Lecture 18
Cointegration

1

Spurious Regression

• Suppose y_t and x_t are I(1). We regress y_t against x_t. What happens?

• The usual *t*-tests on regression coefficients can show statistically significant coefficients, even if in reality it is not so.

• This the *spurious regression problem* (Granger and Newbold (1974)).

• In a Spurious Regression the errors would be correlated and the standard *t-statistic* will be wrongly calculated because the variance of the errors is not consistently estimated.

Note: This problem can also appear with I(0) series –see, Granger, Hyung and Jeon (1998).

Spurious Regression - Examples

Examples:

(1) Egyptian infant mortality rate (Y), 1971-1990, annual data, on Gross aggregate income of American farmers (I) and Total Honduran money supply (M)

$$\hat{y} = 179.9 - .2952\ I - .0439\ M, \quad R^2 = .918,\ DW = .4752,\ F = 95.17$$
$$(16.63)\ \ (-2.32)\ \ (-4.26) \quad Corr = .8858, -.9113, -.9445$$

(2). US Export Index (Y), 1960-1990, annual data, on Australian males' life expectancy (X)

$$\hat{y} = -2943. + 45.7974\ X, \quad R^2 = .916,\ DW = .3599,\ F = 315.2$$
$$(-16.70)\ \ (17.76) \quad Corr = .9570$$

(3) Total Crime Rates in the US (Y), 1971-1991, annual data, on Life expectancy of South Africa (X)

$$\hat{y} = -24569 + 628.9\ X, \quad R^2 = .811,\ DW = .5061,\ F = 81.72$$
$$(-6.03)\ \ (9.04) \quad Corr = .9008$$

Spurious Regression - Statistical Implications

• Suppose y_t and x_t are unrelated I(1) variables. We run the regression:

$$y_t = \beta x_t + \varepsilon_t$$

• True value of $\beta=0$. The above is a spurious regression and $e_t \sim$ I(1).

• Phillips (1986) derived the following results:

- $\hat{\beta}$ not $\xrightarrow{D} 0$. It $\xrightarrow{D}$ non-normal RV not necessarily centered at 0.
  => This is the spurious regression phenomenon.

-The OLS t-statistics for testing $H_{0:}\ \beta=0$ diverge to $\pm\infty$ as $T\rightarrow\infty$. Thus, with a large enough T it will appear that β is significant.

- The usual $R^2 \xrightarrow{D} 1$ as $T\rightarrow\infty$. The model appears to have good fit well even though it is misspecified.

Spurious Regression - Statistical Implications

• Intuition:

With I(1) data sample moments converge to functions of Brownian motion (not to constants).

• Sketch of proof of Phillip's first result.

- Consider two independent RW processes for y_t and x_t. We regress:

$$y_t = \beta x_t + \varepsilon_t$$

- OLS estimator of β:

$$\hat{\beta} = \frac{T^{-2} \sum_{t=1}^{T} y_t x_t}{\left(T^{-2} \sum_{t=1}^{T} x_t^2 \right)} \xrightarrow{d} \left(\sigma_x^2 \int_0^1 W_x(r)^2 dr \right)^{-1} \sigma_x \sigma_y \int_0^1 W_x(r) W_y(r) dr$$

- Then, $\hat{\beta}$ (not) $\xrightarrow{P} 0$. $\hat{\beta} \xrightarrow{D}$ non-normal RV.

Spurious Regression – Detection and Solutions

• Given the statistical implications, the typical symptoms are:
- High R^2, *t*-values, & *F*-values.
- Low DW values.

• Q: How do we detect a spurious regression (between I(1) series)?
- Check the correlogram of the residuals.
- Test for a unit root on the residuals.

• Statistical solution: When series are I(1), take first differences. Now, we have a valid regression. But, the economic interpretation of the regression changes.

. When series are I(0), modify the t-statistic:

$$\frac{\hat{\sigma}_{\hat{\varepsilon}}}{\hat{\lambda}} t \Rightarrow t \text{ - distribution} \ , \quad \text{where} \quad \hat{\lambda} = (\text{long-run variance of } \hat{\varepsilon})^{1/2}$$

Spurious Regression – Detection and Solutions

• The message from spurious regresssion: Regression of I(1) variables can produce nonsense.

Q: Does it make sense a regression between two I(1) variables?
Yes, if the regression errors are I(0). That is, when the variables are cointegrated.

Cointegration

• <u>Integration</u>: In a univariate context, y_t is I*(d)* if its *(d-1)th* difference is I(0). That is, $\Delta^d y_t$ is stationary.

$\quad\quad => y_t$ is I(1) if Δy_t is I(0).

• In many time series, integrated processes are considered together and they form equilibrium relationships:
- Short-term and long-term interest rates
- Inflation rates and interest rates.
- Income and consumption

<u>Idea</u>: Although a time series vector is integrated, certain linear transformations of the time series may be stationary.

Cointegration - Definition

• An mx1 vector time series Y_t is said to be cointegrated of order *(d,b)*, *CI(d,b)* where *0<b≤d*, if each of its component series Y_{it} is *I(d)* but some linear combination $\alpha'Y_t$ is *I(d−b)* for some constant vector $\alpha≠0$.

• α: cointegrating vector or long-run parameter.

• The cointegrating vector is not unique. For any scalar *c*
$$c\, \alpha'Y_t = \alpha^{*'}Y_t \text{ is } I(d−b) \sim I(d−b)$$

• Some normalization assumption is required to uniquely identify α. Usually, α_1=(the coefficient of the first variable) is normalized to 1.

• The most common case is *d=b=1*.

Cointegration - Definition

• If the mx1 vector time series Y_t contains more than 2 components, each being *I(1)*, then there may exist *k (<m)* linearly independent *1*xm vectors α_1', α_2',…, α_k', such that $\alpha'Y_t \sim I(0)$ *k*x1 vector process, where $\alpha = (\alpha_1, \alpha_2',…, \alpha_k)$ is a *k*xm cointegrating matrix.

• <u>Intuition for *I*(1) case</u>
$\alpha'Y_t$ forms a long-run equilibrium. It cannot deviate too far from the equilibrium, otherwise economic forces will operate to restore the equilibrium.

• The number of linearly independent cointegrating vectors is called the cointegrating rank:
$$\Rightarrow Y_t \text{ is cointegrated of rank } k.$$

If the mx1 vector time series Y_t is *CI(k,1)* with *0<k<m* CI vectors, then there are *m-k* common *I(1)* stochastic trends.

Cointegration - Example

<u>Example</u>: Consider the following system of processes

$$x_{1t} = \beta_1 x_{2t} + \beta_2 x_{3t} + \varepsilon_{1t}$$
$$x_{2t} = \beta_3 x_{3t} + \varepsilon_{2t}$$
$$x_{3t} = x_{3,t-1} + \varepsilon_{3t}$$

where the error terms are uncorrelated WN processes. Clearly, all the 3 processes are individually *I(1)*.

- Let $y_t = (x_{1t}, x_{2t}, x_{3t})'$ and $\gamma = (1, -\beta_1, -\beta_2)'$ $=> \gamma' y_t = \varepsilon_{1t} \sim I(0)$.

<u>Note</u>: The coefficient for x_1 is normalized to 1.

- Another CI relationship: x_{2t} & x_{3t}. Let $\gamma^* = (0\ 1, -\beta_3)' => \gamma' y_t = \varepsilon_{2t} \sim I(0)$.

- 2 independent C.I. vectors   => 1 common ST: $\Sigma_t\ \varepsilon_{3t}$.

VAR with Cointegration

• Let Y_t be mx1. Suppose we estimate VAR(p)

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + a_t$$

or

$$Y_t = \Phi(B) Y_{t-1} + a_t.$$

• Suppose we have a unit root. Then, we can write

$$\Phi(B) = \Phi(1) + (1 - B)\Phi^*(B)$$

• This is like a multivariate version of the ADF test:

$$Y_t = \rho Y_{t-1} + \sum_{i-1}^{p} \psi_i \Delta Y_{t-i} + a_t.$$

12

VAR with Cointegration

• Rearranging the equation

$$\Delta Y_t = (\Phi(1) - I)Y_{t-1} + \Phi^*(B)\Delta Y_{t-1} + a_t.$$

where $Rank(\boldsymbol{\Phi}(1) - \boldsymbol{I}) < m$. There are two cases:

1. $\boldsymbol{\Phi}(1) = \boldsymbol{I}$, then we have m independent unit roots, so there is no cointegration, and we should run the VAR in differences.

2. $0 < Rank(\boldsymbol{\Phi}(1) - \boldsymbol{I}) = k < m$, then we can write $\boldsymbol{\Phi}(1) - \boldsymbol{I} = \gamma\alpha'$ where γ and α are mxk. The equation becomes:

$$\Delta \boldsymbol{Y_t} = \boldsymbol{\gamma\alpha'Y}_{t-1} + \boldsymbol{\Phi}^*(B)\Delta\boldsymbol{Y}_{t-1} + \boldsymbol{a}_t.$$

• This is called a *vector error correction model* (VECM). Part of Granger Representation Theorem: "Cointegration implies an ECM."

13

VAR with Cointegration

• <u>Note</u>: If we have cointegration, but we run OLS in differences, then the modeled is misspecified and the results will be biased.

• Q: What can you do?
- If you know the location of the unit roots and cointegration relations, then you can run the VECM by doing OLS of $\Delta\boldsymbol{Y}_t$ on lags of $\Delta\boldsymbol{Y}$ and $\boldsymbol{\alpha'Y}_{t-1}$.
- If you know nothing, then you can either

      (i) run OLS in levels, or

      (ii) test (many times) to estimate cointegrating relations. Then, run VECM.

• The problem with this approach is that you are testing many times and estimating cointegrating relationships. This leads to poor finite sample properties.

14

Residual Based Tests of the Null of No CI

• Procedures designed to distinguish a system without cointegration from a system with at least one cointegrating relationship; they do not estimate the number of cointegrating vectors (the *k*).

• Tests are conditional on pretesting (for unit roots in each variable).

• There are two cases to consider.

• CASE 1 - Cointegration vector is pre-specified/known (say, from economic theory) :
Construct the hypothesized linear combination that is I(0) by theory; treat it as data. Apply a DF unit root test to that linear combination.

• The null hypothesis is that there is a unit root, or no cointegration.

15

Residual Based Tests of the Null of No CI

• CASE 2 - Cointegration vector is unknown. It should be estimated. Thus, if there exists a cointegrating relation, the coefficient on Y_{1t} is nonzero, allowing us to express the "static regression equation as

$$Y_{1t} = \beta' Y_{2t} + u_t$$

• Then, apply a unit root test to the estimated OLS residual from estimation of the above equation, but

- Include a constant in the static regression if the alternative allows for a nonzero mean in u_t

- Include a trend in the static regression if the alternative is stochastic cointegration -i.e., a nonzero trend for $A'Y_t$.

Note: Tests for cointegration using a prespecified cointegrating vector are generally more powerful than tests estimating the vector.

16

Engle and Granger Cointegration

• Steps in cointegration test procedure:

1. Test H_0(unit root) in each component series Y_{it} individually, using the univariate unit root tests, say ADF, PP tests.

2. If the H_0 (unit root) cannot be rejected, then the next step is to test cointegration among the components, i.e., to test whether $\boldsymbol{\alpha'Y}_t$ is I(0).

• In practice, the cointegration vector is unknown. One way to test the existence of cointegration is the regression method –see, Engle and Granger (1986) (EG).

• If $\boldsymbol{Y}_t=(Y_{1t},Y_{2t},...,Y_{mt})$ is cointegrated, $\boldsymbol{\alpha'Y}_t$ is I(0) where $\boldsymbol{\alpha}=(\alpha_1, \alpha_2,..., \alpha_m)$. Then, $(1/\alpha_1)\boldsymbol{\alpha}$ is also a cointegrated vector where $\alpha_1 \neq 0$.

17

Engle and Granger Cointegration

• EG consider the regression model for Y_{1t}

$$Y_{1t} = \delta D_t + \phi_1 Y_{2t} + \cdots + \phi_{m-1} Y_{mt} + \varepsilon_t$$

where D_t: deterministic terms.

• Check whether ε_t is *I(1)* or *I(0)*:

- If $\varepsilon_t \sim I(1)$, then $\boldsymbol{Y}_t$ is not cointegrated.

- If $\varepsilon_t \sim I(0)$, then $\boldsymbol{Y}_t$ is cointegrated with a *normalized cointegrating vector* $\boldsymbol{\alpha'}=(1,\phi_1,..., \phi_{m-1})$.

• Steps:

1. Run OLS. Get estimate $\hat{\alpha} = \left(1, \hat{\phi}_1,\cdots ,\hat{\phi}_{m-1}\right)$

2. Use residuals e_t for unit root testing using the ADF or PP tests without deterministic terms (constant or constant and trend). 18

9

Engle and Granger Cointegration

• Step 2: Use residuals e_t for unit root test.

- Note: $\hat{\phi}_t \xrightarrow{d} t-\text{distribution}, \quad \text{if } \varepsilon_t \sim I(0)$

        If $\varepsilon_t \sim I(1)$, *t-test* has a non-standard distribution.

- H_0 (unit root in residuals): $\lambda=0$ vs H_1: $\lambda <1$ for the model

$$\Delta\varepsilon_t = \lambda\varepsilon_{t-1} + \sum_{j=1}^{p-1}\varphi_j\Delta\varepsilon_{t-j} + a_t$$

- t-statistic: $t_\lambda = \dfrac{\hat{\lambda}}{s_{\hat{\lambda}}}$

- Critical values tabulated by simulation in EG.

• We expect the usual ADF distribution would apply here. But, Phillips and Ouliaris (PO) (1990) show that is not the case. [19]

EG Cointegration – PO Distribution

• Phillips and Ouliaris (PO) (1990) show that the ADF and PP unit root tests applied to the estimated cointegrating residual do not have the usual DF distributions under H_0 (no-cointegration).

• Due to the spurious regression phenomenon under H_0, the distribution of the ADF and PP unit root tests have asymptotic distributions that are functions of Wiener processes that depends on:
- The deterministic terms, D_t, in the regression used to estimate α
- The number of variables, (m-1), in Y_{2t}.

• PO tabulated these distributions. Hansen (1992) improved on these distributions, getting adjustments for different DGPs with trend and/or drift/no drift. [20]

EG Cointegration – Least Square Estimator

• EG propose LS to consistently estimate a normalized CI vector. But, the asymptotic behavior of the LS estimator is non-standard.

• Stock (1987) and Phillips (1991) get the following results:

- $T (\widehat{\alpha} - \alpha) \xrightarrow{D}$ non-normal RV not necessarily centered at 0.

- The LS estimate $\widehat{\alpha} \xrightarrow{p} \alpha$. Convergence is at rate T, not usual $T^{1/2}$. => We say $\widehat{\alpha}$ is *super consistent*.

- $\widehat{\alpha}$ is consistent even if the other *(m-1)* Y_t's are correlated with ε_t.

=> No asymptotic simultaneity bias.

- The OLS formula for computing $aVar(\widehat{\alpha})$ is incorrect

=>usual OLS standard errors are not correct.

- Even though the asymptotic bias →0, as $T→∞$, $\widehat{\alpha}$ can be substantially biased in small samples. LS is also not efficient.

21

EG Cointegration – Least Square Estimator

• The bias is caused by ε_t. If $\varepsilon_t \sim$ WN, there is no asymptotic bias.

• The above results point out that the LS estimator of the CI vector α could be improved upon.

• Stock and Watson (1993) propose augmenting the CI regression of Y_{1t} against the rest *(m-1)* elements in $\mathbf{Y}_t$, say $\mathbf{Y}_t$* with appropriate deterministic terms D_t, with p leads and lags of $\Delta\mathbf{Y}_t$*.

• Estimate the augmented regression by OLS. The resulting estimator of α is called the *dynamic OLS* estimator or $\widehat{\alpha}_{DOLS}$.

• It is consistent, asymptotically normally distributed and, under certain assumptions, efficient.

22

EG Cointegration – Estimating VECM with LS

• Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (Y_{1t}, Y_{2t})$.

- Assume that $\mathbf{Y}_t$ is cointegrated with CI $\boldsymbol{\alpha} = (1, -\alpha_2)$. That is,

$$\boldsymbol{\alpha}'\mathbf{Y}_t = Y_{1t} - \alpha_2 Y_{2t} \sim I(0).$$

- Suppose we have a consistent estimate $\widehat{\boldsymbol{\alpha}}$ (or $\widehat{\boldsymbol{\alpha}}_{\text{DOLS}}$) of $\boldsymbol{\alpha}$.

- We are interested in estimating the VECM for ΔY_{1t} and ΔY_{2t} using:

$$\Delta Y_{1t} = c_1 + \beta_1(Y_{1t} - \alpha_2 Y_{2t}) + \Sigma_j \psi_{11,j} \Delta Y_{1t-j} + \Sigma_j \psi_{12,j} \Delta Y_{2t-j} + u_{1t}$$

$$\Delta Y_{2t} = c_2 + \beta_2(Y_{1t} - \alpha_2 Y_{2t}) + \Sigma_j \psi_{21,j} \Delta Y_{1t-j} + \Sigma_j \psi_{22,j} \Delta Y_{2t-j} + u_{2t}$$

• $\widehat{\boldsymbol{\alpha}}$ is super consistent. It can be treated as known in the ECM. The estimated disequilibrium error $\widehat{\boldsymbol{\alpha}}'\mathbf{Y}_t = Y_{1t} - \alpha_2{}^{\wedge} Y_{2t}$ may be treated like the known $\boldsymbol{\alpha}'\mathbf{Y}_t$.

• All variables are $I(0)$, we can use OLS (or SUR to gain efficiency.)

Johansen Tests

• The EG procedure works well for a single equation, but it does not extend well to a multivariate VAR model.

• Consider a levels VAR(p) model:

$$Y_t = \delta D_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

where $\mathbf{Y}_t$ is a time series $m{\times}1$ vector. of $I(1)$ variables.

• The VAR(p) model is stable if

$$\det(\mathbf{I}_n - \Phi_1 z - \cdots - \Phi_p z^p) = 0$$

has all roots outside the complex unit circle.

• If there are roots on the unit circle then some or all of the variables in $\mathbf{Y}_t$ are I(1) and they may also be cointegrated. 24

Johansen Tests

• If $\mathbf{Y}_t$ is cointegrated, then the levels VAR representation is not the right one, since the cointegrating relations are not explicitly apparent.

• The CI relations appear if the VAR is transformed to the VECM.

• For these cases, Johansen (1988, 1991) proposed two tests: The *trace test* and the *maximal eigenvalue test*. They are based on Granger's (1981) ECM representation. Both tests are easy to implement.

<u>Example</u>: Trace test simple idea:

(1) Assume ε_t are multivariate N($\mathbf{0}, \boldsymbol{\Sigma}$). Estimate the VECM by ML, under various assumptions:

- trend/no trend and/or drift/no drift

- the number k of CI vectors,

(2) Compare models using likelihood ratio tests.

25

Johansen Tests - Intuition

• Consider the VECM

$$\Delta Y_t = \Gamma_0 D_t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \varepsilon_t$$

where

- $\boldsymbol{D}_t$: vector of deterministic variables (constant, trends, and/or seasonal dummy variables);

- $\Gamma_j = -\boldsymbol{I} + \phi_1 + \cdots + \phi_j, \mathrm{j} = 1, \cdots, \mathrm{p}\text{-}1$ are $m \times m$ matrices;

- $\Pi = \gamma \boldsymbol{A}'$ is the long-run impact matrix; $\boldsymbol{A}$ and γ are $m \times k$ matrices;

- ε_t are *i.i.d.* $N_m(\mathbf{0}, \Sigma)$ errors; and

- $\det(\boldsymbol{I} - \sum_{j=1}^{p-1} \Gamma_j \mathrm{B}^j)$ has all of its roots outside the unit circle.

• In this framework, CI happens when Π has *reduced rank*. This is the basis of the test: By checking the rank of Π, we can determine if the system is CI.

26

Johansen Tests - Intuition

• We can also write the ECM using the alternative representation as

$$\Delta Y_t = \Gamma_0 D_t + \Pi * Y_{t-p} + \sum_{j=1}^{p-1} \Gamma *_j \Delta Y_{t-j} + \varepsilon_t$$

where the ECM term is at lag *t-p*. Including a constant and or a deterministic trend in the ECM is also possible.

• Back to original VECM(p).
- Let $Z_{0t} = \Delta \mathbf{Y}_t$, $Z_{1t} = \mathbf{Y}_{t-1}$ and $Z_{2t} = (\Delta \mathbf{Y}_{t-1}, ..., \Delta \mathbf{Y}_{t-p-1}, D_t)'$

• Now, we can write: $Z_{0t} = \gamma \boldsymbol{A'} \mathbf{Y}_{t-1} + \boldsymbol{\Psi} Z_{2t} + \varepsilon_t$
where $\boldsymbol{\Psi} = (\Gamma_1 \Gamma_2 ... \Gamma_p \Gamma_0)$.

• If we assume a distribution for ε_t, we can write the likelihood. 27

Johansen Tests - Intuition

• Assume the VECM errors are independent $N_m(\mathbf{0}, \Sigma)$ distribution,. Then, given the CI restrictions on the trend and/or drift/no drift parameters, the likelihood $L_{max}(k)$ is a function of the CI rank k.

• The *trace test* is based on the log-likelihood ratio:
       LR= 2 *ln*[L_{max}(*Unrestricted*)/L_{max}(*Restricted*)],
which is done sequentially for *k = m-1,...,1,0*.

• The name comes from the fact that the test statistics involved are the *trace* (the sum of the diagonal elements) of a diagonal matrix of generalized eigenvalues.

• The test examines the H_0: CI rank $\leq k$, vs. H_1: CI rank $> k$.
- If the LR is greater than the critical value for a certain rank, then the H_0 is rejected.

28

Johansen Tests - Intuition

• Johansen concentrates all the parameter matrices in the likelihood function out, except for the matrix A. Then, he shows that the MLE of A can be derived as the solution of a generalized eigenvalue problem.

• LR tests of hypotheses about the number of CI vectors can then be based on these eigenvalues. Moreover, Johansen (1988) also proposes LR tests for linear restrictions on these CI vectors.

• <u>Note</u>: The factorization $\Pi = \gamma A'$ is not unique since for any $k \times k$ nonsingular matrix $\mathbf{F}$ we have:

$$\gamma A' = \gamma \ \mathbf{F} \ \mathbf{F}^{-1} \ A' = (\gamma \mathbf{F}) \ (\mathbf{F}^{-1} A') = \gamma^* A^{*'}$$

=> The factorization $\Pi = \gamma A'$ only identifies the space spanned by the CI relations. To get a unique γ and A', we need more restrictions. Usually, we normalize. Finding a good way to do this is hard.

29

Johansen Tests – Sequential Tests

• The Johansen tests examine H_0: Rank(Π) $\leq k$, where k is less than m.

• The unrestricted CI VECM is denoted $H(r)$. The I(1) model $H(k)$ can be formulated as the condition that the rank of Π is less than or equal to k. This creates a nested set of models

$$H\,(0\,) \subset \cdots \subset H\,(k\,) \subset \cdots \subset H\,(m\,)$$

- *H(m)* is the unrestricted, stationary VAR model or *I(0)* model
- *H(0)* non-CI VAR (restriction $\Pi = 0$) => VAR model for differences.

• This nested formulation is convenient for developing a sequential procedure to test for the number k of CI relationships.

30

Johansen Tests – Sequential Tests

• Sequential tests:

       i. H_0: $k=0$,   cannot be rejected →stop →$k=0$

(at most zero coint)   rejected →next test

       ii. H_0: $k<=1$,   cannot be rejected →stop→$k=1$

(at most one coint)   rejected →next test

       iii. H_0: $k \leq 2$,   cannot be rejected →stop→$k=2$

(at most two coint)   rejected →next test

• Possible outcomes:

- **Rank k = m** => All variables in x are $I(0)$, not an interesting case.
- **Rank k = 0** => No linear combinations of Y that are $I(0)$. $\Pi=0$.
     $\Rightarrow$ Model on differenced series
- **Rank k ≤ (m-1)** => Up to $(m\text{-}1)$ cointegration relationships $\alpha' Y_t$

31

Johansen Tests – Sequential Tests

• The Johansen tests examine H_0: Rank(Π) $\leq k$, where k is less than m.

Recall, Rank(Π) = number of non-zero eigenvalues of Π.

• Since $\Pi = \gamma A'$, it is equivalent to test that A and γ are of full column rank k, the number of independent CI vectors that forms the matrix A.

• It turns out the LR test statistic is the trace of a diagonal matrix of generalized eigenvalues from Π.

• These eigenvalues also happen to equal the squared *canonical correlations* between $\Delta \mathbf{Y}_t$ and $\mathbf{Y}_{t-1}$, corrected for lagged $\Delta \mathbf{Y}_t$ and D_t. They are between 0 and 1.

32

16

Johansen Tests – Likelihood

• Back to the VECM(p) representation:

$$\Delta Y_t = \Gamma_0 D_t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \varepsilon_t$$

where D_t may include a drift and a deterministic trend. Including a constant and or a deterministic trend in the ECM is also possible.

• Let $Z_{0t} = \Delta \mathbf{Y}_t$, $Z_{1t} = \mathbf{Y}_{t-1}$, $Z_{2t} = (\Delta \mathbf{Y}_{t-1}, ..., \Delta \mathbf{Y}_{t-p-1}, D_t)'$, and $\Psi = (\Gamma_1 \, \Gamma_2 \, ... \, \Gamma_p \, \Gamma_0)$. Then, we can write:

$$Z_{0t} = \gamma \mathbf{A'} \, \mathbf{Y}_{t-1} + \Psi \, Z_{2t} + \varepsilon_t$$

• Assuming normality for $\varepsilon_t \sim N(\mathbf{0}, \Sigma)$, we can write

$$\ell \;=\; -\frac{kT}{2} \log 2\pi - \frac{T}{2} \log |\Sigma|$$
$$-\frac{1}{2} \sum_{t=1}^{T} (Z_{0t} - \alpha \beta' Z_{1t} - \Psi Z_{2t})' \Sigma^{-1} (Z_{0t} - \alpha \beta' Z_{1t} - \Psi Z_{2t})$$

33

Johansen Tests – Eigenvalues

• Let residuals, R_{0t} and R_{1t}, be obtained by regressing Z_{0t} and Z_{1t} on Z_{2t}, respectively. The (FW) regression equation in residuals is:

$$R_{0t} = \gamma A' R_{1t} + e_t$$

• The crossproducts matrices are computed by

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

• Then, the MLE for $\mathbf{A}$ is obtained from the eigenvectors, $\mathbf{V}$, corresponding to the k largest eigenvalues of the following equation

$$| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} | = 0$$

• These λ's are squared canonical correlations between R_{0t} and R_{1t}. The $\mathbf{V}$'s corresponding to the k largest λ's are k linear combinations of Y_{t-1}.

34

Johansen Tests – Eigenvalues

• The eigenvectors corresponding to the k largest λ's are the k linear combinations of Y_{t-1}, which have the largest squared partial correlations with the I(0) process, after correcting for lags and D_t.

• Computations.

- λ's. Instead of using $\quad | \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} | = 0$,
pre and post multiply the expression by $S_{11}^{-1/2}$ (Cholesky decomposition of S_{11}). Then, we have a standard eigenvalue problem.

$$\Rightarrow \qquad | \lambda I - S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2} | = 0$$

- **V.** The eigenvectors (say, u_i) are usually reported normalized, such that $u_i' u_i = 1$. Then, in this case, we need to use $\hat{v}_i = S_{11}^{-1/2} u_i$

That is, we normalized the eigenvectors such that $\hat{V} S_{11} \hat{V} = I$.

35

Johansen Tests – Eigenvalues

• The tests are based on the λ's from $| \lambda I - S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2} | = 0.$

• Interpretation of the eigenvalue equation.
Using F-W, we regress R_{0t} on R_{1t}, to estimate $\Pi = \gamma A'$. That is

$$\hat{\Pi} = S_{11}^{-1/2} S_{10}.$$

Note that

$$S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2} = S_{11}^{1/2} S_{11}^{-1} S_{10} S_{00}^{-1/2} S_{00}^{-1/2} S_{01} S_{11}^{-1} S_{11}^{1/2} = (S_{11}^{-1/2} \hat{\Pi} S_{00}^{-1/2})(S_{00}^{-1/2} \hat{\Pi} S_{11}^{-1/2})$$

The λ's produced look like eigenvalues of $[\hat{\Pi} \hat{\Pi}]$ after pre-multiplying by $S_{11}^{-1/2}$ and post-multiplying by $S_{00}^{-1/2}$, a normalization.

• Johansen also finds: $\quad \hat{A} = A_{MLE} = [v_1, v_2, ..., v_k]$

$$\hat{\gamma} = \gamma_{MLE} = S_{01} \hat{A}.$$

36

Johansen Tests – Trace Statistic

• Johansen (1988) suggested two tests for H_0: At most k CI vectors:

    - The trace test

    - The maximal eigenvalue test.

• Both tests are based on the λ's from $| \lambda I - S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2} | = 0$. They are LR tests, but they do not have the usual χ^2 asymptotic distribution under H_0. They have non-standard distributions.

37

Johansen Trace Test

• The trace test: $LR_{trace}(k) = -2\ln \Lambda = -T \sum_{i=k+1}^{m} \ln\left(1 - \hat{\lambda}_i\right)$

where $\hat{\lambda}_i$ denotes the descending ordered eigenvalues $\hat{\lambda}_1 > \cdots > \hat{\lambda}_m > 0$ of $| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} | = 0$.

<u>Note</u>: The LR_{trace} statistic is expected to be close to zero if there is at most k (linearly independent) CI vectors.

• If $LR_{trace}(k)>$CV(for rank k), then H_0 (CI Rank=k) is rejected.

• If Rank($\mathbf{\Pi}$) = k_0 then $\hat{\lambda}_{k_0+1}, \cdots, \hat{\lambda}_m$ should all be close to 0. The $LR_{trace}(k_0)$ should be small since $\ln(1 - \hat{\lambda}_i) \approx 0$ for $i > k_0$.

38

Johansen Trace Test – Distribution

• Under H_0, the asymptotic distribution of $LR_{trace}(k_0)$ is not χ^2. It is a multivariate version of the DF unit root distribution, which depends on the dimension m-k_0 and the specification of D_t.

• The statistic $-ln\Lambda$ has a limiting distribution, which can be expressed in terms of a $m-k$ dimensional Brownian motion W as

$$tr\left\{\int_0^1 (dW)\tilde{W}' \left(\int_0^1 \tilde{W}\tilde{W}'dr\right)^{-1} \int_0^1 \tilde{W}(dW)'\right\}$$

$\tilde{W}$ is the Brownian motion itself (W), or the demeaned or detrended W, according to the different specifications for D_t in the VECM

• Using simulations, critical values are tabulated in Johansen (1988, Table 1) and in Osterwald-Lenum (1992) for m-$k_0 = 1, \ldots, 10$.

39

Johansen Maximal EigenvalueTest

• An alternative LR statistic, given by

$$LR_{max}(k) = -2\ln\Lambda = -T\ln\left(1 - \hat{\lambda}_{k+1}\right)$$

is called the *maximal eigenvalue statistic*. It examines the null hypothesis of k cointegrating vectors versus the alternative $k+1$ CI vectors. That is, H_0: CI rank $= k$, vs. H_1: CI rank $= k+1$.

• Similar to the trace statistic, the asymptotic distribution of this LR is not statistic the usual χ^2. It is given by the maximum λ of the stochastic matrix in

$$max\left\{\int_0^1 (dW)\tilde{W}'\left(\int_0^1 \tilde{W}\tilde{W}'dr\right)^{-1} \int_0^1 \tilde{W}(dW)'\right\}$$

which depends on the dimension $m-k_0$ and the specification of the deterministic terms, D_t. See Osterwald-Lenum (1992) for CVs.

40

ML Estimation of the CI VECM

• Suppose we find Rank(Π) = k, $0 < k < m$. Then, the CI VECM:

$$\Delta Y_t = \Gamma_0 D_t + \gamma A' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \varepsilon_t$$

• This is a reduced rank multivariate regression. Johansen derived the ML estimation of the parameters under the reduced rank restriction Rank(Π) = k.

Recall that $\widehat{\mathbf{A}} = \mathbf{A}_{MLE}$ is given by the eigenvectors associated with the λ's.

• The MLEs of the remaining parameters are obtained by OLS of

$$\Delta Y_t = \Gamma_0 D_t + \gamma \hat{A}' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \varepsilon_t$$

41

ML Estimation of the CI VECM

• The factorization $\Pi = \gamma A'$ is not unique. The columns of $\mathbf{A}_{MLE}$ may be interpreted as linear combinations of the underlying CI relations.

• For interpretation, it is convenient to normalize the CI vectors by choosing a specific coordinate system in which to express the variables.

• Johansen suggestion: Solve for the triangular representation of the CI system. The resulting normalized CI vector is denoted $\mathbf{A}_{c,MLE}$.

• The normalization of $\mathbf{A}$ affects the MLE of γ but not the MLEs of the other parameters in the VECM.

• Properties of $\mathbf{A}_{c,MLE}$: asymptotically normal and super consistent. 42

ML Estimation of the CI VECM - Testing

• The Johansen MLE procedure only produces an estimate of the basis for the space of CI vectors.

• It is often of interest to test if some hypothesized CI vector lies in the space spanned by the estimated basis:

$$H_0: \mathbf{A} = [\mathbf{A}_0 \; \varphi] \qquad \text{Rank}(\mathbf{\Pi}) \le k$$

$\mathbf{A}_0$: s✖m matrix of hypothesized CI vectors

φ : $(k-s)$✖m matrix of unspecified CI vectors

• Johansen shows that a LR test can be computed, which is asymptotically distributed as a χ^2 with $s(m-k)$ degrees of freedom.

43

Common Trends

• Following Johansen (1988, 1991) one can choose a set of vectors $\boldsymbol{A}^{\perp}$ such that the matrix $\{\mathbf{A}, \boldsymbol{A}^{\perp}\}$ has full rank and $\mathbf{A}'\, \boldsymbol{A}^{\perp} = 0$. [$\boldsymbol{A}^{\perp}$ read "$\boldsymbol{A}$ perp"]

• That is, the $mx(m-k)$ matrix $\boldsymbol{A}^{\perp}$ is orthogonal to the matrix $\boldsymbol{A}$
      => columns of $\boldsymbol{A}^{\perp}$ are orthogonal to the columns of $\boldsymbol{A}$.

• The vectors $\boldsymbol{A}^{\perp} Y_t$ represents the non-CI part of Y_t. We call $\boldsymbol{A}^{\perp}$ the *common trends loading matrix.*

• We refer to the space spanned by $\boldsymbol{A}^{\perp} Y_t$ as the *unit root space* of Y_t.

• Reference:Stock and Watson (1988).

44

22

Asymptotic Efficient Single Equation Methods

• The EG's two-step estimator is simple, but not asymptotically efficient. Several papers proposed improved, efficient methods.

- Phillips (1991): Regression in the spectral domain.

- Phillips and Loretan (1991): Non-linear EC estimation.

- Phillips and Hansen (1990): IV regression with a correction a la PP.

- Saikkonen (1991): Inclusion of leads and lags in the lag-polynomials of the ECM in order to achieve asymptotic efficiency

- Saikkonen (1992): Simple GLS type estimator

- Park's (1991) CCR estimator transforms the data so that OLS afterwards gives asymptotically efficient estimators

- Engle and Yoo (1991): A 3 step estimator for the EG procedure.

• From all of these estimators, we can get a t-values for the EC term.

Example: Cointegration - Lütkepohl (1993) - SAS

Example (Lütkepohl (1993)): $m=4$ U.S. quarterly macro variables: Log real M1, Log output, 91-day T-bill yield, 20-year T-bond yield.

Period: 1954 to 1987

• Analysis:

1) Dickey-Fuller unit root test

2) Johansen cointegration test integrated order 2,

3) VECM(2) estimation.

• SAS Code
```
proc varmax data=us_money;
id date interval=qtr;
model y1-y4 / p=2 lagmax=6 dftest print=(iarr(3))
      cointtest=(johansen=(iorder=2)) ecm=(rank=1 normalize=y1);
      cointeg rank=1 normalize=y1 exogeneity;
run;
```
46

23

Example: Lütkepohl (1993) – SAS: DF Tests

• Dickey-Fuller Unit Root Tests

Variable	Type	Rho	Pr < Rho	Tau	Pr < Tau
y1	Zero Mean	0.05	0.6934	1.14	0.9343
	Single Mean	-2.97	0.6572	-0.76	0.8260
	Trend	-5.91	0.7454	-1.34	0.8725
y2	Zero Mean	0.13	0.7124	5.14	0.9999
	Single Mean	-0.43	0.9309	-0.79	0.8176
	Trend	-9.21	0.4787	-2.16	0.5063
y3	Zero Mean	-1.28	0.4255	-0.69	0.4182
	Single Mean	-8.86	0.1700	-2.27	0.1842
	Trend	-18.97	0.0742	-2.86	0.1803
y4	Zero Mean	0.40	0.7803	0.45	0.8100
	Single Mean	-2.79	0.6790	-1.29	0.6328
	Trend	-12.12	0.2923	-2.33	0.4170

<u>Note</u>: In all series, we cannot reject H_0 (unit root).

47

Example: Lütkepohl (1993) – SAS: VECM(2)

• The fitted VECM(2) is given as

$$\Delta \mathbf{y}_t = \begin{pmatrix} 0.0408 \\ 0.0860 \\ 0.0052 \\ -0.0144 \end{pmatrix} + \begin{pmatrix} -0.0140 & 0.0065 & -0.2026 & 0.1306 \\ -0.0281 & 0.0131 & -0.4080 & 0.2630 \\ -0.0022 & 0.0010 & -0.0312 & 0.0201 \\ 0.0051 & -0.0024 & 0.0741 & -0.0477 \end{pmatrix} \mathbf{y}_{t-1}$$
$$ + \begin{pmatrix} 0.3460 & 0.0913 & -0.3535 & -0.9690 \\ 0.0994 & 0.0379 & 0.2390 & 0.2866 \\ 0.1812 & 0.0786 & 0.0223 & 0.4051 \\ 0.0322 & 0.0496 & -0.0329 & 0.1857 \end{pmatrix} \Delta \mathbf{y}_{t-1} + \epsilon_t$$

48

Example: Lütkepohl (1993) – SAS: VECM(2)

Cointegration Rank Test for I(2)						
r\k-r-s	4	3	2	1	Trace of I(1)	5% CV of I(1)
0	384.60903	214.37904	107.93782	37.02523	55.9633	47.21
1		219.62395	89.21508	27.32609	**20.6542**	**29.38**
2			73.61779	22.13279	2.6477	15.34
3				38.29435	0.0149	3.84
5% CV I(2)	47.21000	29.38000	15.34000	3.84000		

• <u>Note</u>: System is cointegrated in rank 1 with integrated order 1.

49

Example: Lütkepohl (1993) – SAS: VECM(2)

• The factorization $\Pi = \gamma A'$

A (Beta in SAS)

Variable	1	2	3	4	
y1	**1.00000**	**1.00000**	**1.00000**	**1.00000**	←—— Normalization y1
y2	-0.46458	-0.63174	-0.69996	-0.16140	
y3	14.51619	-1.29864	1.37007	-0.61806	
y4	-9.35520	7.53672	2.47901	1.43731	

γ (Alpha in SAS)

Variable	1	2	3	4
y1	-0.01396	0.01396	-0.01119	0.00008
y2	-0.02811	-0.02739	-0.00032	0.00076
y3	-0.00215	-0.04967	-0.00183	-0.00072
y4	0.00510	-0.02514	-0.00220	0.00016

50

25

Example: Lütkepohl (1993) – SAS: VECM(2)

• The factorization $\Pi = \gamma A'$

Long-Run Parameter Beta Estimates When RANK=1	
Variable	1
y1	1.00000
y2	-0.46458
y3	14.51619
y4	-9.35520

Adjustment Coefficient Alpha Estimates When RANK=1	
Variable	1
y1	-0.01396
y2	-0.02811
y3	-0.00215
y4	0.00510

• Covariance Matrix

Covariances of Innovations

Variable	y1	y2	y3	y4
y1	0.00005	0.00001	-0.00001	-0.00000
y2	0.00001	0.00007	0.00002	0.00001
y3	-0.00001	0.00002	0.00007	0.00002
y4	-0.00000	0.00001	0.00002	0.00002

51

Example: Lütkepohl (1993) – SAS: Diagnostics

Schematic Representation of Cross Correlations of Residuals

Variable/Lag	0	1	2	3	4	5	6
y1	++..	....	++..	....	+...	..--	....
y2	++++	....	....	....	....	....	....
y3	.+++	....	+.-.	..++	-...	....	....
y4	.+++	....	....	..+.	....	....	....

+ is > 2*std error, - is < -2*std error, . is between

Portmanteau Test for Cross Correlations of Residuals

Up To Lag	DF	Chi-Square	Pr > ChiSq
3	16	53.90	<.0001
4	32	74.03	<.0001
5	48	103.08	<.0001
6	64	116.94	<.0001

52

Example: Lütkepohl (1993) – SAS: Diagnostics

Univariate Model ANOVA Diagnostics				
Variable	R-Square	Standard Deviation	F Value	Pr > F
y1	0.6754	0.00712	32.51	<.0001
y2	0.3070	0.00843	6.92	<.0001
y3	0.1328	0.00807	2.39	0.0196
y4	0.0831	0.00403	1.42	**0.1963**

Univariate Model White Noise Diagnostics		Normality		ARCH	
Variable	Durbin Watson	Chi-Square	Pr > ChiSq	F Value	Pr > F
y1	2.13418	7.19	0.0275	1.62	0.2053
y2	2.04003	1.20	0.5483	1.23	0.2697
y3	1.86892	253.76	**<.0001**	1.78	0.1847
y4	1.98440	105.21	**<.0001**	21.01	**<.0001**

• <u>Note</u>: Residuals for y3 & y4 are non-normal. Except the residuals for y4, no ARCH effects on other residuals.

53

Example: Lütkepohl (1993) – SAS: Diagnostics

Univariate Model AR Diagnostics								
	AR1		AR2		AR3		AR4	
Variable	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
y1	0.68	0.4126	2.98	0.0542	2.01	0.1154	2.48	0.0473
y2	0.05	0.8185	0.12	0.8842	0.41	0.7453	0.30	0.8762
y3	0.56	0.4547	2.86	0.0610	4.83	**0.0032**	3.71	**0.0069**
y4	0.01	0.9340	0.16	0.8559	1.21	0.3103	0.95	0.4358

• <u>Note</u>: Except the residuals for y4, no AR effects.

54

Example: Lütkepohl (1993) – SAS: Diagnostics

• If a variable can be taken as "given" without losing information for statistical inference, it is called *weak exogenous*. In the CI model, a variable do not react to a disequilibrium –i.e., the EC term.

Weak exogeneity → Long-run non-causality

Testing Weak Exogeneity of Each Variables			
Variable	DF	Chi-Square	Pr > ChiSq
y1	1	6.55	**0.0105**
y2	1	12.54	**0.0004**
y3	1	0.09	0.7695
y4	1	1.81	0.1786

• <u>Note</u>: Variable $y1$ is not weak exogeneous for the other variables, $y2$, $y3$, & $y4$; variable $y2$ is not weak exogeneous for variables, $y1$, $y3$, & $y4$.