Unit Roots and Co-integration

Topic 5: Co-integration

by Vanessa Berenguer-Rico

University of Oxford

January 2014

Integration

Definition ($I(d)$)

A series with no deterministic components which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $x_t \sim I(d)$.

- In this course, only the values $d = 0$ and $d = 1$ will be considered (e.g. random walk), but many results can be generalized to other cases including the fractional difference model
- Remember: There are substantial differences between a series that is $x_t \sim I(0)$ and another that is $x_t \sim I(1)$. (Random walk!)

Integration

Properties of $x_t \sim I(0)$ (with zero mean):

(i) the variance of x_t is finite

(ii) an innovation has only a temporary effect on the value x_t

(iii) the expected length of times between crossing $x = 0$ is finite

(iv) the autocorrelations, ρ_k, decrease steadily in magnitude for large enough k, so that their sum is finite

Integration

Properties of $x_t \sim I(1)$ ($x_0 = 0$):

(i) the variance of x_t goes to infinity as t goes to infinity

(ii) an innovation has a permanent effect on the value of x_t, as x_t is the sum of all previous changes

(iii) the expected time between crossings of $x = 0$ is infinite

(iv) the theoretical autocorrelations, $\rho_k \rightarrow 1$ for all k as $t \rightarrow \infty$

Co-integration

- If x_t and y_t are both $I(d)$, then it is "**generally**" true that the linear combination
$$z_t = x_t - ay_t,$$
will also be $I(d)$

- David Hendry said to Clive Granger: The difference of two $I(1)$ variables can be $I(0)$...

Co-integration

Engle, R. F. and C. W. J. Granger (1987): “Co-integration and Error Correction: Representation Estimation and Testing,” *Econometrica* 55, 251-276.

Definition (Co-integration)

The components of the vector x_t are said to be co-integrated of order d, b, denoted $x_t \sim CI\,(d, b)$, if (i) all components of x_t are $I\,(d)$; (ii) there exists a vector α $(\neq 0)$ so that $z_t = \alpha' x_t \sim I\,(d - b)$, $b > 0$. The vector α is called the co-integrating vector

Co-integration

- Consider: $d = b = 1$. That is $x_t \sim I(1)$ and $z_t = \alpha' x_t \sim I(0)$
- **Simplest example**: Let $y_t \sim I(1)$, $x_t \sim I(1)$, and

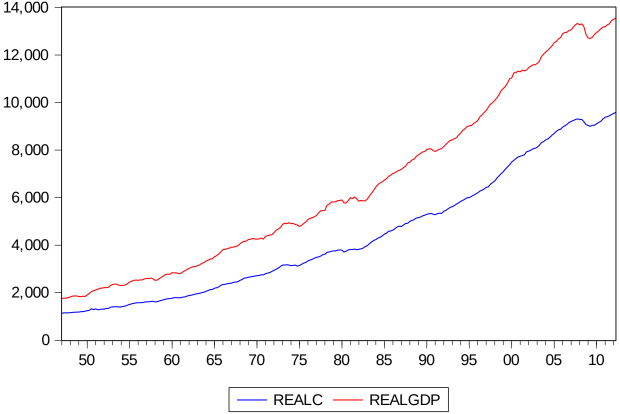
$$y_t = \theta x_t + z_t$$

  with $z_t \sim I(0)$

- Hence, z_t will rarely drift far from zero (if it has zero mean) and will often cross the zero line
- Equilibrium Relationship: $y_t - \theta x_t$; so that u_{yt} represent the stationary deviation from the equilibrium
- Therefore, y_t and x_t will not move too far away from each other

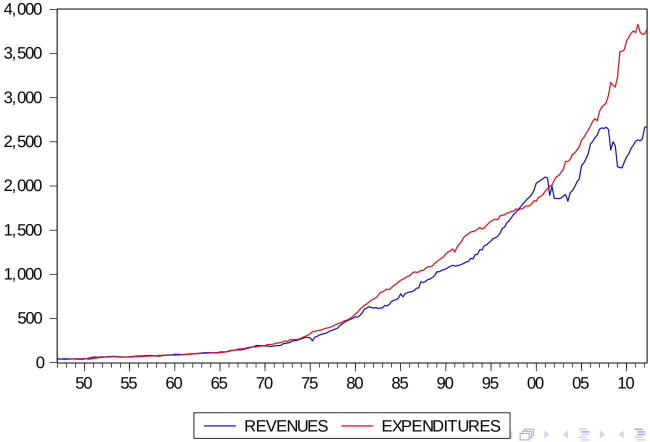
Co-integration

- Real Consumption and Real GDP
- U.S. Quarterly Data from the Federal Reserve Bank of St. Louis: 1947Q1-2012Q2



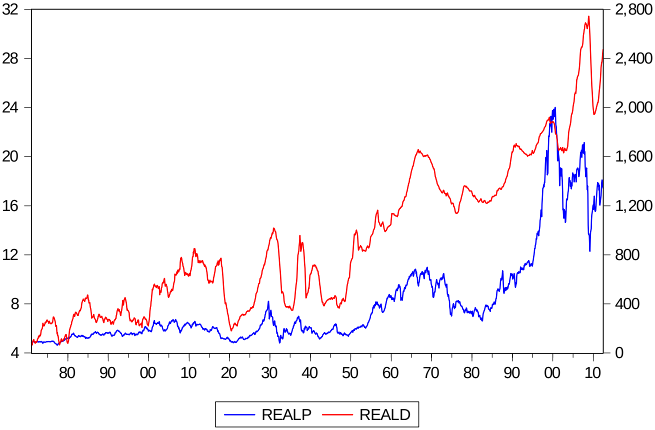
Co-integration

- Government Expenditures and Revenues
- U.S. Quarterly Data from the Federal Reserve Bank of St. Louis: 1947Q1-2012Q2



Co-integration

- Real Stock Prices and Real Dividends
- U.S. Monthly Data from Robert Shiller: 1871m1-2012m6



The common factor explanation

- Example: Let $W_t \sim I(1)$ and consider the following system

$$\begin{aligned} y_t &= aW_t + u_{yt} \\ x_t &= W_t + u_{xt} \end{aligned}$$

where u_{yt} and u_{xt} are both $I(0)$

- Then, $y_t \sim I(1)$ and $x_t \sim I(1)$

- But

$$y_t - ax_t = aW_t + u_{yt} - aW_t + au_{xt} = u_{yt} + au_{xt} \sim I(0)$$

- Cancellation of the common factor W_t!

Estimation and Testing

- Single Equation: Engle and Granger (1987) approach (+DOLS)
- System of Equations: Johansen (1991)

Engle and Granger proposal

- Two stages approach
- *Stage 1*: Estimate the long run (static) relationship by OLS

$$y_t = f(t) + \beta x_t + z_t,$$

and compute the OLS residuals, $\hat{z}_t$. Then, test for co-integration

$$\begin{cases} H_o : \hat{z}_t \sim I(1) \\ H_a : \hat{z}_t \sim I(0) \end{cases}$$

- *Stage 2*: If the null hypothesis is rejected, then estimate the VECM by OLS

$$\Delta y_t = -\gamma \hat{z}_{t-1} + \sum_{i=1}^{p_1} \phi_i \Delta y_{t-i} + \sum_{i=0}^{p_1} \rho_i \Delta x_{t-i} + \varepsilon_t$$

EG Test for Co-integration

- *Stage 1*: Estimate the long run (static) relationship by OLS

$$y_t = f(t) + \beta x_t + z_t,$$

and compute the OLS residuals, $\hat{z}_t$. Then, test for co-integration

$$\begin{cases} H_o : \hat{z}_t \sim I(1) \\ H_a : \hat{z}_t \sim I(0) \end{cases}$$

- **DF test on $\hat{z}_t$**: *Important!* Critical values are not the same as those derived by Dickey and Fuller ($\hat{z}_t$ incorporates the OLS estimates!)

- Critical values tabulated by **MacKinnon (1991)**

EG Estimation Properties

Stage 1:

$$y_t = \theta x_t + z_t$$

In general,

- OLS estimates biased and inefficient (although super-consistent!)
- Nonstandard distributions with nuisance parameters
- There could be more than one cointegrating vector
- Endogeneity

Simplest Example:

$$\begin{aligned} y_t &= \theta x_t + u_t \\ \Delta x_t &= \varepsilon_t \end{aligned}$$

with $u_t \sim i.i.d.N\left(0, \sigma_u^2\right)$, $\varepsilon_t \sim i.i.d.N\left(0, \sigma_\varepsilon^2\right)$, and $E\left[u_t \varepsilon_s\right] = \sigma_{u\varepsilon} 1\left(t = s\right)$.

The OLS estimator of θ satisfies

$$T\left(\hat{\theta}_n - \theta\right) = \frac{T^{-1} \sum_{t=1}^{T} x_t u_t}{T^{-2} \sum_{t=1}^{T} x_t^2}$$

Denominator: Remember $\Delta x_t = \varepsilon_t$ with $\varepsilon_t \sim i.i.d.N\left(0, \sigma_\varepsilon^2\right)$. Hence, we can apply the FCLT and CMT to obtain

$$\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 W_\varepsilon\left(r\right)^2 dr.$$

Numerator: In order to derive the limiting distribution of $T^{-1}\sum_{t=1}^{T} x_t u_t$, it will be convenient to condition u_t on ε_t in the following fashion:

$$u_t = \gamma \varepsilon_t + v_t; \quad \gamma = \sigma_{u\varepsilon}/\sigma_\varepsilon^2; \quad \sigma_v^2 = \sigma_u^2 - \sigma_{u\varepsilon}^2/\sigma_\varepsilon^2.$$

Hence, by construction $E\left[\varepsilon_t v_s\right] = 0 \; \forall t \neq s$. Moreover,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} x_t u_t &= T^{-1}\sum_{t=1}^{T} x_t \left(\gamma \varepsilon_t + v_t\right) \\ &= \gamma\left(T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon_t\right) + \gamma\left(T^{-1}\sum_{t=1}^{T} \varepsilon_t^2\right) \\ &\quad + \left(T^{-1}\sum_{t=1}^{T} x_{t-1} v_t\right) + \left(T^{-1}\sum_{t=1}^{T} \varepsilon_t v_t\right) \end{aligned}$$

Estimation and Testing: Single Equation

Numerator: We already derived that

$$\left(T^{-1}\sum_{t=1}^{T}x_{t-1}\varepsilon_t\right) \xrightarrow{d} \left(\sigma_\varepsilon^2/2\right)\left(W_\varepsilon\left(r\right)^2-1\right)$$

and

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \sigma_\varepsilon^2.$$

By the assumptions above it is easy to see that

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t v_t \xrightarrow{p} 0$$

(why?).

Numerator: Finally, it remains to study

$$T^{-1} \sum_{t=1}^{T} x_{t-1} v_t,$$

which convergences to a stochastic integral! (Kurtz, T.G. & P. Protter, 1991: "Weak limit theorems for stochastic integrals and stochastic differential equations"). Specifically,

$$T^{-1} \sum_{t=1}^{T} x_{t-1} v_t \xrightarrow{d} \sigma_{\varepsilon} \sigma_{v} \int_0^1 W_{\varepsilon}\left(r\right) dW_{v}\left(r\right).$$

Numerator: $\int_0^1 W_\varepsilon(r)\, dW_v(r)$ is an Ito stochastic integral which satisfies

$$\int_0^1 W_\varepsilon(r)\, dW_v(r) \sim N\left(0, \int_0^1 W_\varepsilon^2(r)\, dr\right),$$

see Lemma 5.1 in Park and Phillips (1988) "Statistical Inference with integrated regressors: Part 1"

Numerator: Therefore, the numerator satisfies

$$T^{-1}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} \gamma\left(\sigma_\varepsilon^2/2\right)\left(W_\varepsilon\left(r\right)^2-1\right)+\gamma\sigma_\varepsilon^2+\sigma_\varepsilon\sigma_v\int_0^1 W_\varepsilon\left(r\right)dW_v\left(r\right).$$

Estimation and Testing: Single Equation

OLS:

$$T\left(\hat{\theta}_n-\theta\right)\xrightarrow{d}\frac{\left(\gamma\sigma_\varepsilon^2/2\right)\left(W_\varepsilon\left(r\right)^2-1\right)+\gamma\sigma_\varepsilon^2+\sigma_\varepsilon\sigma_v\int_0^1 W_\varepsilon\left(r\right)dW_v\left(r\right)}{\int_0^1 W_\varepsilon\left(r\right)^2 dr}$$

t-test:

$$t_{\theta=0}\xrightarrow{d}\left(\gamma\sigma_\varepsilon\sigma_u^{-1}/2\right)\left(W_\varepsilon\left(r\right)^2+1\right)\left(\int_0^1 W_\varepsilon\left(r\right)^2 dr\right)^{-1/2}+\sigma_v\sigma_u^{-1}N\left(0,1\right)$$

Estimation and Testing: Single Equation

t-test:

$$t_{\theta=0} \xrightarrow{d} \left(\gamma\sigma_{\varepsilon}\sigma_{u}^{-1}/2\right)\left(W_{\varepsilon}(r)^{2}+1\right)\left(\int_{0}^{1} W_{\varepsilon}(r)^{2}\,dr\right)^{-1/2} + \sigma_{v}\sigma_{u}^{-1}N(0,1)$$

- In general, therefore, the t-ratio of $\hat{\theta}_n$ will not have a standard normal distribution unless $\gamma = \sigma_{u\varepsilon}/\sigma_{\varepsilon}^{2} = 0$ (that is, unless x_t is strongly exogenous for the estimation of θ)
- When $\gamma \neq 0$ the first term in the asymptotic distribution gives rise to "second order" or "endogeneity" bias, which although asymptotically negligible in estimating θ due to super consistency, can be important in finite samples

EG Estimation Properties

Stage 1:

$$y_t = \theta x_t + z_t$$

In general,

- OLS estimates biased and inefficient (although super-consistent!)
- Nonstandard distributions with nuisance parameters
- There could be more than one cointegrating vector
- Endogeneity

DOLS: Dynamic OLS

- Saikkonen (1991): "Asymptotically Efficient Estimation of Cointegration Regressions"
- Stock and Watson (1993): "A Simple Estimator of Co-integrating Vectors in Higher Order Integrated Systems"

DOLS: Dynamic OLS

The DOLS estimator is based on a regression function that incorporates leads and lags of the first differences of the regressors, that is,

$$y_t = x_t\theta + \sum_{j=-k_1}^{k_2} \Delta x_{t-j}\phi_j + u_t.$$

- Accounts for the dynamics when estimating the cointegrating vector
- Accounts for the correlation between the regressors and the error term
- Optimal (consistent and efficient) estimation of the cointegrating vector (mixed-gaussian distributions)
- Standard t-tests, Wald

Co-integration

Recall:

Engle, R. F. and C. W. J. Granger (1987): "Co-integration and Error Correction: Representation Estimation and Testing," *Econometrica* 55, 251-276.

Definition (Co-integration)

The components of the vector x_t are said to be co-integrated of order d, b, denoted $x_t \sim CI(d, b)$, if (i) all components of x_t are $I(d)$; (ii) there exists a vector α $(\neq 0)$ so that $z_t = \alpha' x_t \sim I(d - b)$, $b > 0$. The vector α is called the co-integrating vector

Co-integration

- Consider: $d = b = 1$. That is $x_t \sim I(1)$ and $z_t = \alpha' x_t \sim I(0)$
- **Recall**: Simplest example: Let $y_t \sim I(1)$, $x_t \sim I(1)$, and

$$y_t = \theta x_t + z_t$$

  with $z_t \sim I(0)$

- Hence, z_t will rarely drift far from zero (if it has zero mean) and will often cross the zero line
- Equilibrium Relationship: $y_t - \theta x_t$; so that u_{yt} represent the stationary deviation from the equilibrium
- Therefore, y_t and x_t will not move too far away from each other

Remarks:

- If x_t has $m > 2$ components, then there can be more than one cointegrating vector α ($\alpha' x_t \sim I(0)$)
- It is assumed that there are exactly r linearly independent co-integrating vectors with $r \leq m - 1$
- It is useful to gather the co-integrating vectors together into the $m \times r$ array α
- So, by construction the rank of α will be r and it will be called the "co-integrating rank" of x_t

Co-integration: Representation

<u>**Representations of a Co-integrated System**</u>: (Watson, 1994)

- VAR: $x_t = \sum_{i=1}^{p} \Pi_i x_{t-i} + \varepsilon_t$
- VECM: $\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta x_{t-i} + \varepsilon_t$
- MA: $\Delta x_t = C(L)\, \varepsilon_t$
- Common Trends: $x_t = C(1) \sum_{s=1}^{t} \varepsilon_s + C^*(L)\, \varepsilon_t + x_0$
- Triangular: $x_t = (x'_{1t}, x'_{2t})'$ with $\Delta x_{1t} = u_{1t}$ and $x_{2t} - \beta x_{1t} = u_{2t}$

Co-integration: Representation

<u>**Simplest Example**</u>: Let x_t be a 2×1 vector, $x_t = (x_{1t}, x_{2t})'$.

- <u>*Triangular Representation*</u>

$$\begin{aligned} x_{1t} &= \theta x_{2t} + \varepsilon_{1t} \\ x_{2t} &= x_{2t-1} + \varepsilon_{2t} \end{aligned}$$

- <u>*MA Representation*</u>

$$\begin{aligned} \Delta x_{1t} &= \Delta \varepsilon_{1t} + \theta \varepsilon_{2t} \\ \Delta x_{2t} &= \varepsilon_{2t} \end{aligned}$$

or equivalently

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} (1-L) & \theta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Hence,

$$\Delta x_t = C(L)\, \varepsilon_t$$

Co-integration: Representation

- *Common Trends Representation*: From the MA representation

$$\Delta x_t = C(L)\varepsilon_t = C(1)\varepsilon_t + [C(L) - C(1)]\varepsilon_t$$

and solving backward for the levels of x_t

$$x_t = C(1)\sum_{s=1}^{t}\varepsilon_s + C^*(L)\varepsilon_t + x_0$$

where $C^*(L) = (1-L)^{-1}[C(L) - C(1)]$. In particular, for $x_0 = 0$,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 0 & \theta \\ 0 & 1 \end{pmatrix} \sum_{s=1}^{t}\varepsilon_s + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

or equivalently

$$x_{1t} = \theta\sum_{s=1}^{t}\varepsilon_{2s} + \varepsilon_{1t} \text{ and } x_{2t} = \sum_{s=1}^{t}\varepsilon_{2s}$$

Co-integration: Representation

- *VECM Representation*: From the triangular representation

$$
\begin{aligned}
x_{1t} &= \theta x_{2t} + \varepsilon_{1t} \\
x_{2t} &= x_{2t-1} + \varepsilon_{2t}
\end{aligned}
$$

or equivalently

$$
\begin{aligned}
\Delta x_{1t} &= -x_{1t-1} + \theta x_{2t-1} - \theta x_{2t-1} + \theta x_{2t} + \varepsilon_{1t} \\
\Delta x_{2t} &= \varepsilon_{2t}
\end{aligned}
$$

that is

$$
\begin{aligned}
\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} \varepsilon_{1t-1} + \begin{pmatrix} 1 & \theta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\
&= \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & -\theta \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 & \theta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.
\end{aligned}
$$

- *VECM Representation*

That is

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \varepsilon_{1t-1} + \begin{pmatrix} 1 & \theta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$
$$= \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & -\theta \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 & \theta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Hence,

$$\begin{aligned} \Delta x_t &= \Pi x_{t-1} + \tilde{\varepsilon}_t \\ &= \alpha \beta' x_{t-1} + \tilde{\varepsilon}_t \end{aligned}$$

Co-integration: Representation

- *VAR Representation*

From the VECM

$$x_t = (\Pi - I)\, x_{t-1} + \tilde{\varepsilon}_t,$$

or equivalently

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} -2 & \theta \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \tilde{\varepsilon}_{1t} \\ \tilde{\varepsilon}_{2t} \end{pmatrix},$$

that is a VAR(1)

$$x_t = \Pi_1 x_{t-1} + \tilde{\varepsilon}_t,$$

Granger Representation Theorem

- It proves, in a general context, that a co-integrated system of variables can be represented in three (four) main forms: the vector autoregressive (VAR), error correction, and moving-average (common factors) forms
- These representations are all isomorphic to each other
- The theorem establishes the restrictions that hold between the lag-polynomial matrices in each representation of the process
- Engle and Granger (1987) or Johansen (1991)
- Common factor representation (Stock and Watson, 1988); Triangular representation (Phillips, 1991)

Co-integration: Representation

<u>**Granger Representation Theorem**</u> (Johansen, 1991)

Consider a general VAR model with Gaussian errors written in the error correction form

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \Phi D_t + \mu + \varepsilon_t,$$

where D_t are seasonal dummies orthogonal to the constant term. Further, ε_t are independent p-dimensional Gaussian variables with mean zero and variance matrix Λ. This model can be rewritten as

$$\Pi(L) X_t = -\Pi X_t + \Psi(L) \Delta X_t = \varepsilon_t + \mu + \Phi D_t,$$

where $\Psi(L) = (\Pi(L) - \Pi(1)) / (1 - L)$. Note that $-\Pi = \Pi(1)$ and $-\Psi = -\Psi(1)$ is the derivative of $\Pi(z)$ for $z = 1$.

Co-integration: Representation

Theorem (*Granger Representation Theorem*): *Let the process X_t be as defined above and let*

$$\Pi = \alpha\beta',$$

for α and β of dimensions $p \times r$ and rank r, and let

$$\alpha_{\perp}' \Psi \beta_{\perp},$$

have full rank $p - r$. We define

$$C = \beta_{\perp} \left(\alpha_{\perp}' \Psi \beta_{\perp}\right)^{-1} \alpha_{\perp}'.$$

Then, ΔX_t and $\beta' X_t$ can be given initial distributions, such that

(i) *ΔX_t is stationary*

(ii) *$\beta' X_t$ is stationary*

(iii) *X_t is nonstationary, with linear trend $\tau t = C\mu t$*

Co-integration: Representation

Further,

(iv) $E\left(\beta' X_t\right) = -\left(\alpha'\alpha\right)^{-1}\alpha'\mu + \left(\alpha'\alpha\right)^{-1}\alpha'\Psi\beta_{\perp}\left(\alpha'_{\perp}\Psi\beta_{\perp}\right)^{-1}\alpha'_{\perp}\mu,$

(v) $E\left(\Delta X_t\right) = \tau,$

apart from terms involving the seasonal dummies. If $\alpha'_{\perp}\mu = 0$*, then* $\tau = 0$ *and the linear trend disappears.*

...

Co-integration: Representation

(vi) *If the initial distributions are expressed in terms of the doubly infinite sequence* $\{\varepsilon_t\}$, *then* ΔX_t *has a representation*

$$\Delta X_t = C(L)(\varepsilon_t + \mu + \Phi D_t),$$

with $C(1) = C$

(vii) *For* $C_1(L) = (C(L) - C(1))/(1-L)$, *so that* $C(L) = C(1) + (1-L)C_1(L)$, *the process* X_t *has the representation*

$$X_t = X_0 + C\sum_{i=1}^{t}\varepsilon_i + \tau t + C(L)\Phi\sum_{i=1}^{t}D_i + S_t - S_0,$$

where $S_t = C_1(L)\varepsilon_t$, *and* $\beta' X_0 = \beta' S_0$.

Estimation and Testing

- Single Equation: Engle and Granger (1987) approach (+DOLS)

- System of Equations: Johansen (1991)

Estimation and Testing: System of Equations

System of Equations: Johansen's Approach

- Uses likelihood methods for the analysis of cointegration in VAR models with Gaussian errors
- Likelihood ratio test of cointegration rank (nonstandard inference)
- Tests of structural hypothesis about cointegrating relationships (standard inference)

The Cointegrated VAR

Let $Y_t = (y_{1t}, y_{2t}, ..., y_{mt})'$ be generated as a VAR(p)

$$Y_t = \Pi_1 Y_{t-1} + ... + \Pi_p Y_{t-p} + \varepsilon_t,$$

with $\varepsilon_t \sim i.i.d.N(0, \Omega)$. (No deterministic terms). The VECM representation is

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t,$$

where

$$\Pi = -\left(I - \Pi_1 - ... - \Pi_p\right),$$

and

$$\Gamma_i = -\left(\Pi_{i+1} + \Pi_{i+2} + ... + \Pi_p\right).$$

(Try a VAR(2)!)

Johansen's Methodology in Practice

1. Specify and estimate a VAR(p) model for Y_t

2. Construct likelihood ratio tests for the rank of Π to determine the number of cointegrating vectors

3. If necessary, impose normalization and identifying restrictions on the cointegrating vectors

4. Given the normalized cointegrating vectors estimate the resulting cointegrated VECM by maximum likelihood

5. Test relevant hypothesis

Estimation and Testing via a ML approach

The VECM is

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t,$$

with $\varepsilon_t \sim i.i.d.N\left(0, \Omega\right)$. Hence, the log-likelihood function is

$$\log L\left(\Gamma, \Pi, \Omega\right) = -\frac{1}{2} T \log|\Omega| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \varepsilon_t.$$

Concentrate the log-likelihood function

Step 1: Concentrate $\log L$ with respect Γ

Step 2: Concentrate $\log L$respect Ω

Step 3: Concentrate $\log L$ with respect α

Estimation and Testing via a ML approach

Recall: The log-likelihood function is

$$\log L(\Gamma, \Pi, \Omega) = -\frac{1}{2} T \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \varepsilon_t.$$

Notice that the VECM can be rewritten as

$$Z_{0t} = -\Pi Z_{1t} + \Gamma Z_{2t} + \varepsilon_t,$$

where $Z_{0t} = \Delta Y_t$, $Z_{1t} = Y_{t-1}$, $Z_{2t} = \left(\Delta Y_{t-1}', ..., \Delta Y_{t-p+1}'\right)'$, $\Pi = -\alpha\beta'$, and Γ consist of the parameters $\left(\Gamma_1, ..., \Gamma_{p-1}\right)$.

Estimation and Testing: System of Equations

Step 1: Concentrate $\log L$ with respect Γ

The first order conditions for Γ are

$$\sum_{t=1}^{T} \left(Z_{0t} - \alpha\beta' Z_{1t} + \Gamma Z_{2t} \right) Z_{2t}' = 0,$$

or equivalently

$$M_{02} = \alpha\beta' M_{12} + \Gamma M_{22}.$$

This implies that, for fixed α and β,

$$\hat{\Gamma}\left(\alpha, \beta\right) = M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1},$$

and hence

$$\log L\left(\alpha, \beta, \Omega\right) = -\frac{1}{2} T \log\left|\Omega\right| - \frac{1}{2} \sum_{t=1}^{T} \left(R_{0t} - \alpha\beta' R_{1t} \right) \Omega^{-1} \left(R_{0t} - \alpha\beta' R_{1t} \right),$$

where R_{0t} and R_{1t} are the residuals obtained from regressing Z_{0t} and Z_{1t} against Z_{2t}, respectively.

Estimation and Testing: System of Equations

Step 2: Concentrate $\log L$ with respect Ω

$$\begin{aligned}\log L(\alpha,\beta) &= -\frac{Tm}{2}\log(2\pi) - \frac{Tm}{2} \\ &\quad -\frac{T}{2}\log\left|\frac{1}{T}\sum_{t=1}^{T}\left(R_{0t}-\alpha\beta' R_{1t}\right)\left(R_{0t}-\alpha\beta' R_{1t}\right)\right|.\end{aligned}$$

(Why? Hint: $\partial \log|\Omega| / \partial\Omega = (\Omega')^{-1}$,
$\partial tr\left(BA^{-1}C\right)/\partial A = -\left(A^{-1}CBA^{-1}\right)'$,
$\sum_{t=1}^{T}\left(R_{0t}-\alpha\beta' R_{1t}\right)\hat{\Omega}^{-1}\left(R_{0t}-\alpha\beta' R_{1t}\right) = Tm$).

Step 3: Concentrate $\log L$ with respect α and notice that maximizing $\overline{\log L}(\beta)$ is equivalent to minimize

$$\left| S_{00} - S_{01}\beta \left(\beta' S_{11}\beta\right)^{-1} \beta S_{10} \right|,$$

where $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}'$, $i, j = 0, 1$.
(Why?)

Estimation and Testing: System of Equations

Moreover,

$$\left|S_{00} - S_{01}\beta\left(\beta' S_{11}\beta\right)^{-1}\beta S_{10}\right| = \frac{|S_{00}|\left|\beta'\left(S_{11} - S_{10}S_{00}S_{01}\right)^{-1}\beta\right|}{\left|\beta' S_{11}\beta\right|}$$

$$= |S_{00}|\prod_{i=1}^{r}\left(1 - \lambda_i\right),$$

where λ_i, $i = 1, ..., r$ denote the r largest eigenvalues obtained from

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

(*Reduced Rank regression*: Anderson, 1951)

Estimation and Testing: System of Equations

For a given r and largest eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_r$, we obtain $\hat{\beta}_i$, $i = 1, ..., r$, the corresponding eigenvectors, from

$$\left(\hat{\lambda}_i S_{11} - S_{10} S_{00} S_{01}^{-1}\right) \hat{\beta}_i = 0 \quad i = 1, ..., r$$

and

$$\begin{aligned} \hat{\alpha} &= S_{01} \hat{\beta} \\ \hat{\Pi} &= -\hat{\alpha} \hat{\beta}' \\ \hat{\Omega} &= S_{00} - \hat{\alpha} \hat{\alpha}'. \end{aligned}$$

In practice, the number of cointegrating vectors, r, is unknown!

Testing for Co-integration

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t,$$

Three situations of interest:

(i) The rank of Π is zero: There are no co-integrating relationships

(ii) The rank of Π is m: All variables in Y_t are stationary

(iii) The rank of Π is $r < m$: There exist r cointegrating vector and $\Pi = -\alpha\beta'$ where α and β are $(m \times r)$ matrices

Estimation and Testing: System of Equations

- Hence, testing for co-integration is equivalent to test for reduced rank of Π. In other words, testing for cointegration is equivalent to find the number of r linearly independent columns of Π

- Johansen's maximum likelihood approach to solve this problem amounts to a reduced rank regression which provides m eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_m$, and their corresponding eigenvector $\hat{V} = (\hat{v}_1, \hat{v}_2, ..., \hat{v}_m)$

- Those r elements in $\hat{V}$ which determine the linear combinations of stationary relationships can be denoted $\hat{\beta} = (\hat{v}_1, \hat{v}_2, ..., \hat{v}_r)$. The last $(m - r)$ combinations indicate the non-stationary combinations

- Each eigenvector $\hat{v}_i$ has a corresponding eigenvalue $\hat{\lambda}_i$ and, in particular, the eigenvectors corresponding to the non-stationary part of the model equal are equal zero

Estimation and Testing: System of Equations

Two alternative likelihood ratio tests:

(a) **Trace test**: This procedure tests that there are at most r cointegrating vectors (and thus $(m - r)$ unit roots). Hence, $H_o : \lambda_i = 0$ $i = r + 1, ..., m$. The test statistic is

$$\lambda_{trace}\left(r\right) = -T \sum_{i=r+1}^{m} \ln\left(1 - \hat{\lambda}_i\right)$$

(b) **Max eigenvalue test**: This tests that there are r cointegrating vectors against the alternative that there are $r + 1$. The test statistic in this case is

$$\lambda_{\max}\left(r\right) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right) \qquad r = 0, 1, 2, ..., m - 1$$

Remark 1: Sequential procedures
Remark 2: Both statistics have nonstandard distributions (functionals of Brownian motions)!

Estimation and Testing: System of Equations

Both statistics are based on the Likelihood Ratio test. Notice that

$$\max \log L = -\frac{Tm}{2} \log (2\pi) - \frac{Tm}{2} - \frac{T}{2} \left[\log |S_{00}| + \sum_{i=1}^{r} \log (1 - \lambda_i) \right].$$

Therefore,

$$\begin{aligned} LR(r_0, r_1) &= 2 [\log L(r_1) - \log L(r_0)] \\ &= T \left[-\sum_{i=1}^{r_1} \log (1 - \lambda_i) + \sum_{i=1}^{r_0} \log (1 - \lambda_i) \right] \\ &= -T \sum_{i=r_0+1}^{r_1} \log (1 - \lambda_i) \end{aligned}$$