Advanced Econometrics: Time Series

TOPIC 4

Cointegration

Plan

- Define Cointegration
- Work through properties of a cointegrated system
  - Common Trend Representation
  - Vector Error Correction Mechanism representation
  - Estimation and Tests

Another resource:

- I found another excellent set of lecture notes from Eric Zivot
  - He is co-author of one of the most used stationarity tests under an arbitrary structural break
  - https://faculty.washington.edu/ezivot/

Multivariate Concern: Spurious Regression

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + e_t$$

- Regressing two ***independent*** I(1) processes on each other gives misleading results
  - Easily find significant coefficients, high R^2
  - When the processes are pure random walks (no drift) the β_1 estimates can *still* be centred around the truth (i.e. a zero) (at least in my simulations)
  - The distribution is not standard (wider than a t-distribution), however, so the standard t tests have the wrong *size*
  - I.e. using standard t-tests, one will *under-reject* the null hypothesis of zero coefficient
  - Phillips (1986) shows that in the general case (allowing for drift and more than one I(1) regressor, the OLS t-stats diverge and the R-squared converges to 1

Monte Carlo Experiment: Spurious regression

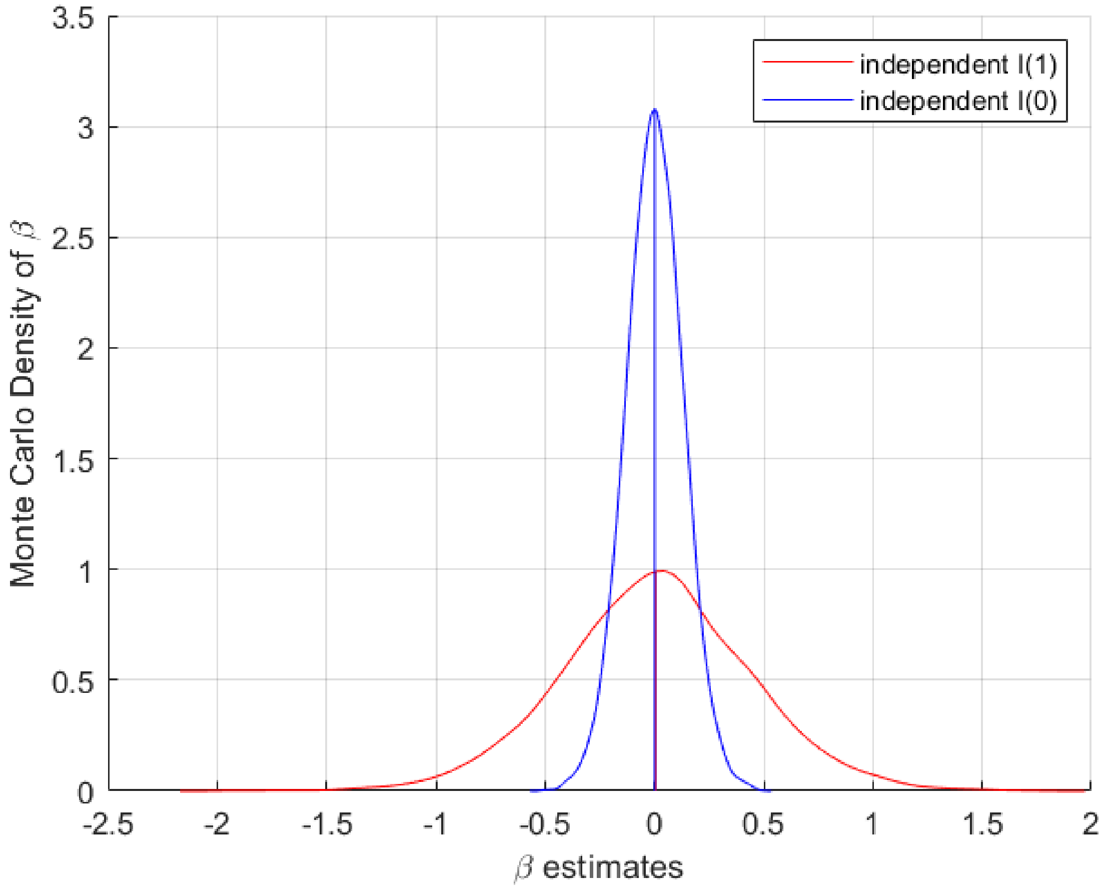
I generated 5000 independent samples of 100 periods:

- Two independent AR(1) processes with coefficient 0.5
- Two independent random walks with drift

Then performed regressions:

- the stationary variables on the stationary variables, plus constant and linear time trend
- the non-stationary variables on the nonstationary variables, plus constant and linear time trend

I plot the kernel density estimate of the distributions of each set of coefficients



Multivariate Concern: Spurious Regression

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + e_t$$

- When two I(1) process are not independent, but ***cointegrated*** a regression of one on the other gives valid results
  - It is defined by a situation where there exists a linear combination of the I(1) variables that is stationary
  - Then a regression is valid, but *super-consistent* – to obtain the limiting distribution, one must scale with T rather than $\sqrt{T}$:
  - The limiting distribution is not standard

Monetary example

- Consider a central bank that follows a policy that implies a money supply that is a random walk with drift:

- How would aggregate money and prices behave if production was stationary and prices perfectly flexible?

Monetary example

- Consider a central bank that follows a policy that implies the log of the money supply that is a random walk with drift, i.e. m_t is an $I(1)$ process, or equivalently *integrated*

$$m_t = a_0 + m_{t-1} + \varepsilon_{M,t}$$

- How would aggregate money M_t and prices P_t behave if production Y_t was stationary and prices perfectly flexible and the velocity of money is constant V?
  - If prices are perfectly flexible, it is optimal for firms to immediately internalize an increase in money supply by increasing their prices proportionally
  - In simple models we have the following prediction:

$$M_t V = P_t Y_t$$

  - In logarithms:

$$m_t + v = p_t + y_t$$
$$p_t = v + m_t - y_t$$

  - Thus the log-price series "inherits" the unit root of the log money supply
  - Moreover, a linear combination of log money supply and log price level is stationary. They are called **co**integrated.

$$p_t - m_t = v - y_t$$

Cointegration Defined

- Simplest case: two variables y_{1t} and y_{2t} are defined to be linearly *cointegrated* if:
  - Each variable has a unit root

$$y_{1t}, y_{2t} \sim I(1)$$

  - And there exists a linear combination (b_1, b_2) of them which is stationary:

$$b_1 y_{1t} + b_2 y_{2t} \sim I(0)$$

The typical notation for this is $CI(1,1)$

Cointegration Defined

Definition: the $I(p)$ process $\mathbf{y}_t$ is said to be *linearly cointegrated* of order q denoted as $\mathbf{y}_t \sim CI(p, q)$with $p, q \in \mathbb{Z}^{++}$, $p > q \geq 1$ if:
$\exists$ at least one non-zero vector $\mathbf{b} \in \mathbb{R}^n$, unique up to a scalar multiple, such that $\mathbf{b}\mathbf{y}_t$ is $I(p-q)$

What are we not looking at?

- We are only considering *linear* cointegration. In principle a non-linear relationship between two $I(1)$ processes can be stationary. I have not seen this employed yet

- We will focus on $CI(1,1)$ processes. Enders gives some analysis and references of higher order applications in the literature.

General two variable case

Carefully consider the properties of the cointegrating vector

- Scaling
- Uniqueness

Let

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \sim CI(1,1)$$

and let the cointegrating relationship be:

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}$$

Thus:

$$y_{1t} \sim I(1) \text{ and } y_{2t} \sim I(1), \text{ but } \mathbf{by}_t = b_1 y_{1t} + b_2 y_{2t} \sim I(0).$$

General two variable case

Define: $u_t = b_1 y_{1t} + b_2 y_{2t}$

By definition, u_t is stationary

- Note: we can allow for a constant in the cointegrating relationship, which will mean that the expected value of u_t could be a non-zero constant independent of time: $E(u_t) = \mu_u$

Thus:

$$E(u_t) = E(u_s) = \mu_u \quad \forall \quad t, s$$

$$E(u_t)^2 = E(u_s)^2 = \sigma_u^2 \quad \forall \quad t, s$$

$$E[u_t u_{t-s}] = \sigma_s \quad \forall \quad t, s$$

Now consider the properties of λu_t for any non-zero $|\lambda| < \infty$.

General two variable case

Define: $$u_t = b_1 y_{1t} + b_2 y_{2t}$$

We find that , λu_t is also stationary

$$E(\lambda u_t) = E(\lambda u_s) = \lambda \mu_u \quad \forall \quad t, s$$
$$E(\lambda u_t - \lambda \mu_u)^2 = E(\lambda u_s - \lambda \mu_u)^2 = \lambda^2 \sigma_u^2 \quad \forall \quad t, s$$
$$E[\lambda u_t \lambda u_{t-s}] = \lambda^2 \sigma_s \quad \forall \quad t, s$$

Thus if $\mathbf{b}$ is a cointegrating relationship, so is $\lambda \mathbf{b}$

- Cointegrating relationships are unique only up to scale
- Thus we will always normalize one of the coefficients to 1

Exercise:

- Show that any linear combination of two stationary processes is also a stationary process

Cointegrating Relationships vs Common Trends

- We will show that
  - In a system $n > 1$ variables, there can be up to $(n - 1)$ cointegrating relationships
    - If there are $r \in \{1,2, \dots, n - 1\}$ cointegrating relationships in a system
    - There are $n - r$ independent common stochastic trends in the system
  - if two processes are $CI(1,1)$, they share a common stochastic trend

Number of cointegrating relationships: 2 variables

Proof by contradiction:

Suppose $\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \sim CI(1,1)$ and there are two distinct cointegrating relationships: $\mathbf{b}_1 = \begin{bmatrix} 1 & \beta_1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 1 & \beta_2 \end{bmatrix}$.

Since they are distinct, we must have $\beta_1 \neq \beta_2$.

Since $\mathbf{b}_1$ and $\mathbf{b}_2$ are cointegrating relationships, both $\mathbf{b}_1\mathbf{y}_t \sim I(0)$ and $\mathbf{b}_2\mathbf{y}_t \sim I(0)$.

Recall that any linear combination of two stationary processes is also a stationary process.

Consider the following linear combination:

$$\left(\frac{\beta_2}{\beta_2 - \beta_1}\right)\mathbf{b}_1\mathbf{y}_t - \frac{\beta_1}{\beta_2}\left(\frac{\beta_2}{\beta_2 - \beta_1}\right)\mathbf{b}_2\mathbf{y}_t \quad \sim \quad I(0)$$

Number of cointegrating relationships: 2 variables

Proof by contradiction:

Suppose $\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \sim CI(1,1)$ and there are two distinct cointegrating relationships: $\mathbf{b}_1 = \begin{bmatrix} 1 & \beta_1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 1 & \beta_2 \end{bmatrix}$.

Since they are distinct, we must have $\beta_1 \neq \beta_2$.

Since $\mathbf{b}_1$ and $\mathbf{b}_2$ are cointegrating relationships, both $\mathbf{b}_1\mathbf{y}_t \sim I(0)$ and $\mathbf{b}_2\mathbf{y}_t \sim I(0)$.

Recall that any linear combination of two stationary processes is also a stationary process.

Consider the following linear combination:

$$\left(\frac{\beta_2}{\beta_2 - \beta_1}\right)\mathbf{b}_1\mathbf{y}_t - \frac{\beta_1}{\beta_2}\left(\frac{\beta_2}{\beta_2 - \beta_1}\right)\mathbf{b}_2\mathbf{y}_t = y_{1t} \sim I(0)$$

This contradicts the supposition that $y_{1t} \sim I(1)$. We conclude that there could not have been two distinct cointegrating relationships.

n-variable case

- We will show that there can be at most $n-1$ cointegrating relationships
- First: any linear combination of cointegrating relationships is also a cointegrating relationship
- Let the n variables $\mathbf{y}_t$ be individually $I(1)$, and let $\mathbf{b}_1$ and $\mathbf{b}_2$ be two distinct (i.e. linearly independent) cointegrating relationships. Then, for any $\alpha \in [0,1]$:

$$\alpha \mathbf{b}_1 \mathbf{y}_t + (1-\alpha)\mathbf{b}_2 \mathbf{y}_t \sim I(0)$$

$$[\alpha \mathbf{b}_1 + (1-\alpha)\mathbf{b}_2]\mathbf{y}_t \sim I(0)$$

- Thus $[\alpha \mathbf{b}_1 + (1-\alpha)\mathbf{b}_2]$ is also a cointegrating relationship

n-variable case

- Let the n variables $\mathbf{y}_t$ be individually I(1), and let $\mathbf{B}$ be an [n x n] matrix containing all the cointegrating relationships
- We are interested in the number of distinct (i.e. linearly independent) cointegrating relationships in $\mathbf{B}$
  - I.e. the number of independent rows of $\mathbf{B}$,
  - Or equivalently, the rank of $\mathbf{B}$
- Cases:
  - $rank(\mathbf{B}) = 0$
  - $rank(\mathbf{B}) = n$
  - $rank(\mathbf{B}) = r$ with $0 < r < n$

n-variable case

- Cases:
  - $rank(\mathbf{B}) = 0 \Rightarrow \mathbf{B} = \mathbf{0}$
    - no linear relationship is stationary
    - The only way to make $\mathbf{y}_t$ stationary by linear operation is to multiply by a matrix of zeros

n-variable case

- Cases:
  - $rank(\mathbf{B}) = n$
    - $\mathbf{B}$ has n non-zero eigenvalues and $\mathbf{B}^{-1}$ exists,
    - If $\mathbf{B}\mathbf{y}_t$ is stationary, so is any linear combination, so $\mathbf{B}^{-1}\mathbf{B}\mathbf{y}_t$ is I(0)
    - So $\mathbf{I}\mathbf{y}_t$ is stationary, i.e. each component of $\mathbf{y}_t$ is stationary, which contradicts our starting point
    - Hence there cannot be n linearly independent cointegrating relationships
    - Intuition from Linear Algebra:

$\mathbf{B}$ forms a *basis* of $\mathbf{R}^n$ - any $x \in \mathbf{R}^n$ is a linear combination of rows of $\mathbf{B}$

n-variable case

- We conclude:
- If the n variables $\mathbf{y}_t$ are CI(1,1) and $\mathbf{B}$ is the [n x n] matrix containing all the cointegrating relationships:
  - $rank(\mathbf{B}) = r$ with $0 < r < n$
  - $\mathbf{B}$ has only r non-zero eigenvalues
  - There are r cointegrating relationships with $r \in \{1,2,\dots,n-1\}$

Common Trends Representation in 2 variables

- Let $\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \sim CI(1,1)$ with cointegrating relationship: $\mathbf{b} = \begin{bmatrix} 1 & \beta \end{bmatrix}$

- Since $y_{2t} \sim I(1)$ we know we can represent it as:

$$\begin{aligned} y_{2t} &= \mu_t + e_{2t} \\ \text{with} & \\ \mu_t &\sim I(1) \\ e_{2t} &\sim I(0) \text{ but not necessarily white noise} \end{aligned}$$

- By assumption $u_t = y_{1t} + \beta y_{2t}$ is stationary

- Now we substitute and rearrange:

Common Trends Representation in 2 variables

- Now we substitute and rearrange:

$$
\begin{aligned}
u_t &= y_{1t} + \beta y_{2t} \\
y_{1t} &= -\beta y_{2t} + u_t \\
&= -\beta\left(\mu_t + e_{2t}\right) + u_t \\
&= -\beta \mu_t - \beta e_{2t} + u_t
\end{aligned}
$$

- Thus two variables are cointegrated if (and only if) their stochastic trends are identical up to scaling
- Exercises:
  - Show the “only if” direction of the proof: if two variables share the same trend up to scaling, they are cointegrated
  - Show that if two I(1) processes have independent stochastic trends, they can be cointegrated with a third I(1) under a specific linear relationship

Common Trends Representation

- In N variables, just the result:
  - If there are $0<r<n$ cointegrating relationship
  - There are n-r independent stochastic trends

The Vector Error Correction Mechanism (VECM) Representation

- Start from the n-variable case this time
- Consider a VAR(2)
  - Enders shows the same for a VAR(p)

The Vector Error Correction Mechanism (VECM) Representation

Let $\underset{[n\times 1]}{\mathbf{y}_t} \sim CI\left(1,1\right)$

Consider the standard reduced form VAR(2) (abstracting from constants):

$$\mathbf{y}_t = A_1\mathbf{y}_{t-1} + A_2\mathbf{y}_{t-2} + e_t$$

Note:

- If the variables are cointegrated, the residuals from this regression must be stationary – the cointegrating relationship will be captured “somewhere” in the linear coefficients above
- Put differently: if the residuals are not stationary, there is no cointegrating relationship between the n variables
  - What about subsets?

The Vector Error Correction Mechanism (VECM) Representation

$$\mathbf{y}_t = A_1\mathbf{y}_{t-1} + A_2\mathbf{y}_{t-2} + \boldsymbol{e}_t$$

Add and subtract $A_2\mathbf{y}_{t-1}$ from the RHS:

$$\begin{aligned} \mathbf{y}_t &= A_1\mathbf{y}_{t-1} + A_2\mathbf{y}_{t-2} - A_2\mathbf{y}_{t-1} + A_2\mathbf{y}_{t-1} + \boldsymbol{e}_t \\ &= (A_1 + A_2)\,\mathbf{y}_{t-1} - A_2\triangle\mathbf{y}_{t-1} + \boldsymbol{e}_t \end{aligned}$$

Subtract $\mathbf{y}_{t-1}$ from both sides:

$$\begin{aligned} \mathbf{y}_t - \mathbf{y}_{t-1} &= (A_1 + A_2)\,\mathbf{y}_{t-1} - \mathbf{y}_{t-1} - A_2\triangle\mathbf{y}_{t-1} + \boldsymbol{e}_t \\ \triangle\mathbf{y}_t &= -(I - A_1 - A_2)\,\mathbf{y}_{t-1} - A_2\triangle\mathbf{y}_{t-1} + \boldsymbol{e}_t \\ &= \Pi\mathbf{y}_{t-1} + C_1\triangle\mathbf{y}_{t-1} + \boldsymbol{e}_t \end{aligned}$$

Vector Error Correction Mechanism (VECM) Representation

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + C_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Thus Π contains the cointegrating relationships/vectors

Recall: $rank(\Pi) = r < n$

- We will use the estimated eigenvalues of Π to construct an empirical test of the rank of the matrix (i.e. the number of cointegrating relationships.
- Additionally, the rank condition implies the following decomposition exists

$$\underset{[n \times n]}{\Pi} = \underset{[n \times r]}{\boldsymbol{\alpha}} \; \underset{[r \times n]}{\beta}$$

- Where $\boldsymbol{\beta}$ contains the cointegrating vectors in its r independent rows
- And $\beta \mathbf{y}_t = \underset{[r\times 1]}{\mathbf{u}_t}$ are the "equilibrium errors" or "deviations from the long run equilibrium" (what is "long run"?)

Vector Error Correction Mechanism (VECM) Representation

$$\triangle \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + C_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Additionally, this representation gives another result on appropriate time series modelling:

- *Unless,* both theoretically and empirically $rank(\Pi) = 0$, estimating a VAR in differences when the levels are integrated is a misspecification, and will yield inconsistent estimates.

Vector Error Correction Mechanism (VECM) Representation

$$\begin{aligned}\triangle \mathbf{y}_t &= \Pi \mathbf{y}_{t-1} + C_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{e}_t \\ &= \boldsymbol{\alpha}\left(\beta \mathbf{y}_{t-1}\right) + C_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{e}_t \\ &= \boldsymbol{\alpha}\left(\mathbf{u}_{t-1}\right) + C_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{e}_t\end{aligned}$$

The matrix $\boldsymbol{\alpha}$ contains the *loading factors* on the past long run error

- E.g. the entry α_{ij} measures the size of the response of Δy_{it} to the deviation from long run equilibrium relationship j in period $t-1$
- Hence: how much of the change in y_{it} is a *correction* of the j^{th} long run error from the previous period
- For the process to be cointegrated, one or more of the variables *must* adjust to deviation from any long run error, otherwise the process will drift arbitrarily far from the proposed long run relationship

Vector Error Correction Mechanism (VECM) Representation

$$\begin{bmatrix} \triangle y_{1t} \\ \triangle y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} (y_{1t-1} + \beta y_{2t-1}) + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$
$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} u_{t-1} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

- In the two variable system, or when there is only one cointegrating relationship, the $\boldsymbol{\alpha}$ vector is called the *speed of adjustment* coefficients
  - Some always call them that, but it doesn't quite work a neatly in higher order systems.

- Some texts also argue that one must consider the values of these coefficients to ensure stability:
  - Suppose $\beta < 0$, then $u_{t-1} > 0$ implies that either y_{1t-1} was too large, or y_{2t-1} was too small, so either y_{1t} must decrease, i.e. $\alpha_1 < 0$ or y_{2t} must increase, i.e. $\alpha_2 > 0$ or both, otherwise the errors are not corrected
  - In practice, if there is strong empirical cointegration, this holds naturally – since the system *tries* to find stationary errors by construction
  - If the errors from the VECM are stationary, the cointegration relationships are necessary stable

Vector Error Correction Mechanism (VECM) Representation

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + C_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Note that, like in the two variable case, the decomposition is not unique:

$$\underset{[n \times n]}{\Pi} = \underset{[n \times r]}{\boldsymbol{\alpha}} \; \underset{[r \times n]}{\boldsymbol{\beta}}$$

- Consider any $[r \times r]$ matrix H with $rank(H) = r$ then

$$\Pi = \alpha\beta = (\alpha H)(H^{-1}\beta) = \alpha^* \beta^*$$

- β^* is still a valid set of cointegrating relationships, since any linear combination of stationary processes is also a stationary process

- Thus, we usually impose a **normalization** on β

Vector Error Correction Mechanism (VECM) Representation

Typical normalization of β:

- Consider an arbitrary set of cointegrating relationships in a three variable CI(1,1) system:

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix}$$

- If the first [2x2] entries $H = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$ are full rank, then

$$\beta^* = H^{-1}\beta = \begin{bmatrix} 1 & 0 & \beta_{13}^* \\ 0 & 1 & \beta_{23}^* \end{bmatrix}$$

- Is also a valid set of cointegrating vectors,

- Note that this is a normalization, ***not a restriction*** – the first true restriction we impose on the cointegrating space is an *additional* one over and above this normalization

- Since the order of the variables are arbitrary at this point, this normalization can make it difficult to a link the estimated cointegrating relationships to theoretical economic equilibria of interest.

Weak Exogeneity and ARDL models

$$\begin{bmatrix} \triangle y_{1t} \\ \triangle y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} u_{t-1} + \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- If $\alpha_2 = 0$ all the “adjusting to the long run error” occurs through the dynamics of y_{1t}
- If, in addition, there are no contemporaneous effects from y_{1t} to y_{2t}, i.e. when $f_{21} = 0$, we call y_{2t} **weakly exogenous** to the system
- Note that the first condition is testable, whereas the second is not

- When one is confident that *all* variables, other than one of interest, are weakly exogenous, one can validly estimate the first equation as a cointegrated ARDL model (i.e. a single equation model)
  - Then one validly can test for cointegration in a single equation, even with a mixture of I(0) and I(1) variables via the Pesaran, Shin and Smith (2001) **bounds test**
  - Be careful – most applications just do it without considering the necessary conditions for the method to work.

Plan

- Weak exogeneity in a cointegrated system
- Estimation Approaches:
  – Engle Granger approach (sequential, single-equation, OLS based)
  – Johansen Method (simultaneous, VAR, ML based)

Weak Exogeneity and ARDL models

$$\begin{bmatrix} \triangle y_{1t} \\ \triangle y_{2t} \end{bmatrix} = \begin{bmatrix} a_1 \\ 0 \end{bmatrix} u_{t-1} + \begin{bmatrix} f_{11} & f_{12} \\ 0 & f_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Or:

$$\begin{aligned} \triangle y_{1t} &= a_1 \left(y_{1t-1} + \beta y_{2t-1} \right) + \left[f_{11}\varepsilon_{1t} + f_{12}\varepsilon_{2t} \right] \\ \triangle y_{2t} &= f_{22}\varepsilon_{2t} \end{aligned}$$

Then noting: $\varepsilon_{2t} = \frac{\triangle y_{2t}}{f_{22}}$ and defining $c = \frac{f_{12}}{f_{22}}$ we obtain:

$$\triangle y_{1t} = a_1 \left(y_{1t-1} + \beta y_{2t-1} \right) + c\triangle y_{2t} + f_{11}\varepsilon_{1t}$$

Which can be generalized to:

$$\triangle y_{1t} = \beta_1 y_{1t-1} + \beta_2 y_{2t-1} + C_1 (L) \triangle y_{1t-1} + C_2 (L) \triangle y_{2t} + e_{1t}$$

- Again, for this to be a valid approach y_{2t} **must** be weakly exogenous: $\mathrm{E}(y_{2t} e_{1t}) = 0$, which requires both $\alpha_2 = 0$ and $f_{21} = 0$

The Engle and Granger (1987) approach

- The first practical approach to estimating a cointegration relationship
  - Eventually got them a Nobel, although it probably should have been shared with David Hendry…
- We will just briefly consider this approach as the Johansen (1988,1992) method is a more general, simultaneous ML approach that is currently the most used
- The key problems with the EG method:
  - One has to pick the “LHS” variable explicitly, and the results may depend on this choice
  - It is a two step process – any error in the first stage carries over and influences the second step

The Engle and Granger (1987) approach

1. Pre-test variables for order of integration.
   - If all I(1), proceed. (Enders also gives clear coverage of multi-cointegration
2. Estimate the long run relationship

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + e_t$$

   - Remember if there is cointegration, OLS estimates are *super consistent,* they converge faster than I(0) regressions.
   - Test whether $\hat{e}_t$ series is I(0), if yes – conclude cointegration
     - Note, this is *not* just a standard ADF test, as the estimation above attempts to minimize squared residuals – thus biased towards stationarity
     - Engle and Granger constructed new tabulations for this test
3. Estimate the error correction model

$$\Delta \mathbf{y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\alpha} \hat{e}_{t-1} + A(L)\Delta \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

   - This is just a VAR in first differences with an additional term: the lagged deviation from the long run relationship estimated in step 2
4. Assess Model Adequacy

The Johansen (1988, 1992) method

- Johansen developed a novel, maximum likelihood based test of cointegration in a VAR setting
  - Since it is maximum likelihood based, it avoids the small sample issues by construction
  - BUT, maximum likelihood is sensitive to the assumption of the correct density function
  - Probably still subject to small sample issues

$$\triangle \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + C_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{e}_t$$

- The core of the method is a test of the number of cointegrating relationships, which is equivalent to the number of non-zero eigenvalues of the matrix Π
  - In an empirical estimation (with noise), Π will be algebraically full rank – a random real-valued square matrix is almost always invertible
  - Johansen derived tests for the number of *statistically significant* eigenvalues of the Π matrix
  - These boil down to a multivariate version of the Dickey-Fuller test
  - Again – non-standard distribution tabulated via Monte Carlo methods

The Johansen (1988, 1992) method

$$\Delta \mathbf{y}_t = \Phi \mathbf{D}_t + \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{C}(L) \Delta \mathbf{y}_{t-1} + e_t$$

The steps:

0. Choose the specification of the deterministic parts: Here the $\mathbf{\Phi D}_t = \mu_t = \mu_0 + \mu_1 t$ term contains all deterministic parts, i.e. constants and trends. As in the ADF test, the critical values of the tests depend on these terms.

   – In the cointegration setting there are 5 cases:

The Johansen (1988, 1992) method

No constant: $\mu_t = 0$:
$\mathbf{y}_t$ is I(1) without drift, $\beta\mathbf{y}_t$ is mean zero (no LR constant)

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha\beta}\mathbf{y}_t + A(L)\Delta\mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Restricted constant: $\mu_t = \mu_0 = \boldsymbol{\alpha\rho}_0$
$\mathbf{y}_t$ is I(1) without drift, $\beta\mathbf{y}_t$ has non-zero means

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}\mathbf{y}_t + \boldsymbol{\rho}_0) + A(L)\Delta\mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Unrestricted constant: $\mu_t = \mu_0$
$\mathbf{y}_t$ is I(1) with drift vector μ_0, $\beta\mathbf{y}_t$ may have non-zero means

$$\Delta\mathbf{y}_t = \mu_0 + \boldsymbol{\alpha\beta}\mathbf{y}_t + A(L)\Delta\mathbf{y}_{t-1} + \boldsymbol{e}_t$$

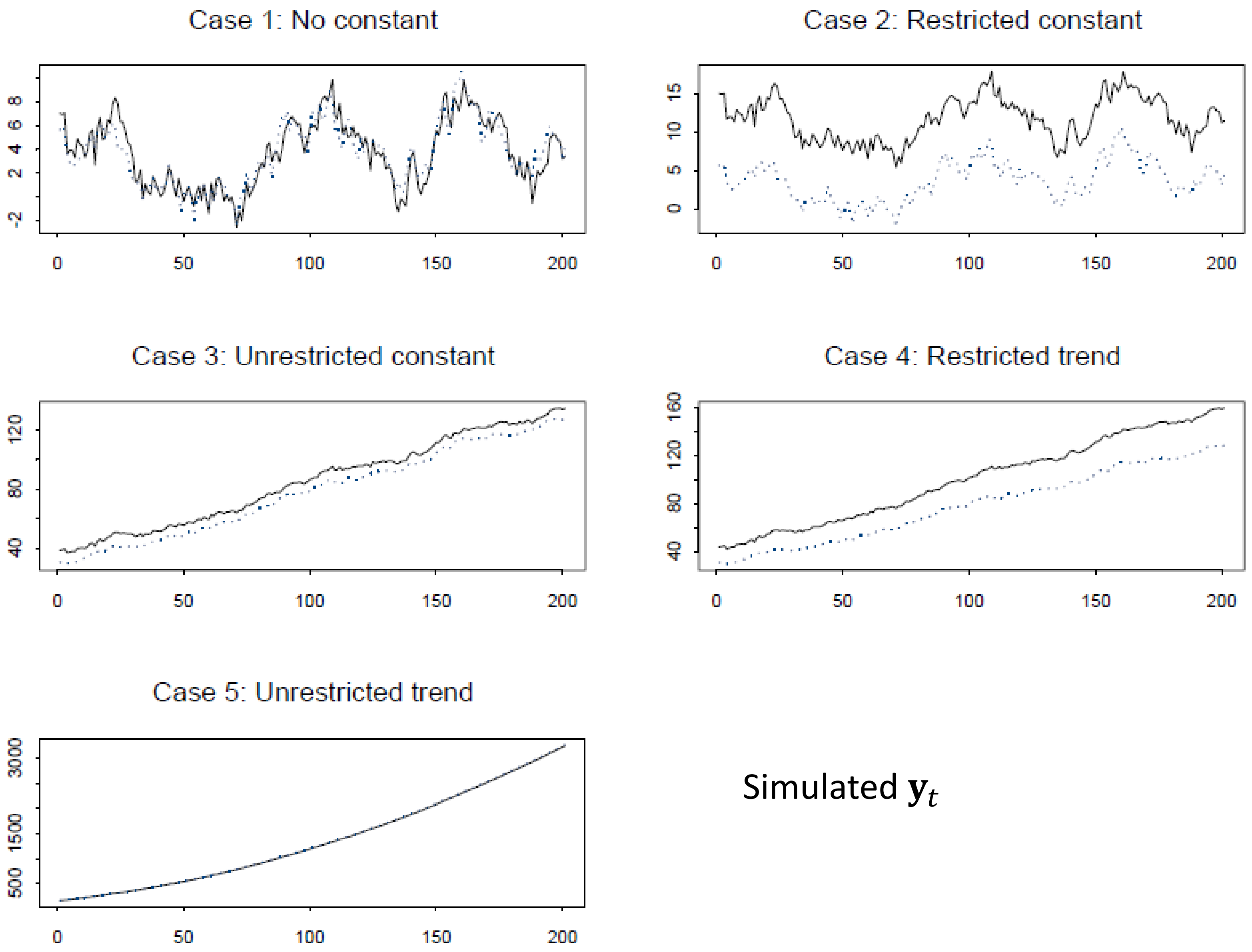
Restricted trend: $\mu_t = \mu_0 + \boldsymbol{\alpha\rho}_1 t$
$\mathbf{y}_t$ is I(1) with drift vector μ_0, $\beta\mathbf{y}_t$ has linear trend

$$\Delta\mathbf{y}_t = \mu_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}\mathbf{y}_t + \boldsymbol{\rho}_1 t) + A(L)\Delta\mathbf{y}_{t-1} + \boldsymbol{e}_t$$

Unrestricted trend: $\mu_t = \mu_0 + \mu_1 t$
$\mathbf{y}_t$ is I(1) with linear trend and drift vector μ_0 - quadratic trend in levels,
$\beta\mathbf{y}_t$ has linear trend

$$\Delta\mathbf{y}_t = \mu_0 + \mu_1 t + \boldsymbol{\alpha\beta}\mathbf{y}_t + A(L)\Delta\mathbf{y}_{t-1} + \boldsymbol{e}_t$$

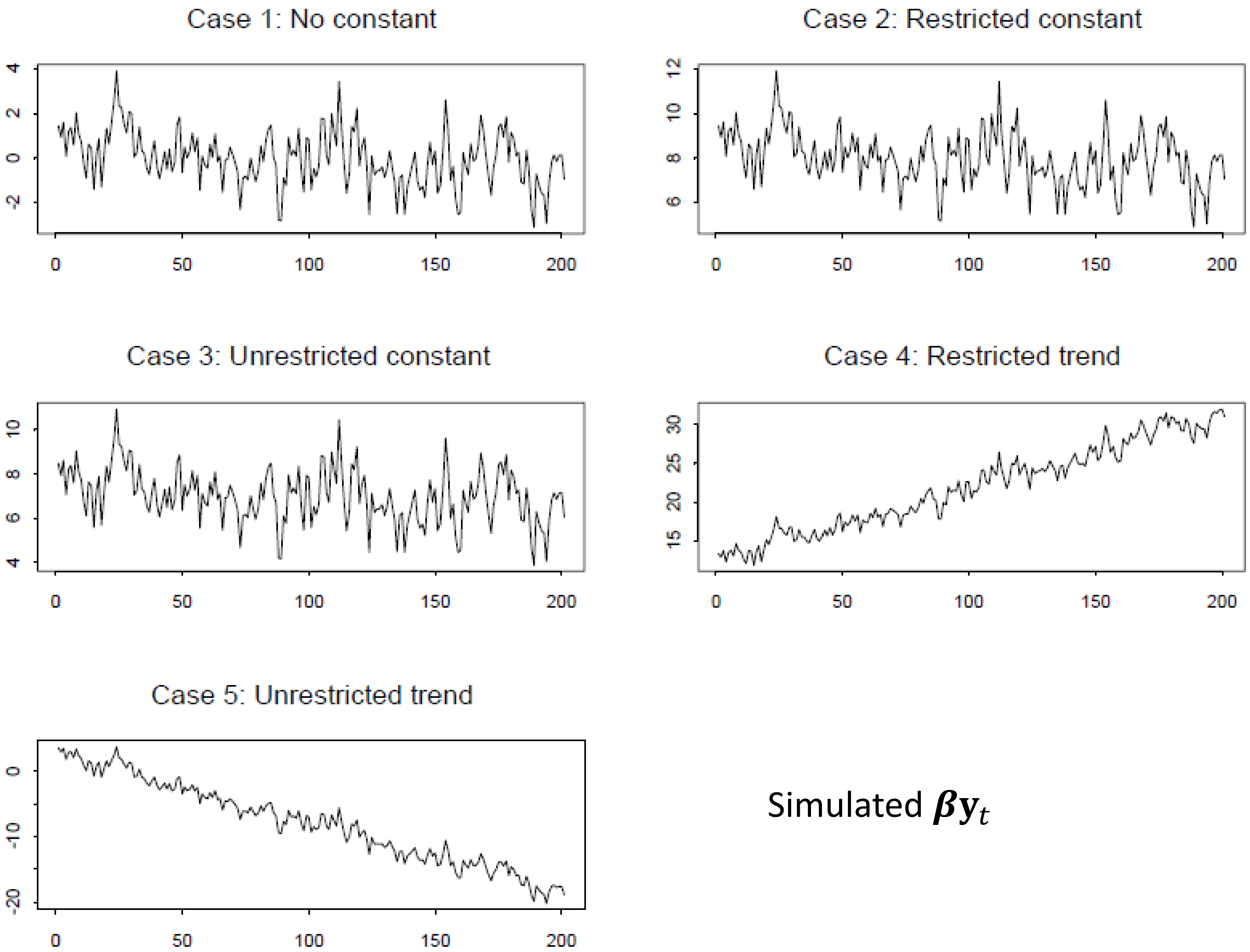
The Johansen (1988, 1992) method



Simulated $\mathbf{y}_t$

Source: https://faculty.washington.edu/ezivot/econ584/notes/cointegration.pdf

The Johansen (1988, 1992) method



Simulated $\boldsymbol{\beta y}_t$

Source: https://faculty.washington.edu/ezivot/econ584/notes/cointegration.pdf

The Johansen (1988, 1992) method

The steps:

1. Pre-test the variables to conclude that they are (or may be) I(1)
   – Note that there are also tests for individual unit roots *within* the system that have different properties

The Johansen (1988, 1992) method

The steps:

2. Estimate an unrestricted VAR in levels and check the adequacy of specification of the model

$$\mathbf{y}_t = \Phi \mathbf{D}_t + \mathbf{A}(L)\mathbf{y}_{t-1} + e_t$$

- If there is cointegration, some linear combination(s) of the endogenous variables are stationary, and this will be captured in the coefficient matrices
- A critical point is to test that errors are **white noise**
  - The Johansen Method is a ML method, i.e. limiting distributions are derived assuming **normal** errors
  - Robust to some deviations: some non-normality, heteroscedasticity
  - **Requires:** i.i.d. errors with finite variance
  - **Hence the following is unacceptable:** autocorrelated residuals, time-varying parameters, structural breaks

The Johansen (1988, 1992) method

The steps:

3. Determine Cointegration Rank (number of cointegrating relationships)

   Perform Likelihood based tests on the eigenvalues of the Π matrix

$$\triangle \mathbf{y}_t \quad = \quad \Phi \mathbf{D}_t + \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{C}(L)\triangle \mathbf{y}_{t-1} + e_t$$

- Johansen shows how a consistent estimate of Π can be recovered from the covariance matrices of two regressions: the level of Y on its lagged differences, and the difference of Y on its lagged differences
- Find the estimated eigenvalues and order them from large to small:

$$\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$$

- There are two tests:
  - The trace test
  - The maximum eigenvalue test
- Johansen shows that the following iterative scheme is consistent:
  Start from the hypothesis that there are zero significant eigenvalues, if rejected, there must be at least one. Then test for 1 against the alternative of more, and so on.

The Johansen (1988, 1992) method

The steps:

3.

- The trace test

$$H_0(r): r = r_0 \text{ vs. } H_1(r_0): r > r_0$$

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$$

  Intuition: if there are r_0 non-zero eigenvalues, the sum of the rest should be small

- The maximum eigenvalue test

$$H_0(r_0): r = r_0 \text{ vs. } H_1(r_0): r_0 = r_0 + 1$$

$$LR_{\max}(r_0) = -T \ln(1 - \hat{\lambda}_{r_0+1})$$

  Intuition: if there are r_0 non-zero eigenvalues, the eigenvalue $r_0 + 1$ should be small

The Johansen (1988, 1992) method

The steps:

4. Impose the number of cointegrating relationships r and perform a reduced rank regression of:

$$\triangle \mathbf{y}_t \quad = \quad \Phi \mathbf{D}_t + \mathbf{\Pi} \mathbf{y}_{t-1} + \mathbf{C}(L) \triangle \mathbf{y}_{t-1} + e_t$$

   - Again, Johansen shows how to obtain consistent ML estimates.
   - Note that *now* we are imposing restrictions on the Π matrix, so the estimates will differ from the unrestricted VAR/VECM we started with
     - Again evaluate the model adequacy

The Johansen (1988, 1992) method

The steps:

5. Normalize the cointegrating relationships, evaluate/interpret the economic content/forecasts of the model. Test further hypothesized restrictions
   - As in previous models, restrictions are tested analogously to an F test:
     - Estimate the restricted model, Compare the reduction in fit. If the restriction is valid, the reduction in fit should be small
     - Again – the critical values are found by simulation
   - A key question is usually whether a variable is weakly exogenous (in terms of the loading factor)
   - Impulse Response Functions have been derived
     - But now there are two types of shock: permanent and temporary