr(\boldsymbol{u}_1,\boldsymbol{z}_1) \equiv R_1$

- R_1 is called the (first) canonical *correlation coefficient*.
- The next $n-1$ pairs of canonical variates are defined in the same way as the first (unit variance in particular)

Canonical variates and correlations VII

- In addition it is required that all pairs are uncorrelated.
- The pairs are ordered by the size of the associated R_i^2.

Canonical correlation and ML I

- The eigenvalue problem (40) is the same problem as the ML approach was leading to, compare (15).
- Hence, there are several important relationships between Johansen's method and canonical correlations and variates:

$$\rho_1 = \max_{a,b} Corr(\boldsymbol{u},\boldsymbol{z}) = R_1^2$$

$$\rho_i = R_i^2, i = 1, 2, \ldots, n$$

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_r] = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_r, \ldots, \boldsymbol{b}_n]\boldsymbol{P} \equiv \boldsymbol{VP}$$

Canonical correlation and ML II

- Finally

$$\boldsymbol{V}^{'}\boldsymbol{S}_{11}\boldsymbol{V} = \mathbf{I}_{n\times n} \text{ and } \hat{\beta}^{'}\boldsymbol{S}_{11}\hat{\beta} = \boldsymbol{I}_{r\times r}$$

see (17) and (19), hold by definition for canonical variates, variance equal to 1 and uncorrelated.

Canonical correlation for open systems

- Canonical correlation analysis is not limited to cases with equal number of variables in the two groups $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$.
- If there are p variables in $\mathbf{X}_0$ and q variables in $\mathbf{X}_1$, the maximum number of eigenvalues becomes $min\{p, q\}$.
- Assume for example the we include non-modelled ("exogenous") $I(1)$ variables in the $\boldsymbol{X}_1$ matrix, so that $p < q$ $\Longrightarrow$ maximum number of cointegrating vectors is p
- And the rank of the relevant $\mathbf{\Pi}$ matrix is full.

Canonical correlations and single equation OLS I

- Canonical correlation coefficients are interpretable as generalizations of usual correlations coefficients
- Consider the case of $p = 1$ and $q > 1$. Linear regression gives canonical variates $\boldsymbol{X}_0$ and $\boldsymbol{b}'\boldsymbol{X}_1$and

$$\max_{\boldsymbol{b}} \text{Corr}(\boldsymbol{X}_0, \boldsymbol{b}'\boldsymbol{X}_1) = R$$

implies that

$$R^2 = \rho_1$$

From the formulae above, for $p = 1$ and $q > 1$

$$R = \frac{1}{s_0}\sqrt{\boldsymbol{S}_{01}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}} \quad (43)$$

where $s_0 = \boldsymbol{S}_{00}$, is consistent for σ_0^2, the variance of Δy_t. $\boldsymbol{S}_{01}$ is a $1 \times q$ vector

Canonical correlations and single equation OLS II

Assuming normality, the conditional variance of Δy_t given $\boldsymbol{y}_{t-1}$ er

$$\sigma_1^2 = \sigma_0^2 - \boldsymbol{S}_{01}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}$$

showing that R^2 from (43) is an estimate of the relative unexplained variation:

$$\frac{\sigma_0^2 - \sigma_1^2}{\sigma_0^2} = R^2$$

$$\sigma_0^2 = \sigma_1^2 \iff R^2 = 0$$
$$\sigma_1^2 = 0 \iff R^2 = 1$$

Canonical correlations and single equation OLS III

- When $p = 1$, both canonical correlation and the Johansen method are equivalent to usual regression analysis
- When $p > 1$, R_1^2 is larger than any R-squares from individual regressions.
- The interpretation is that more than one variable equilibrium corrects,
- This information is not used when we do single equation regression, which is therefore inefficient when weak exogeneity does not hold for the variables in the cointegrating vector.

Testing the hypotheses about reduced rank I

From (13), and neglecting additive constants, the maximized likelihood is

$$L^* = -\frac{T}{2} ln \left| \hat{\beta}^{'} (\mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{01}) \hat{\beta} \right| \tag{44}$$

From

$$\hat{\beta}^{'} \mathbf{S}_{11} \hat{\beta} = \mathbf{I}_{r \times r}$$

and

$$\hat{\beta} (\mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{01}) \hat{\beta} = \boldsymbol{\rho}_{rxr}$$

Testing the hypotheses about reduced rank II

where $\boldsymbol{\rho}_{rxr}$ is the diagonal matrix with the eigenvalues of $\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$, we obtain

$$L^* = -\frac{T}{2} ln \left| \boldsymbol{I}_{r\times r} - \boldsymbol{\rho}_{rxr} \right| = -\frac{T}{2} \sum_{i=1}^{r} \ln(1-\rho_i). \quad (45)$$

When $\boldsymbol{\Pi}$ is estimated freely, we have

$$L^{**} = -\frac{T}{2} \sum_{i=1}^{n} \ln(1-\rho_i) \quad (46)$$

Testing the hypotheses about reduced rank III

The Likelihood-ratio test of the hypothesis that there are at most r cointegrating vectors $0 \leq r < n$, and $n - r$ unit-roots:

$$\begin{aligned} \eta_r &= 2\{L^*(\hat{\beta}) - L^*(\mathbf{V})\} \qquad (47) \\ &= -T \sum_{i=r+1}^{n} \ln(1-\rho_i),\ r = 0, 1, 2, \dots, n-1 \end{aligned}$$

is called the *trace-test*. It's distribution is tabulated in Hamilton Table B.10, "Case 1".

- The testing is sequential; $\eta_0, \eta_1, \dots, \eta_{n-1}$.
- Not that if the largest squared correlation coefficient ρ_0 is small, the whole sequence $\eta_0, \eta_1, \dots, \eta_{n-1}$ will be small values as a result of low multivariate correlation between the $I(0)$ variables in $\Delta \boldsymbol{y}_t$ and the I(1) variables in $\boldsymbol{y}_{t-1}$.

Testing the hypotheses about reduced rank IV

- The number of cointegrating vector is $r+1$ if the last significant test is η_r (the H_0 of $n-r$ unit-roots is rejected).
- Since $\boldsymbol{y}_t$ is a multivariate $\mathrm{I}(1)$ process for the whole sequence of H_0 s,
- Therefore η_r will be a function of Brownian motions (and not a Chi-square).
- An alternative test hypothesis formulation is H_0: $r = r^*$, against H_1: $r = r^* + 1$ This leads to the *maximal-eigenvalue* test:

$$\zeta_r = -T \ln(1-\rho_{r+1}),\ r = 0, 1, 2, \ldots, n-1. \qquad (48)$$

Testing the hypotheses about reduced rank V

- The *trace-test* ha become most common in applied modelling.
- An interesting special case is:H_0: $r = r^* = n - 1$ against H_1: $r = r^* = n$.
- This becomes in effect a test of $I(1)$, with a single common trend under H_0.
- In this case the two tests coincide, and it has the asymptotic distribution of the square of the Dickey-Fuller t-statistic.

VAR(p)

- In the usual way, a VAR(p) can be written in terms of differences and lagged levels.
- With reference to FLW theorem: regress out the effect of the $p-1$ differenced variables from Δy_t and $\boldsymbol{y}_{t-1}$ and proceed to analyse these OLS residuals
- Can be interpreted as yet another "layer" of likelihood concentration.
- Based on the residuals the derivation of the ML-estimators is as above.
- Note the similarity to the ADF test

Deterministic terms I

- Lecture 11/12: It matters a great deal whether the constant is restricted to be in the cointegrating space or not
- It also affects the distributions of the tests
- μ free in DGP *and* in the model: Table B.10 Case 3.

μ in the model but $\mu = \alpha\mu_0$ in DGP: Table B.10, Case 2.
$\mu = \alpha\mu_0$ in the model *and* in DGP: Not tabulated in Hamilton.

- Restricted linear trends seem relevant for economic data, again the distributions for the tests are affected, the book by Juselius is a good reference but MacKinnon, Haug and Michelis even more comprehensive (and with programs).
- For those who use OxMetrics, the tutorial in Ch 4 in the manual: Pc-Give II (Modelling Dynamic Systems), see Doornik and Hendry (2013) is an excellent starting point.

Similarity in rank testing

- Since the distributions are different for different forms of deterministic non-stationarity, there is a premium on the test procedure that gives similarity
- The advise for data with visible drift in levels:
  - include an deterministic trend as *restricted* together with an unrestricted constant.
  - After rank determination, can test significance of the restricted trend with standard inference
- Shift in levels
  - Include restricted step dummy and a free impulse dummy.
- Exogenous variables, see table and program by MacKinnon, Haug and Michelis (1999), and Doornik (1998,2003).

I(0) variables in the VAR I

- A misunderstanding that sometimes occurs is that "there can be no stationary variables in he cointegrating relationships".
- Consider for example:

$$-y_{1t} + \beta_{12}y_{2t} + \beta_{13}y_{3t} + \beta_{14}y_{4t} = ecm_{1t} \quad (49)$$

$$\beta_{21}y_{1t} - y_{2t} + \beta_{23}y_{3t} + \beta_{24}y_{4t} = ecm_{2t} \quad (50)$$

- If y_1 is the log of real-wages, y_2 productivity, y_3 relative import prices, and y_4 the rate of unemployment, then the first relationship may be a bargaining based wage equation (which can be identified by restrictions on the system).
- Since y_{4t} is the rate of unemployment we may assume $y_{4t} \sim I(0)$. but we want to estimate and test the theory $\beta_{14} = 0$.

I(0) variables in the VAR II

- It is therefore appealing to specify the VAR with y_{4t} included in the cointegration space(*Restricted* in the cointegration analysis in PcGive Multivariate modelling, see Doornik and Hendry (2013).

From I(1) to I(0) I

- When the rank has been determined, we are back in the stationary-case.
- The distribution of the identified cointegration coefficents are "mixed normal" so that conventional asymptotic inference can be performed on this $\boldsymbol{\beta}$.
- The determination of rank allows us to move from the $I(1)$ VAR, to the cointegrated VAR that contains only $I(0)$ variables
- Another name for this $I(0)$ model is the vector equilibrium correction model, VECM.
- The VECM can usually be analysed further, using the tools of the stationary model (simultaneous equations, recursive models, etc).

From I(1) to I(0) II

- In sum, co-integration analysis is an important step in dynamic system modelling, but only one step.

References I

Anderson, T. W. (1951) Estimating Linear Restrictions on Regressions Coefficients for Multivariate Normal Distributions, *Annals of Mathematical Statistics*, 22, 327-51

Doornik, J.A. (1998) Approximations to the asymptotic distribution of cointegration tests.Journal of Economic Surveys 12, 573–593.

Doornik, J.A. (2003) Asymptotic tables for cointegration tests based on the gamma distribution approach. Nuffield College,University of Oxford

Doornik, J.A. and D.F. Hendry (2013). PcGiveTM 14 Volume II, Modelling Dynamic Systems, Timberlake Consultants, London.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press

Juselius, K (2004) *The Cointegrated VAR Model, Methodology and Applications*, Oxford University Press

References II

MacKinnon, J., A. A. Haug and L. Michelis (1999) Numerical Distributions Functions of Likelihood Ratio Tests for Cointegration, with programs.