Lecture 18
Cointegration

1

Spurious Regression

• Suppose y_t and x_t are non-stationary, $I(1)$. That is, we differentiate them and the changes become stationary, or $I(0)$. 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)): We find a statistically significant relation between unrelated variables.

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

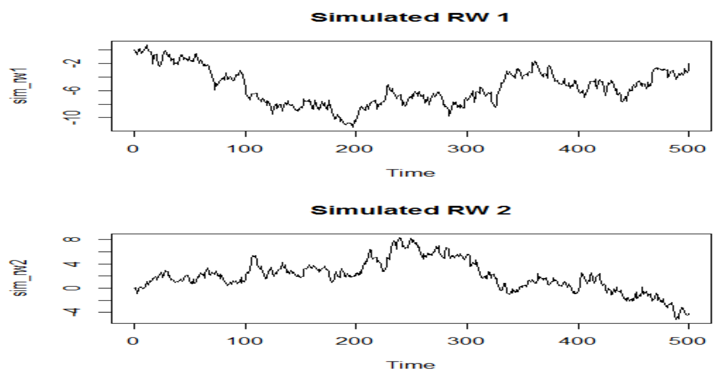
Spurious Regression: Simulated Example

Example: We simulate two independent RW ($RW_{1,t}$, $RW_{2,t}$) and then regress one against the other:

$$RW_{1,t} = \mu + \beta\, RW_{2,t} + \varepsilon_t$$

With test the true H_0: β = 0.

```
sim_rw1 <- arima.sim(list(order=c(0,1,0)), sd=.5, n=500)    # simulate RW series 1
sim_rw2 <- arima.sim(list(order=c(0,1,0)), sd=.5, n=500)    # simulate RW series 2
```



Spurious Regression: Simulated Example

Example:

```
fit_sim_rw <- lm(sim_rw1 ~ sim_rw2)              # Regression of two RWs
res_sim_rw <- fit_sim_rw$residuals               # Extract residuals
> summary(fit_sim_rw)

Coefficients:
             Estimate  Std. Error t value Pr(>|t|)
(Intercept) -4.61541   0.13188  -35.00   <2e-16 ***
sim_rw2     -0.47384   0.04076  -11.62   <2e-16 ***       ⇒ Reject H0: β = 0.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.356 on 499 degrees of freedom
Multiple R-squared:  0.2131,    Adjusted R-squared:  0.2115
F-statistic: 135.1 on 1 and 499 DF,  p-value: < 2.2e-16
```

Note: Very significant t-value (& F-goodness of fit stat), and a good R^2. But, the model makes no sense.

Spurious Regression: Real Examples

Examples:

(1) Egyptian infant mortality rate (Y_t), 1971-1990, annual data, on Gross aggregate income of American farmers (I_t) and Total Honduran money supply (M_t):

$\widehat{Y}_t$ = 179.9 - .2952 I_t − .0439 M_t, R^2 = **.918**, DW = **.4752**, F = 95.17

(16.63) (-2.32) (-4.26) Corr(Y_t, $X_{i,t}$) = .8858, -.9113, -.9445

(2). US Export Index (Y_t), 1960-1990, annual data, on Australian males' life expectancy (X_t):

$\widehat{Y}_t$ = -2943. + 45.7974 X_t, R^2 = **.916**, DW = **.3599**, F = 315.2

(-16.70) (17.76) Corr(Y_t, X_t) = .9570

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

$\widehat{Y}_t$ = -24569 + 628.9 X_t, R^2 = **.811**, DW = **.5061**, F = 81.72

(-6.03) (9.04) Corr(Y_t, X_t) = .9008

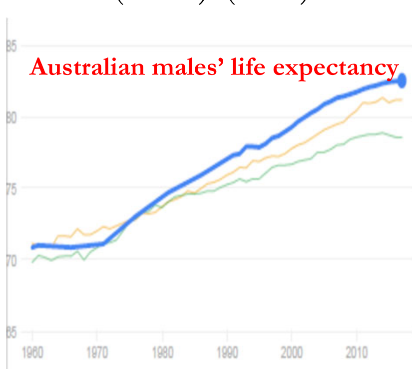
Spurious Regression: Real Examples

Examples (continuation):

(2). US Export Index (Y_t), 1960-1990, annual data, on Australian males' life expectancy (X_t):

$\widehat{Y}_t$ = -2943. + 45.7974 X_t, R^2 = **.916**, DW = **.3599**, F = 315.2

(-16.70) (17.76) Corr(Y_t, X_t) = .9570



Note: It looks like the trend is the common element between Y_t & X_t.

Spurious Regression: Statistical Implications

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

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

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

• <u>Technical points</u>: Phillips (1986) derived the following results:

- $\hat{\beta} \xrightarrow{p} \neq 0$ & $\hat{\beta} \xrightarrow{d}$ Non-normal RV not necessarily centered at 0.

 ⇒ This is the *spurious regression* phenomenon.

- The OLS *t*-statistics for testing H_0: β=0 diverge to ±∞ as $T \to \infty$. Thus, with a large enough T it will appear that β is significant.

- The usual $R^2 \xrightarrow{p} 1$ as $T \to \infty$. The model appears to have good fit well, even though it is a bad (nonsense) model.

Spurious Regression: Detection and Solutions

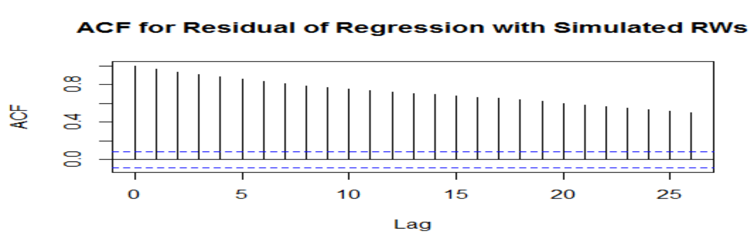
• Given the statistical implications, the typical symptoms are:

- 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. (This lecture.)

Spurious Regression: Detection and Solutions

• Statistical solution:
When series (y_t & $\boldsymbol{x}_t$) are $I(1)$, work with first differences, instead:

$$\Delta y_t = y_t - y_{t-1} \qquad \& \qquad \Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$$

If the relation between the series y_t & $\boldsymbol{x}_t$ exists, β should be the same in levels (y_t, $\boldsymbol{x}_t$) or first differences (Δy_t, $\Delta \boldsymbol{x}_t$).

Levels: $\qquad y_t = \beta\, x_t + \varepsilon_t \qquad$ (*)
Lagged Levels: $\qquad y_{t-1} = \beta\, x_{t-1} + \varepsilon_{t-1} \qquad$ (**)

Subtract (**) from (*): We have a regression with 1st differences:
First Differences: $\qquad \Delta y_t = \beta\, \Delta \boldsymbol{x}_t + u_t, \qquad$ where $u_t = \varepsilon_t - \varepsilon_{t-1}$

Now, we have a valid regression, since both regressors are I(0). But, the economic interpretation of the regression changes.

Spurious Regression: Simulated Example

Example: We regress the two RW in first differences:

```
diff_rw1 <- diff(sim_rw1)                      # First differences for RW 1
diff_rw2 <- diff(sim_rw2)                      # First differences for RW 2
fit_diff_rw <- lm(diff_rw1 ~ diff_rw2)
> summary(fit_diff_rw)

Coefficients:
              Estimate    Std. Error t value Pr(>|t|)
(Intercept) -0.003199   0.023481   -0.136   0.8917
diff_rw2     0.106339   0.044773    2.375   0.0179 *          ⇒ Reject H0: β = 0.
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.525 on 498 degrees of freedom
Multiple R-squared:  0.0112,    Adjusted R-squared:  0.009215
F-statistic: 5.641 on 1 and 498 DF,  p-value: 0.01792
```

Note: Still significant (by chance), but the β coefficient changes sign. Clear indication that something is wrong with regression.

Spurious Regression: Remarks

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

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

In this cointegration case, there is a linear combination of the $I(1)$ processes, $\boldsymbol{Y}_t$, such that $\boldsymbol{\alpha}'\boldsymbol{Y}_t \sim I(0)$.

We call $\boldsymbol{\alpha}$ the **cointegrating vector** or **long-run parameter**.

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.

$$\Rightarrow y_t \text{ is } I(1) \text{ if } \Delta y_t \text{ is } I(0)$$

• 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.
- Spot and Forward rates.
- Dividends and Earnings.

<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 $\boldsymbol{Y}_t$ is said to be **cointegrated** of order (d, b), $CI(d, b)$, where $0 < b \leq d$, if each of its component series $\boldsymbol{Y}_{it}$ is $I(d)$ but some linear combination $\boldsymbol{\alpha}'\boldsymbol{Y}_t$ is $I(d-b)$ for some constant vector $\boldsymbol{\alpha} \neq \mathbf{0}$. ($\boldsymbol{\alpha}$: **cointegrating vector**).

• The cointegrating vector is not unique. For any scalar c

$$c\, \boldsymbol{\alpha}'\boldsymbol{Y}_t = \boldsymbol{\alpha}^{*\prime}\boldsymbol{Y}_t \sim I(d-b)$$

• Some normalization assumption is required to uniquely identify $\boldsymbol{\alpha}$. Usually, α_1 (=the coefficient of the first variable) is normalized to 1. Look at the previous example, where the cointegrating vector is [1 -β].

• The most common case is $d = b = 1$.

Cointegration: Definition

• If the mx1 vector time series $\boldsymbol{Y}_t$ contains more than 2 components, each being $I(1)$, then there may exist k $(< m)$ linearly independent 1x*m* vectors $\boldsymbol{\alpha}_1{}'$, $\boldsymbol{\alpha}_2{}'$,..., $\boldsymbol{\alpha}_k{}'$, such that $\boldsymbol{\alpha}'\boldsymbol{Y}_t \sim I(0)$ kx1 vector process, where $\boldsymbol{\alpha}$ = ($\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$,..., $\boldsymbol{\alpha}_k$) is a kxm **cointegrating matrix**.

• The number of linearly independent cointegrating vectors is called the **cointegrating rank**: $\boldsymbol{Y}_t$ is cointegrated of rank k.

If the mx1 vector time series $\boldsymbol{Y}_t$ is *CI*$(k, 1)$ with $0 < k < m$ CI vectors, then, we say that there are $m - k$ common $I(1)$ ***stochastic trends***.

Cointegration: Example

Example: Consider the following system ($m = 3$):

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

where the error terms are uncorrelated WN processes. Since x_{3t} is a RW –i.e., $I(1)$–, clearly, all the 3 processes are individually $I(1)$.

- One CI relationship: x_{1t}, x_{2t}, & x_{3t}.
Let $\boldsymbol{Y}_t$ = $(x_{1t}, x_{2t}, x_{3t})'$ & $\boldsymbol{\alpha}$ =(1, $-\beta_1$, $-\beta_2$)' $\Rightarrow \boldsymbol{\alpha}' \boldsymbol{Y}_t = \varepsilon_{1t} \sim I(0)$

<u>Note</u>: The coefficient for x_{1t} (=α_1) is normalized to 1.

- A second CI relationship: x_{2t}, & x_{3t}
Let $\boldsymbol{\alpha}^* = (0, 1, -\beta_3)'$ $\quad \Rightarrow \boldsymbol{\alpha}^{*\prime} \boldsymbol{Y}_t = \varepsilon_{2t} \sim I(0)$.

Cointegration: Example

Example (continuation):
We have 2 C.I. vectors: $\boldsymbol{\alpha}$ & $\boldsymbol{\alpha}^*$. They are independent, that is,

$$\text{rank}\begin{bmatrix} 1 & 0 \\ -\beta_1 & 1 \\ -\beta_2 & -\beta_3 \end{bmatrix} = 2$$

(This is the *cointegrating rank*.)

• Summary for the system (with three time series –i.e., $m = 3$):

We have: $k = 2$ independent C.I. vectors: $\boldsymbol{\alpha}$ & $\boldsymbol{\alpha}$
$m - k = 1$ common stochastic trend (ST): $\sum_{t=1}^{T} \varepsilon_{3t}$.

Cointegration: Long-Run Relation

• Intuition for $I(1)$ case

$\boldsymbol{\alpha}'\boldsymbol{Y}_t$ forms a ***long-run equilibrium***. The system cannot deviate too far from the equilibrium, otherwise economic forces, say arbitrage or competition, will operate to restore the equilibrium. We think of cointegrated variables as variables that "*move together*."

Example: In the previous example, we have two long-run relationships –i.e., two CI relationships:

1) $\varepsilon_{1t} = x_{1t} - \beta_1 x_{2t} - \beta_2 x_{3t}$.
2) $\varepsilon_{2t} = x_{2t} - \beta_3 x_{3t}$

Interpretation: Let's look at CI relation 2). In the *long-run*, when x_{3t} changes by 1 unit, x_{2t} changes by β_3 units. (This is why the CI vector is referred as **long-run parameters**.)

VAR with Cointegration

• Let $\boldsymbol{Y}_t$ be mx1 . Suppose we estimate VAR(p)

$$(\boldsymbol{Y}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}_1 (\boldsymbol{Y}_{t-1} - \boldsymbol{\mu}) + \ldots + \boldsymbol{\Phi}_p (\boldsymbol{Y}_{t-p} - \boldsymbol{\mu}) + a_t$$

or, setting $\boldsymbol{\mu} = 0$,

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}(L)\, \boldsymbol{Y}_{t-1} + \boldsymbol{a}_t$$

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

$$\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}(1) + (1 - L)\, \boldsymbol{\Phi}^*(L)$$

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

$$\boldsymbol{Y}_t = \rho\, \boldsymbol{Y}_{t-1} + \textstyle\sum_{i=1}^{p} \Psi_i\, \Delta\, \boldsymbol{Y}_{t-1} + \boldsymbol{a}_t$$

18

VAR with Cointegration

• Rearranging the equation

$$\Delta \boldsymbol{Y}_t = [\boldsymbol{\Phi}(1) - \boldsymbol{I}]\ \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}^*(L)\ \Delta \boldsymbol{Y}_{t-1} + \boldsymbol{a}_t$$

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

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

2. Cointegration. $0 < Rank[\boldsymbol{\Phi}(1) - \boldsymbol{I}] = k < m$, then we can write

$$[\boldsymbol{\Phi}(1) - \boldsymbol{I}] = \gamma \boldsymbol{\alpha}'$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ are mxk. The equation becomes:

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

• This is called a **vector error correction model** (**VECM**).

19

VAR with Cointegration

• Note: 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 $\Delta \boldsymbol{Y}_t$ and $\boldsymbol{\alpha}'\boldsymbol{Y}_{t-1}$.
- If you know nothing, then you can either

  (i) run OLS in levels, or

  (ii) test (many times) to estimate CI 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.

20

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 - CI vector is known (or pre-specified, say, from 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.

21

Residual Based Tests of the Null of No CI

CASE 2 - CI vector is unknown. It should be estimated.
Thus, if there exists a cointegrating relation, the coefficient on Y_{2t} 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 $\boldsymbol{A}'\boldsymbol{Y}_t$.

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

22

Engle and Granger Cointegration

• Steps in cointegration test procedure:

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

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

23

EG Cointegration: Step 2

• **Step 2:** EG consider the regression model for Y_{1t}:

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

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** $\boldsymbol{\alpha}' = (1, \phi_1, \ldots, \phi_{m-1})$.

• Steps:

1. Run OLS. Get estimate $\hat{\boldsymbol{\alpha}} = (1, \hat{\phi}_1, \ldots, \hat{\phi}_{m-1})$ & residuals, e_t.
2. Use residuals e_t for unit root testing using the ADF or PP tests without deterministic terms (constant or constant and trend).

24

EG Cointegration: Step 2 – Check residuals

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

- Note: $\hat{\phi}_1 \xrightarrow{d}$ t-distribution, 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

$$\Delta e_t = \lambda e_{t-1} + \sum_{i=1}^{p-1} \phi_i \, \Delta e_{t-1} + a_t$$

- 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. Again, we have a **non-standard distribution** (not the DF distribution).

25

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 $\boldsymbol{\alpha}$
- 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.

26

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:

- $T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d}$ non-normal RV not necessarily centered at 0.
- The LS estimate $\hat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}$. Convergence is at rate T, not usual $T^{1/2}$.
  ⇒ We say $\hat{\boldsymbol{\alpha}}$ is **super consistent**.
- $\hat{\boldsymbol{\alpha}}$ is consistent even if the other $(m-1)$ $\boldsymbol{Y}_t$'s are correlated with ε_t.
  ⇒ No asymptotic simultaneity bias.
- The OLS formula for computing aVar($\hat{\boldsymbol{\alpha}}$) is incorrect
  ⇒ usual OLS standard errors are not correct.
- Even though the asymptotic bias → 0, as $T \to \infty$, $\hat{\boldsymbol{\alpha}}$ can be substantially biased in small samples. LS is also not efficient.

27

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 $\boldsymbol{\alpha}$ could be improved upon.

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

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

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

28

EG Cointegration: Estimating VECM with LS

• Consider a bivariate $I(1)$ vector $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})$.
- Assume that $\boldsymbol{Y}_t$ is cointegrated with CI $\boldsymbol{\alpha} = (1, -\alpha_2)$. That is,

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

- Suppose we have a consistent estimate $\hat{\boldsymbol{\alpha}}$ (or $\hat{\boldsymbol{\alpha}}_{\text{DOLS}}$) of $\boldsymbol{\alpha}$.
- We are interested in estimating the VECM for ΔY_{1t} and ΔY_{2t} using:

$$\Delta Y_{1t} = c_1 + \beta_1\, \boldsymbol{\alpha}'\boldsymbol{Y}_{t-1} + \sum_{j=1}^{p-1} \psi_{11,j}\, \Delta Y_{1t-j} + \sum_{j=1}^{p-1} \psi_{12,j}\, \Delta Y_{2t-j} + u_{1t}$$

$$\Delta Y_{2t} = c_2 + \beta_2\, \boldsymbol{\alpha}'\boldsymbol{Y}_{t-1} + \sum_{j=1}^{p-1} \psi_{21,j}\, \Delta Y_{1t-j} + \sum_{j=1}^{p-1} \psi_{22,j}\, \Delta Y_{2t-j} + u_{2t}$$

• $\hat{\boldsymbol{\alpha}}$ is super consistent. It can be treated as known in the ECM. The estimated disequilibrium error $\widehat{\boldsymbol{\alpha}'}\boldsymbol{Y}_t = Y_{1t} - \hat{\alpha}_2\, Y_{2t}$ may be treated like the known $\boldsymbol{\alpha}'\boldsymbol{Y}_t$.

• All variables are $I(0)$, we can use OLS (or SUR to gain efficiency.) 29

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:

$$\boldsymbol{Y}_t = \delta \boldsymbol{D}_t + \phi_1 \boldsymbol{Y}_{t-1} + \ldots + \phi_p \boldsymbol{Y}_{t-p} + \varepsilon_t$$

where $\boldsymbol{Y}_t$ is a time series mx1 vector of $I(1)$ variables.

• The VAR(p) model is stable if

$$\det(\mathbf{I} - \phi_1 z - \cdots - \phi_p z^p) = 0$$

has all roots outside the complex unit circle.

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

30

Johansen Tests

• If $\boldsymbol{Y}_t$ 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** & the **maximal eigenvalue test**. They are based on Granger's (1981) ECM representation. Tests are easy to implement.

Example: Trace test simple idea:

(1) Assume ε_t are multivariate $N_m(\mathbf{0}, \boldsymbol{\Sigma})$. 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.

Johansen Tests: Intuition

• Consider the VECM

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma_0}\, \boldsymbol{D}_t + \boldsymbol{\Pi}\, Y_{t-1} + \boldsymbol{\Gamma}_1\, \Delta Y_{t-1} + \dots + \boldsymbol{\Gamma}_p\, \Delta Y_{t-p+1} + \varepsilon_t$$

where

- $\boldsymbol{D}_t$: vector of deterministic variables (constant, trends, and/or seasonal dummy variables);
- $\boldsymbol{\Gamma}_j = -\boldsymbol{I} + \boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_j, \quad j = 1, 2, \dots, p-1$ are $m \times m$ matrices;
- $\boldsymbol{\Pi} = \boldsymbol{\gamma} \boldsymbol{A}'$ is the long-run impact matrix; $\boldsymbol{A}$ & $\boldsymbol{\gamma}$ are $m \times k$ matrices;
- ε_t are *i.i.d.* $N_m(\mathbf{0}, \boldsymbol{\Sigma})$ errors;
- $\det(\boldsymbol{I} - \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j\, L^j)$ has all of its roots outside the unit circle.

• In this framework, CI happens when $\boldsymbol{\Pi}$ has **reduced rank**. This is the basis of the test: By checking the rank of $\boldsymbol{\Pi}$, we can determine if the system is CI.

Johansen Tests: Intuition

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

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma_0}\, \boldsymbol{D}_t + \boldsymbol{\Pi}^* Y_{t-p} + \sum_{j=1}^{p-1} \boldsymbol{\Gamma}^*{}_j\ \Delta Y_{t-j} + \varepsilon_t$$

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 $Z_{0t} = \Delta \boldsymbol{Y}_t$, $Z_{1t} = \boldsymbol{Y}_{t-1}$ and $Z_{2t} =$ ($\Delta \boldsymbol{Y}_{t-1}$,..., $\Delta \boldsymbol{Y}_{t-p-1}$, D_t)'

• Now, we can write: $Z_{0t} = \boldsymbol{\gamma A'} Z_{1t} + \boldsymbol{\Psi}\, Z_{2t} + \boldsymbol{\varepsilon}_t$

where $\boldsymbol{\Psi}$ =($\boldsymbol{\Gamma}_1$ $\boldsymbol{\Gamma}_2$... $\boldsymbol{\Gamma}_p$ $\boldsymbol{\Gamma}_0$).

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

33

Johansen Tests: Intuition

• Assume the VECM errors are independent $N_m(\mathbf{0}, \boldsymbol{\Sigma})$ 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 $k = m-1, \ldots, 1, 0$.

• 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 ≤ k, vs. H_1: CI rank > k.

- If the LR > critical value for a certain rank, ⇒ reject H_0.

34

Johansen Tests: Intuition

• Johansen concentrates all the parameter matrices in the likelihood function out, except for the matrix $\boldsymbol{A}$.

• Then, he shows that the MLE of $\boldsymbol{A}$ 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.

Note: The factorization $\mathbf{\Pi} = \boldsymbol{\gamma}\boldsymbol{A}'$ is not unique since for any kxk nonsingular matrix $\mathbf{F}$ we have:

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

35

Johansen Tests: Intuition

Note: The factorization $\mathbf{\Pi} = \boldsymbol{\gamma}\boldsymbol{A}'$ is not unique since for any kxk nonsingular matrix $\mathbf{F}$ we have:

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

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

36

Johansen Tests: Sequential Tests

• The Johansen tests examine H_0: Rank($\mathbf{\Pi}$) $\leq k$, 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 $\mathbf{\Pi}$ is less than or equal to k. This creates a nested set of models

$$H(0) \subset \ldots . \subset H(k) \subset \ldots .. \subset H(m)$$

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

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

37

Johansen Tests – Sequential Tests

• Sequential tests:

i. H_0: $k = 0$,	cannot be rejected →stop → k=0
(at most zero cointegration)	rejected →next test
ii. H_0: $k \leq 1$,	cannot be rejected →stop → k=1
(at most one cointegration)	rejected →next test
iii. H_0: $k \leq 2$,	cannot be rejected →stop → k=2
(at most two cointegration)	rejected →next test

• Possible outcomes:
- ***Rank*** $k = m \Rightarrow$ All variables in $\boldsymbol{Y}_t$ are $I(0)$, not an interesting case.
- ***Rank*** $k = 0$ $\Rightarrow$ No linear combinations of $\boldsymbol{Y}_t$ are $I(0)$. ($\mathbf{\Pi} = 0$.)
  $\Rightarrow$ Model on differenced series
- ***Rank*** $k \leq (m-1) \Rightarrow$ Up to $(m-1)$ CI relationships $\boldsymbol{\alpha}'\boldsymbol{Y}_t$
  $\Rightarrow$ VECM

38

Johansen Tests: Sequential Tests

• The Johansen tests examine H_0: Rank($\mathbf{\Pi}$) $\leq k$, where k is less than m

Recall, Rank($\mathbf{\Pi}$) = number of non-zero eigenvalues of $\mathbf{\Pi}$.

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

• It turns out the LR test statistic is the trace of a diagonal matrix of generalized eigenvalues from $\mathbf{\Pi}$.

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

39

Johansen Tests: Likelihood

• Back to the VECM(p) representation:

$$\Delta\boldsymbol{Y}_t = \mathbf{\Gamma_0}\,\boldsymbol{D}_t + \mathbf{\Pi}Y_{t-p} + \textstyle\sum_{j=1}^{p-1}\mathbf{\Gamma}_j\ \Delta Y_{t-j} + \varepsilon_t$$

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 $Z_{0t} = \Delta\boldsymbol{Y}_t$, $Z_{1t} = \boldsymbol{Y}_{t-1}$ & $Z_{2t} = (\Delta\boldsymbol{Y}_{t-1},..., \Delta\boldsymbol{Y}_{t-p-1}, D_t)$'

Now, we can write: $Z_{0t} = \boldsymbol{\gamma}\boldsymbol{A}'Z_{1t} + \mathbf{\Psi}\, Z_{2t} + \boldsymbol{\varepsilon}_t$

where $\mathbf{\Psi}$ =($\mathbf{\Gamma}_1\ \mathbf{\Gamma}_2 ...\ \mathbf{\Gamma}_p\ \mathbf{\Gamma}_0$).

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

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

40

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:

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

• The calculations are based on the sample cross-products matrices:

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

• Then, the MLE for **A** is obtained from the eigenvectors, **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} & R_{1t}. The **V**'s corresponding to the k largest λ's are k linear combinations of $\boldsymbol{Y}_{t-1}$.

41

Johansen Tests: Eigenvalues

• The eigenvectors corresponding to the k largest λ's are the k linear combinations of $\boldsymbol{Y}_{t-1}$, 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 $|\lambda S_{11} - S_{10}\, S_{00}{}^{-1} S_{01}| = 0,$ pre and post multiply the expression by $S_{11}{}^{-1/2}$ (Cholesky decomposition of S_{11}). Then, we have a standard eigenvalue problem.

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

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

42

Johansen Tests: Eigenvalues

• The tests are based on the λ's from

$$|\lambda \boldsymbol{I} - S_{11}{}^{-1/2} S_{10}\, S_{00}{}^{-1} S_{01} S_{11}{}^{-1/2}| = 0.$$

• Interpretation of the eigenvalue equation.

Using F-W, we regress R_{0t} on R_{1t}, to estimate $\boldsymbol{\Pi} = \boldsymbol{\gamma A'}$. That is

$$\widehat{\boldsymbol{\Pi}} = S_{11}{}^{-1} S_{10}$$

Note that

$$S_{11}{}^{-1/2} S_{10}\, S_{00}{}^{-1} S_{01} S_{11}{}^{-1/2} =$$

$$= S_{11}{}^{1/2} S_{11}{}^{-1} S_{10}\, S_{00}{}^{-1/2} S_{00}{}^{-1/2} S_{01} S_{11}{}^{-1} S_{11}{}^{1/2}$$

$$= S_{11}{}^{1/2}\, \widehat{\boldsymbol{\Pi}}\, S_{00}{}^{-1/2} S_{00}{}^{-1/2} \widehat{\boldsymbol{\Pi}}\, S_{11}{}^{1/2}$$

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

43

Johansen Tests: Trace Statistic

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

$$\widehat{\boldsymbol{A}} = \boldsymbol{A}_{MLE} = [v_1, v_2, \ldots,, v_k]$$

• Johansen also finds: $\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_{MLE} = S_{01}\widehat{\boldsymbol{A}}$.

• 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

$$|\lambda \boldsymbol{I} - S_{11}{}^{-1/2} S_{10}\, S_{00}{}^{-1} S_{01} S_{11}{}^{-1/2}| = 0.$$

They are LR tests. They do not have the χ^2 asymptotic distribution.

44

Johansen Trace Test

• The tests are LR tests with non-standard distributions.

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

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

Note: 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(Π) $= k_0$ then $\hat{\lambda}_{k_0+1}, \ldots, \hat{\lambda}_m$ should all be close to 0. The $LR_{trace}(k_0)$ should be small since $\ln(1-\hat{\lambda}_i) \approx 0$ for $i > k_0$.

45

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 $\boldsymbol{W}$ as

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

$\widetilde{\boldsymbol{W}}$ is the Brownian motion itself ($\boldsymbol{W}$), or the demeaned or detrended $\boldsymbol{W}$, 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 $m - k_0 = 1, \ldots, 10$.

46

Johansen Maximal EigenvalueTest

• An alternative LR statistic, given by

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

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

$$max\{\int_0^1 (d\boldsymbol{W})\tilde{\boldsymbol{W}}'(\int_0^1 \tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}}'dr)^{-1}\int_0^1 \tilde{\boldsymbol{W}}(d\boldsymbol{W})'\}$$

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

47

Johansen CI Tests: Example

Example: We test for units roots on the Schiller's historical monthly data set (1871 – 2024) for stock prices (SP), earnings (E) and dividends (D). We use the R package **tseries**.

```
Sh_da <- read.csv("https://www.bauer.uh.edu/rsusmel/4397/Shiller_data.csv",
head=TRUE, sep=",")
SP <- Sh_da$P                         # Extract P = S&P500 series
D <- Sh_da$D                          # Extract D = Dividend S&P500 seriesE <-
Sh_da$E                               # Extract E = Earnings S&P500 series
i <- Sh_da$Long_i                     # Extract Long_i = 10-year interest rate series
T <- length(SP)
###  SP = SP500
t0 <-1                                # t0=926 (1948:Jan)
x <- SP[t0:T]
adf.test(x, k=12)
pp.test(x,  type = c("Z(alpha)"))
pp.test(x,  type = c("Z(t_alpha)"))
kpss.test(x, null=c("Level", "Trend"))
```

Autoregressive Unit Root – Testing: Examples

Example (continuation):

	ADF(12)	ADF(8)	PP-Z_{α_0}	PP-Z_t	KPSS
SP	**5.1758** (**0.99**)	**4.9134** (**0.99**)	**18.229** (**0.99**)	**7.458** (**0.99**)	**9.2041** (**0.01**)
E	**3.803** (**0.99**)	**0.37927** (**0.99**)	**4.1169** (**0.99**)	**0.97931** (**0.99**)	**10.123** (**0.01**)
D	**8.1557** (**0.99**)	**4.8629** (**0.99**)	**9.6253** (**0.99**)	**10.092** (**0.99**)	**10.942** (**0.01**)
i	**-2.3349** (**0.44**)	**0.4897** (**0.49**)	**-8.6799** (**0.63**)	**-2.1078** (**0.53**)	**3.1092** (**0.01**)

Conclusion: Very strong evidence for the unit root hypothesis

Johansen CI Tests: Example

Example: We test for cointegration among for $I(1)$ series: SP, E, D and 10-year interest rates, i. All taken from Shiller's historical data. We use the R package urca, function ca.jo.

```
> x_c <- data.frame(SP,D,E,i)
> co_jo <- ca.jo(x_c, ecdet = "const", type="trace", K=2)
> summary(co_jo)
```

Test type: **trace statistic** , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 9.109170e-02 6.862034e-02 5.691714e-03 2.806550e-03 6.928071e-18

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct	
r <= 3 |	**5.19**	7.52	**9.24**	12.97	
r <= 2 |	**15.73**	17.85	**19.96**	24.60	⇒ We cannot reject H_0 (two or less CI relations)
r <= 1 |	**146.95**	32.00	**34.91**	41.07	⇒ We reject H_0 (one or less CI relations)
r = 0 |	**323.27**	49.65	**53.12**	60.16	⇒ We reject H_0 (no cointegration)

Note: Similar conclusion if we use ecdet = "trend". ⇒ Rank($\mathbf{\Pi}$) = 2.

50

Johansen CI Tests: Example

Example (continuation): Now, we use the $LR_{max}(k)$, using function ca.jo, with type= "eigen". setup for a constant & trend.

```
> x_c <- data.frame(SP,D,E,i)
> co_eigen <- ca.jo(x_c, ecdet = "trend", type="eigen", K=2)
> summary(co_eigen)
```

Test type: **maximal eigenvalue statistic (lambda max)**, without linear trend in cointegration

Eigenvalues (lambda):
[1] 9.109170e-02 6.862034e-02 5.691714e-03 2.806550e-03 6.928071e-18

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct	
r <= 3 |	**5.13**	10.49	**12.25**	16.26	
r <= 2 |	**18.26**	16.85	**18.96**	23.65	⇒ We cannot reject H_0 (two or less CI relations)
r <= 1 |	**127.55**	23.11	**25.54**	30.34	⇒ We reject H_0 (one or less CI relations)
r = 0 |	**177.06**	29.12	**31.46**	36.65	⇒ We reject H_0 (no cointegration)

Note: Similar conclusion if we use ecdet = "trend".

51

ML Estimation of the CI VECM

• Suppose we find Rank($\mathbf{\Pi}$) = k, $0 < k < m$. Then, the CI VECM:

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma_0}\, \boldsymbol{D}_t + \boldsymbol{\gamma A}'\boldsymbol{Y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j\ \Delta \boldsymbol{Y}_{t-j} + \varepsilon_t$$

• This is a reduced rank multivariate regression. Johansen derived the ML estimation of the parameters under the reduced rank restriction Rank($\mathbf{\Pi}$) = k.

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

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

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma_0}\, \boldsymbol{D}_t + \boldsymbol{\gamma}\hat{\mathbf{A}}'\boldsymbol{Y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j\ \Delta \boldsymbol{Y}_{t-j} + \varepsilon_t$$

52

ML Estimation of the CI VECM

• Factorization $\mathbf{\Pi} = \boldsymbol{\gamma}\boldsymbol{A}'$ 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 $\mathbf{A}_{\boldsymbol{c},MLE}$.

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

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

Johansen CI Tests: Example

Example (continuation): ca.jo reports the decomposition of $\mathbf{\Pi}$.

```
> summary(co_eigen)

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             SP.l2        D.l2        E.l2         i.l2     trend.l2
SP.l2     1.0000000    1.000000   1.0000000    1.0000000   1.0000000
D.l2     50.7585632 -1053.464186 -54.0274352 -34.2698659 -29.1954350
E.l2    -54.1158522  181.277712  -5.1522761   -6.7771361  -4.4754885
i.l2     11.2117969   -5.069838 -25.9188786  206.0186272  30.4983441
trend.l2  0.0603914    2.995943   0.3712533   -0.1085308  -0.8544632

Weights W:
(This is the loading matrix)

             SP.l2         D.l2          E.l2          i.l2      trend.l2
SP.d -8.125802e-03 -1.340277e-03 -6.858083e-03 -7.469823e-04  4.272233e-18
D.d  -3.388041e-05  1.325556e-07  6.745596e-06  4.970874e-07 -1.577001e-20
E.d   4.305228e-04 -1.664109e-05  2.247754e-04  1.437162e-05 -3.286190e-19
i.d   2.059088e-06  2.840250e-07  3.218113e-05 -1.918360e-05 -2.697507e-20
```

54

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: $\mathbf{A} = [\mathbf{A}_0\ \boldsymbol{\phi}]$ Rank($\boldsymbol{\Pi}$) $\leq k$

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

$\boldsymbol{\phi}$: $(k-s)\times m$ 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.

55

Common Trends

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

• That is, the mx$(m-k)$ matrix $\mathbf{A}^{\perp}$ is orthogonal to the matrix **A**

$\Rightarrow$ columns of $\mathbf{A}^{\perp}$ are orthogonal to the columns of **A**.

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

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

• Reference: Stock and Watson (1988).

56

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.

57

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;
```

58

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

Note: In all series, we cannot reject H_0 (unit root).

59

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

• The fitted VECM(2) is given as

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

60

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		

• Note: System is cointegrated in rank 1 with integrated order 1.

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

• The factorization $\boldsymbol{\Pi} = \boldsymbol{\gamma A}'$

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

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

- The factorization $\mathbf{\Pi} = \boldsymbol{\gamma} \boldsymbol{A}'$

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

63

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

64

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

Variable	Durbin Watson	Normality		ARCH	
		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**

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

65

Example: Lütkepohl (1993) – SAS: Diagnostics

Univariate Model AR Diagnostics

Variable	AR1		AR2		AR3		AR4	
	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

• Note: Except the residuals for y4, no AR effects.

66

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

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