# Chapter Five

# Cointegration Analysis

Tingguo Zheng (zhengtg@gmail.com)
Department of Statistics and Data Science, School of Economics, Xiamen University

(For Undergraduate Program, 2025-2026, Spring Semester)

- *Reference: “Analysis of Financial Time Series (3ed)” by Ruey S. Tsay, 2010. (Chapter 8)*

- **Outline:**
  - Cointegration
  - Cointegrated VAR Models
  - Discussion of a Simple Model
  - Specification, Estimation and Cointegration Tests

# 1. Cointegration

## 1.1 Spurious Regression

- If some or all of the variables in a regression are I(1) then the usual statistical results may or may not hold.

- One important case is *spurious regression* when all the regressors are I(1) and not cointegrated.

Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of I(1) time series that are not cointegrated.

Write $\boldsymbol{y}_t = (y_{1t}, \boldsymbol{y}'_{2t})'$, and consider regressing of $y_{1t}$ on $\boldsymbol{y}_{2t}$ giving

$$y_{1t} = \widehat{\boldsymbol{\beta}}'_2 \boldsymbol{y}_{2t} + \hat{u}_t.$$

Since $y_{1t}$ is not cointegrated with $\boldsymbol{y}_{2t}$,

- True value of $\boldsymbol{\beta}_2$ is zero;

- The above is a spurious regression and $\hat{u}_t \sim I(1)$.

**Example 1** (Spurious regression). Consider regressing $y_t$ on the completely unrelated series $x_t$ where

$$y_t = \alpha_y + y_{t-1} + \epsilon_t^y, \quad x_t = \alpha_x + x_{t-1} + \epsilon_t^x.$$

The OLS estimator is

$$\widehat{\beta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} = \frac{\sum_{t=1}^T (\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0)(\alpha_y t + \sum_{k=1}^t \epsilon_k^y + y_0)}{\sum_{t=1}^T (\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0)^2}$$

As $T$ gets large, the terms in $t^2$ will dominate all other terms. Re-writing this as

$$\widehat{\beta} = \frac{\frac{1}{T^3}\sum_{t=1}^T (\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0)(\alpha_y t + \sum_{k=1}^t \epsilon_k^y + y_0)}{\frac{1}{T^3}\sum_{t=1}^T (\alpha_x t + \sum_{k=1}^t \epsilon_k^x + x_0)^2},$$

then all of the terms that are not of the form $\frac{1}{T^3}\sum_{t=1}^T t^2$ will go to zero.

This means that as $T$ gets large

$$\widehat{\beta} \xrightarrow{p} \frac{\alpha_x \alpha_y}{\alpha_x^2} = \frac{\alpha_y}{\alpha_x}$$

Therefore,

- The OLS estimator will tend towards the ratio of the two drift terms.
- In addition, the $t$ statistics,

$$t\text{-ratio} = \frac{\widehat{\beta}}{\widehat{\mathrm{std}}(\beta)} = \widehat{\beta}\sigma^{-1}(\sum_{t=1}^{T} x_t^2)^{1/2} \to \pm\infty$$

as $T \to \infty$, will generally indicate that there is a highly statistically significant relationship.

The following results about the behavior of $\widehat{\boldsymbol{\beta}}_2$ in the spurious regression are due to Phillips (1986):

- $\widehat{\boldsymbol{\beta}}_2$ does *not converge in probability to* zero but instead *converges in distribution to* a non-normal random variable not necessarily centered at zero. This is the *spurious regression phenomenon*.

- The usual OLS $t$-statistics for testing that the elements of $\boldsymbol{\beta}_2$ are zero *diverge to* $\pm\infty$ *as* $T \to \infty$. Hence, with a large enough sample it will appear that $\boldsymbol{y}_t$ is cointegrated when it is not if the usual asymptotic normal inference is used.

- The usual $R^2$ from the regression *converges to unity* as $T \to \infty$ so that the model will appear to fit well even though it is misspecified.

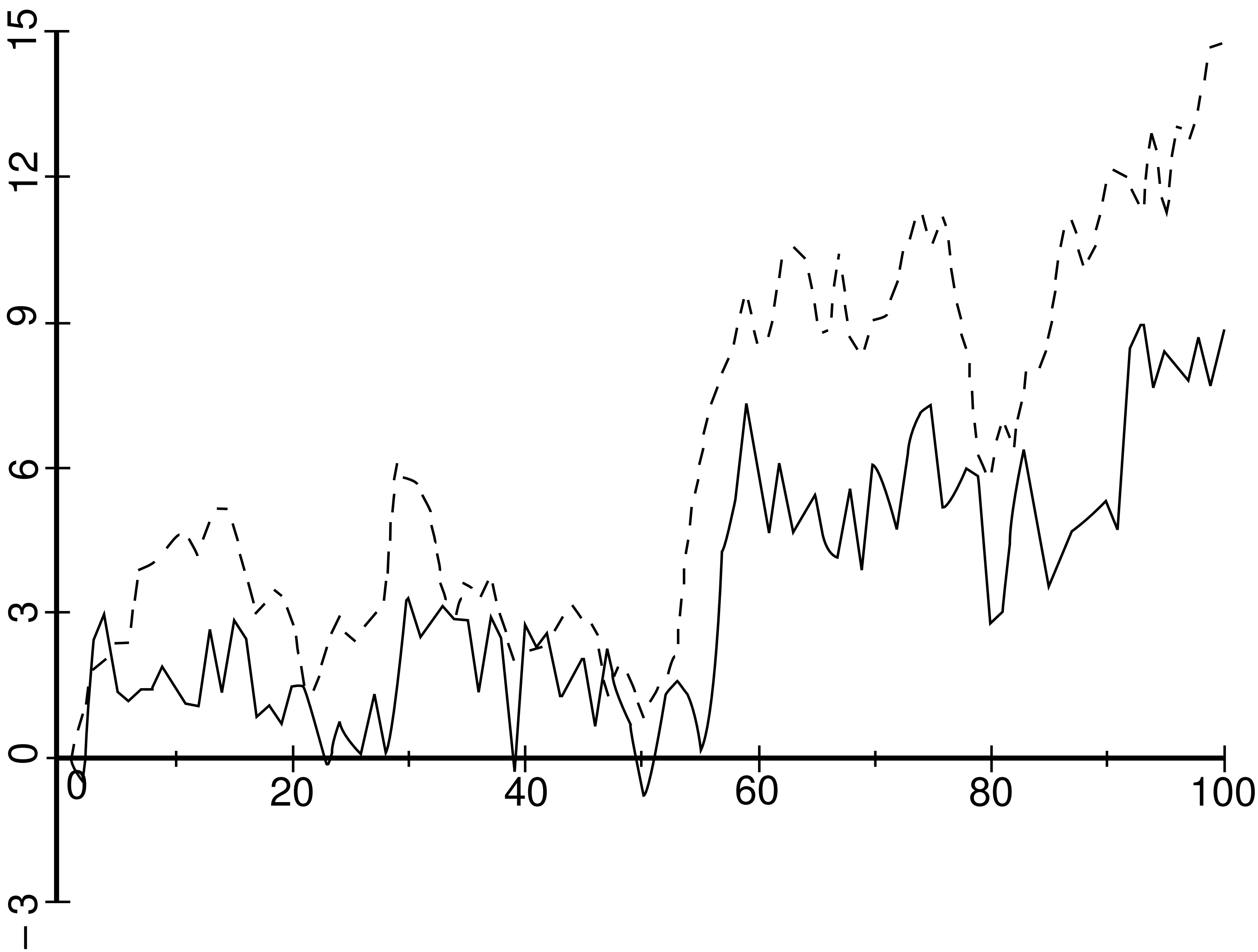
## 1.2 Cointegration

**Definition 1.** *If there exists a linear combination of the series that is of integration order lower than that of the series with higher integration order, the two series are cointegrated.*

In the bivariate example, if both $x_t$ and $y_t$ are I(1), but there exists a $\beta$ such that $y_t - \beta x_t$ is I(0), then $x_t$ and $y_t$ are cointegrated, and $(1, -\beta)$ is the *cointegrating vector*.

**Definition 2** (Cointegration)**.** *More generally, the elements of the $n$-dimensional vector $\boldsymbol{y}_t$ are cointegrated of order $(d, b)$, denoted with $CI(d, b)$, if*

- *All variables in $\boldsymbol{y}_t$ are $I(d)$;*
- *There exists a vector $\boldsymbol{\beta} \neq 0$ such that $\boldsymbol{\beta}'\boldsymbol{y}_t$ is $I(d-b)$, with $b > 0$.*


15
12
9
6
3
0
−3
0
20
40
60
80
100

Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of $I(1)$ time series. $\boldsymbol{y}_t$ is cointegrated if there exists an $(n \times 1)$ vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ such that

$$\boldsymbol{\beta}'\boldsymbol{y}_t = \beta_1 y_{1t} + \cdots + \beta_n y_{nt} \sim I(0).$$

In words, the nonstationary $I(1)$ time series in $\boldsymbol{y}_t$ are cointegrated if there is a linear combination of them that is stationary or $I(0)$.

- The linear combination $\boldsymbol{\beta}'\boldsymbol{y}_t$ is often motivated by economic theory and referred to as a *long-run equilibrium relationship*.

- Intuition: $I(1)$ time series with a long-run equilibrium relationship cannot drift too far apart from the equilibrium because *economic forces* will act to restore the equilibrium relationship.
  *That is, $w_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ is mean-reverting*.

## 1.3 Superconsistency

In the case of cointegration, it turns out that OLS estimates of $\beta$ are *consistent*. Moreover, this property is known as *superconsistency*.

**Example 2.** Consider again the case where $x_t$ is a unit root with drift

$$x_t = \alpha_x + x_{t-1} + \epsilon_t^x,$$

but in this case the variable $y_t$ is cointegrated with $x_t$ so that

$$y_t = \beta x_t + u_t,$$

where $u_t$ is mean-zero I(0) series.

We can calculate the properties of the OLS estimator as follows:

$$\widehat{\beta} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0) u_t}{\sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0)^2} \tag{1}$$

In this case, the terms in $t^2$ will dominate as $T \rightarrow \infty$ so that the denominator of the last term will grow faster than the numerator. This means that $\widehat{\beta} \xrightarrow{p} \beta$.

Now multiply both sides of (1) by $T$ but do this to the right hand side by dividing the numerator by $T^2$ and the denominator by $T^3$:

$$T(\widehat{\beta} - \beta) = \frac{\frac{1}{T^2} \sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0) u_t}{\frac{1}{T^3} \sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0)^2} \xrightarrow{p} 0.$$

The numerator converges to zero (the sum $\frac{1}{T^2} \sum_{t=1}^{T} \alpha_x t \rightarrow \frac{\alpha_x}{2}$ but is multiplied by an uncorrelated mean zero series $u_t$) while the denominator converges to $\frac{\alpha_x^2}{3}$.

We have seen cases before where $\widehat{\beta} - \beta$ converges in distribution to a mean zero series when multiplied by $\sqrt{T}$. In this case, the gap between the estimator and the true value converges in probability to zero even when multiplied by $T$. This property is known as superconsistency.

## 1.4 Normalization

The cointegration vector $\boldsymbol{\beta}$ is not unique since for any scalar $c$

$$c \cdot \boldsymbol{\beta}' \boldsymbol{y}_t = \boldsymbol{\beta}^{*\prime} \boldsymbol{y}_t \sim I(0).$$

Hence, some *normalization assumption* is required to *uniquely* identify $\boldsymbol{\beta}$. A typical normalization is

$$\boldsymbol{\beta} = (1, -\beta_2, \ldots, -\beta_n)'$$

so that

$$\boldsymbol{\beta}' \boldsymbol{y}_t = y_{1t} - \beta_2 y_{2t} - \cdots - \beta_n y_{nt} \sim I(0)$$

or

$y_{1t} = \beta_2 y_{2t} + \ldots + \beta_n y_{nt} + u_t, \quad u_t \sim I(0)$ is the *cointegrating residual*.

## 1.5 Error correction model (ECM)

**Theorem 1** (*Granger representation theorem* (Granger, 1983, Engle and Granger, 1987))**.** *If the variables in $Y$ are cointegrated, then they can be represented as an ECM: And, vice versa, if $Y$ has an ECM representation, then the variables in $Y$ are cointegrated.*

In a simple bivariate case, consider the ADL(1,1) (*autoregressive distributed lag*) model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

and its ECM representation

$$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + (\alpha_1 - 1)\left(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1}\right) + u_t,$$

where the short-run and long-run features are modelled separately.

Note also that the ECM can be rewritten as

$$\begin{aligned}\Delta y_t &= \beta_0 \Delta x_t + (\alpha_1 - 1)\left(y_{t-1} - \frac{\alpha_0}{1-\alpha_1} - \frac{\beta_0+\beta_1}{1-\alpha_1} x_{t-1}\right) \\ &= \beta_0 \Delta x_t + (\alpha_1 - 1)(y_{t-1} - y^*_{t-1}) + u_t,\end{aligned}$$

expressing the change in $y$ as a function of the change in $x$ plus an error-correction term, where $\beta_0$ describes the short-run relationship and $(\alpha_1 - 1)$ the speed of adjustment to the equilibrium.

It should be noted that the results of above ADL(1,1) model can be extended to a general ADL($p, q$) model

$$\alpha(L) y_t = c + \beta(L) x_t + u_t,$$

where $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$ and $\beta(L) = \beta_0 + \beta_1 L + \cdots + \beta_q L^q$.

## 1.6 Engle-Granger methodology

In the bivariate case, testing for cointegration is relatively straightforward. For the model

$$y_t = \alpha + \beta x_t + \varepsilon_t, \tag{2}$$

test for a cointegrating relationship by testing for a unit root in the residuals, and use $\widehat{\beta}_{OLS}$ as a (super) consistent estimator of $\beta$.

ADF-type unit root tests on the residuals are often adopted, but it requires adjusting the critical values that stationarity is "induced" by OLS in the first step.

If the null of integrated errors is rejected, so that the null of no-cointegration between $y_t$ and $x_t$ is rejected, $\widehat{\beta}_{OLS}$ is a super consistent estimator of $\beta$, i.e. $T(\widehat{\beta} - \beta)$ converges to 0.

**TABLE 8.2**
**Asymptotic critical values for cointegration tests**

| $k$* | Test statistic† | 1% | 5% | 10% |
|---|---|---|---|---|
| 2 | *c* | −3.90 | −3.34 | −3.04 |
| | *ct* | −4.32 | −3.78 | −3.50 |
| 3 | *c* | −4.29 | −3.74 | −3.45 |
| | *ct* | −4.66 | −4.12 | −3.84 |
| 4 | *c* | −4.64 | −4.10 | −3.81 |
| | *ct* | −4.97 | −4.43 | −4.15 |
| 5 | *c* | −4.96 | −4.42 | −4.13 |
| | *ct* | −5.25 | −4.72 | −4.43 |
| 6 | *c* | −5.25 | −4.71 | −4.42 |
| | *ct* | −5.52 | −4.98 | −4.70 |

Reprinted by permission from Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993, 722.
*The heading $k$ indicates the number of variables in the estimated cointegrating equation.
†The letters $c$ and $ct$ indicate whether that equation contains a constant or a constant plus a linear time trend.

Figure 1 Engle-Granger critical Values.

Summarizing the Engle-Granger approach:

1. Use the *ADF tests* to verify that individual variables are I(1).

2. Estimate the long run relationship $y_t = \alpha + \beta x_t + \varepsilon_t$ by OLS.

3. Use the ADF test on the residuals from this regression to *see if they are indeed stationary*. If so, we can assume that the variables are cointegrated; if not, then we can't make that assumption.

4. If the null is rejected, use the residuals in the ECM model to estimate the remaining parameters, e.g.: $\Delta y_t = \beta_0 \Delta x_t + \gamma \widehat{\varepsilon}_{t-1} + u_t$. *(The parameters of the ECM model can be consistently estimated)*

## 1.7 Multiple Cointegrating Relationships

Now suppose that we've established that several non-stationary variables are cointegrated.

If the $(n \times 1)$ vector $\boldsymbol{y}_t$ is cointegrated there may be $0 < r < n$ linearly independent cointegrating vectors. For example, let $n = 3$ and suppose there are $r = 2$ *cointegrating vectors*

$$\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})', \quad \boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \beta_{23})'$$

Then

$$\boldsymbol{\beta}_1' \boldsymbol{y}_t = \beta_{11} y_{1t} + \beta_{12} y_{2t} + \beta_{13} y_{3t} \sim I(0)$$

$$\boldsymbol{\beta}_2' \boldsymbol{y}_t = \beta_{21} y_{1t} + \beta_{22} y_{2t} + \beta_{23} y_{3t} \sim I(0)$$

and the $(3 \times 2)$ matrix

$$\boldsymbol{\beta}' = \begin{pmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix}$$

forms a basis for the space of *cointegrating vectors*.

**Remark**:

- The linearly independent vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the cointegrating basis $\boldsymbol{\beta}$ are not unique *unless* some normalization assumptions are made.

- Any linear combination of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, e.g. $\boldsymbol{\beta}_3 = c_1\boldsymbol{\beta}_1 + c_2\boldsymbol{\beta}_2$ where $c_1$ and $c_2$ are constants, is also a cointegrating vector.

# 2. Cointegrated VAR Models

## 2.1 Vector error-correction representation

Consider a $n$-dimensional VAR($p$) time series $\boldsymbol{y}_t$ with possible *time trend*

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$. Write $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$.

- If all roots of $|\boldsymbol{\Phi}(z)| = 0$ are outside the unit circle, then $\boldsymbol{y}_t$ is *unit-root stationary*, i.e., $\boldsymbol{y}_t$ is a vector of $I(0)$ series.
- If $|\boldsymbol{\Phi}(1)| = 0$, i.e., $z = 1$ is a root, then $\boldsymbol{y}_t$ is *unit-root nonstationary*.
- For simplicity, we usually assume that all variables in $\boldsymbol{y}_t$ is an $I(1)$ process.

An error-correction model (ECM) for the VAR($p$) process $\boldsymbol{y}_t$ is

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{4}$$

where

$$\boldsymbol{\Phi}_j^* = -\sum_{i=j+1}^{p} \boldsymbol{\Phi}_i, \quad j = 1, \ldots, p-1$$

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = -\boldsymbol{\Phi}(1) = \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_{p-1} + \cdots + \boldsymbol{\Phi}_1 - \boldsymbol{I}.$$

- We refer to the term $\boldsymbol{\Pi} \boldsymbol{y}_{t-1}$ as the *error-correction term*.

- The matrices $\boldsymbol{\Phi}_i$ can be recovered from the ECM representation via

$$\boldsymbol{\Phi}_1 = \boldsymbol{I} + \boldsymbol{\Pi} + \boldsymbol{\Phi}_1^*, \quad \boldsymbol{\Phi}_i = \boldsymbol{\Phi}_i^* - \boldsymbol{\Phi}_{i-1}^*, \quad i = 2, \ldots, p,$$

where $\boldsymbol{\Phi}_p^* = \boldsymbol{0}$, the zero matrix.

## Advantages over VAR

Again

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{4}$$

- Combines levels and differences.
- Gives more intuitive explanation of data, as effects categorised into long-run and short-run.
- All long-run information confined to $\boldsymbol{\Pi}$ matrix, so can focus on that.
- $\boldsymbol{\Phi}_j^*$ matrices capture short-run dynamics of the data.

## The rank of $\boldsymbol{\Pi}$

Again

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}' = -\boldsymbol{\Phi}(1) = \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_{p-1} + \cdots + \boldsymbol{\Phi}_1 - \boldsymbol{I}.$$

If $\boldsymbol{y}_t$ contains unit roots ($z = 1$), then $|\boldsymbol{\Phi}(1)| = 0$ so that $\boldsymbol{\Pi} = -\boldsymbol{\Phi}(1)$ is singular.

Therefore, three cases are of interest in considering the ECM:

1. $\mathrm{Rank}(\boldsymbol{\Pi}) = 0$ *(Difference Stationary)*.
   This implies $\boldsymbol{\Phi}(1) = \boldsymbol{0}$ and $\boldsymbol{y}_t$ is not cointegrated. The ECM of Eq. (4) reduces to

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

   so that $\Delta \boldsymbol{y}_t$ follows a VAR($p - 1$) model with deterministic trend $\boldsymbol{\mu}_t$.

2. $\mathrm{Rank}(\mathbf{\Pi}) = n$ (full rank) *(Stationary)*.

   This implies that $|\mathbf{\Phi}(1)| \neq 0$ and $\boldsymbol{y}_t$ contains no unit roots; that is, $\boldsymbol{y}_t$ is $I(0)$. The ECM model is not informative and one studies $\boldsymbol{y}_t$ directly.

3. $0 < \mathrm{Rank}(\mathbf{\Pi}) = m < n$ *(Cointegration)*. In this case, one can write $\mathbf{\Pi}$ as

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}', \tag{5}$$

   where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times m$ matrices with $\mathrm{Rank}(\boldsymbol{\alpha}) = \mathrm{Rank}(\boldsymbol{\beta}) = m$. The ECM of Eq. (4) becomes

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \mathbf{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \mathbf{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t. \tag{6}$$

   This means that $\boldsymbol{y}_t$ is cointegrated with $m$ linearly independent cointegrating vectors, $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{y}_t \sim I(0)$, and has $n - m$ unit roots that give $n - m$ *common stochastic trends* of $\boldsymbol{y}_t$.

**Common stochastic trends** - cointegrated with $\mathrm{Rank}(\mathbf{\Pi}) = m$

- A simple way to obtain a presentation of the $n - m$ common trends is to obtain an *orthogonal complement matrix* $\boldsymbol{\alpha}_\perp$ of $\boldsymbol{\alpha}$; that is, $\boldsymbol{\alpha}_\perp$ is a $n \times (n - m)$ matrix such that $\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha} = \mathbf{0}$, and use

$$\boldsymbol{c}_t = \boldsymbol{\alpha}'_\perp \boldsymbol{y}_t.$$

- *Premultiply* the ECM by $\boldsymbol{\alpha}'_\perp$ and use $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, we see that there would be *no error-correction term* in the resulting equation, that is,

$$\boldsymbol{\alpha}'_\perp \Delta \boldsymbol{y}_t = \boldsymbol{\alpha}'_\perp \boldsymbol{\mu}_t + \boldsymbol{\alpha}'_\perp \mathbf{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\alpha}'_\perp \mathbf{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\alpha}'_\perp \boldsymbol{\varepsilon}_t.$$

- Consequently, the $(n - m)$-dimensional series $\boldsymbol{c}_t$ should have $n - m$ unit roots. (Recall that $\Delta y_t = \eta_t = \phi^{-1}(L)\varepsilon_t$)

## Normalization and constraints

The factorization $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is not unique, because for any $m \times m$ orthogonal matrix $\boldsymbol{\Omega}$ satisfying $\boldsymbol{\Omega}\boldsymbol{\Omega}' = \boldsymbol{I}$, we have

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\boldsymbol{\Omega}\boldsymbol{\Omega}'\boldsymbol{\beta}' = (\boldsymbol{\alpha}\boldsymbol{\Omega})(\boldsymbol{\beta}\boldsymbol{\Omega})' \equiv \boldsymbol{\alpha}_*\boldsymbol{\beta}_*'.$$

Additional constraints are needed to uniquely identify $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

- It is common to require that $\boldsymbol{\beta}' = [\boldsymbol{I}_m, \boldsymbol{\beta}_1']$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix and $\boldsymbol{\beta}_1$ is a $(n-m) \times m$ matrix.

- In practice, this may require reordering of the elements of $\boldsymbol{y}_t$ such that the first $m$ components all have a unit root.

- The elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ must also satisfy other constraints for the process $w_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ to be unit-root stationary.

**Example 3.** Consider the case of a bivariate VAR(1) model with one cointegrating vector. Here $n = 2$, $m = 1$, and the ECM is

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & \beta_1 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

Premultiplying the prior equation by $\boldsymbol{\beta}' = [1 \quad \beta_1]$, using $w_{t-i} = \boldsymbol{\beta}' \boldsymbol{y}_{t-i}$, and moving $w_{t-1}$ to the right-hand side of the equation, we obtain

$$\begin{aligned} \boldsymbol{\beta}'(\boldsymbol{y}_t - \boldsymbol{y}_{t-1}) = w_t - w_{t-1} &= \boldsymbol{\beta}'\boldsymbol{\mu}_t + (\alpha_1 + \alpha_2\beta_1) w_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t \\ \Longrightarrow w_t &= \boldsymbol{\beta}'\boldsymbol{\mu}_t + (1 + \alpha_1 + \alpha_2\beta_1) w_{t-1} + \eta_t, \end{aligned}$$

where $\eta_t = \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$. This implies that $w_t$ is a stationary AR(1) process. Consequently, $\alpha_i$ and $\beta_1$ must satisfy the stationarity constraint $|1 + \alpha_1 + \alpha_2\beta_1| < 1$.

# 3. Discussion of a Simple Model

## 3.1 Granger representation or MA representation

Consider a cointegrated VAR model with no lagged differences:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{7}$$

- **Stationary representation:** Multiplying $\boldsymbol{\beta}'$ from the left to model (7), we obtain

$$\begin{aligned}
\boldsymbol{\beta}'\Delta \boldsymbol{y}_t &= \boldsymbol{\beta}'\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t \\
\boldsymbol{\beta}'\boldsymbol{y}_t &= (\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t \\
\boldsymbol{\beta}'\boldsymbol{y}_t &= \sum_{i=0}^{\infty}(\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})^i\boldsymbol{\beta}'\boldsymbol{\varepsilon}_{t-i}
\end{aligned}$$

$\boldsymbol{\beta}'\boldsymbol{y}_t \sim I(0)$, if eigenvalues of $(\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})$ within unit circle.

- **Random walk representation:** Multiplying $\boldsymbol{\alpha}_\perp'$ from the left to model (7),

$$\boldsymbol{\alpha}_\perp' \Delta \boldsymbol{y}_t = \boldsymbol{\alpha}_\perp' \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\alpha}_\perp' \boldsymbol{\varepsilon}_t = \boldsymbol{\alpha}_\perp' \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\alpha}_\perp' \boldsymbol{y}_t = \boldsymbol{\alpha}_\perp' \sum_{i=0}^{t} \boldsymbol{\varepsilon}_i + \boldsymbol{\alpha}_\perp' \boldsymbol{y}_0.$$

- **Granger representation:** Combine two representations mentioned before, use

$$\boldsymbol{I} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}_\perp' + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}',$$

we have

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}_\perp' \boldsymbol{y}_t + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}' \boldsymbol{y}_t \\
&= \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp)^{-1} \left( \boldsymbol{\alpha}_\perp' \sum_{i=0}^{t} \boldsymbol{\varepsilon}_i + \boldsymbol{\alpha}_\perp' \boldsymbol{y}_0 \right) + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1} \left( \sum_{i=0}^{\infty} (\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})^i \boldsymbol{\beta}' \boldsymbol{\varepsilon}_{t-i} \right) \\
&= C \sum_{i=0}^{t} \boldsymbol{\varepsilon}_i + \sum_{i=0}^{\infty} C_i^* \boldsymbol{\varepsilon}_{t-i} + A
\end{aligned}$$

$$\boldsymbol{y}_t = C\sum_{i=0}^{t}\boldsymbol{\varepsilon}_i + \sum_{i=0}^{\infty} C_i^* \boldsymbol{\varepsilon}_{t-i} + A$$

where $C = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}'_\perp$, $C_i^* = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}(\boldsymbol{I}+\boldsymbol{\beta}'\boldsymbol{\alpha})^i\boldsymbol{\beta}'$, and $A$ is initial values.

– This formula is called as *Granger representation* (Johansen, 1995), or *moving average representation* for cointegrated VAR.

– The cointegrated VAR can be split into *random walk component* $(C\sum_{i=0}^{t}\boldsymbol{\varepsilon}_i)$, a *stationary component* $(\sum_{i=0}^{\infty} C_i^* \boldsymbol{\varepsilon}_{t-i})$, and initial values, $A$.

– So shocks to system can have both permanent effect and a transitory effect.

– Random walk parts, $\boldsymbol{\alpha}'_\perp\boldsymbol{y}_t$, are stochastic/common trends.

– Stationary parts, $\boldsymbol{\beta}'\boldsymbol{y}_t$, are cointegrating relations.

## Impulse response function

The Granger representation enables the impulse response analysis

$$\mathbf{\Psi}_h = \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{\varepsilon}_t} = C + C_h^*$$

- The impulse response function tend to $C$, i.e. $\mathbf{\Psi}_h \to C$, as the coefficient $C_i^* \to 0$ is on the stationary component.

- Hence permanent and transitory effect of impulse to the system, corresponding to the random walk and stationary parts of the process.

## 3.2 Deterministic Terms

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \underline{\boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_t} \tag{8}$$

- Granger representation:

$$\boldsymbol{y}_t = \underbrace{C \sum_{i=0}^{t} (\boldsymbol{\varepsilon}_i + \boldsymbol{\mu}_0)}_{\text{Trend}} + \underbrace{\sum_{i=0}^{\infty} C_i^* (\boldsymbol{\varepsilon}_{t-i} + \boldsymbol{\mu}_0)}_{\text{Constant}} + A$$

and further

$$\Delta \boldsymbol{y}_t = C(\boldsymbol{\varepsilon}_t + \boldsymbol{\mu}_0) + \sum_{i=0}^{\infty} C_i^* \Delta \boldsymbol{\varepsilon}_{t-i}$$

Then the equilibrium growth rate

$$E(\Delta \boldsymbol{y}_t) = C\boldsymbol{\mu}_0 = \boldsymbol{\gamma}_0.$$

- Stationary representation: We can also transform Eq (8) into

$$\boldsymbol{y}_t = (\boldsymbol{I} + \boldsymbol{\alpha}\boldsymbol{\beta}')\boldsymbol{y}_{t-1} + \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\beta}'\boldsymbol{y}_t = (\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\mu}_0 + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\beta}'\boldsymbol{y}_t = \sum_{i=0}^{\infty}(\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})^i\boldsymbol{\beta}'(\boldsymbol{\varepsilon}_{t-i} + \boldsymbol{\mu}_0)$$

Taking expectations:

$$E(\boldsymbol{\beta}'\boldsymbol{y}_t) = \sum_{i=0}^{\infty}(\boldsymbol{I} + \boldsymbol{\beta}'\boldsymbol{\alpha})^i\boldsymbol{\beta}'\boldsymbol{\mu}_0 = -(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\mu}_0$$

We write $\quad (\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\mu}_0 = \boldsymbol{\beta}_0.$

This is expression for *equilibrium mean* in the cointegrated VAR model.

- Consider the following identity again

$$\boldsymbol{I} = \boldsymbol{\beta}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\beta}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'$$

and apply to $\boldsymbol{\mu}_0$, we get

$$\boldsymbol{\mu}_0 = \boldsymbol{\alpha}\underline{(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\mu}_0} + \underline{\boldsymbol{\beta}'_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\beta})^{-1}\boldsymbol{\alpha}'_{\perp}\boldsymbol{\mu}_0} = \boldsymbol{\alpha}\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0.$$

Using this result, we can rewrite the cointegrated VAR

$$\Delta\boldsymbol{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

as follows:

$$\Delta\boldsymbol{y}_t - \boldsymbol{\gamma}_0 = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}_0) + \boldsymbol{\varepsilon}_t.$$

# 4 Estimation and Cointegration Tests

## 4.1 Specification of the Deterministic Function

Consider the following VECM

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t. \qquad (6)$$

1. No deterministics: $\boldsymbol{\mu}_t = 0$ (i.e. $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = 0$):

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t.$$

In this case, all the component series of $\boldsymbol{y}_t$ are $I(1)$ without drift (*no growth in data*) and the stationary series $\boldsymbol{w}_t = \boldsymbol{\beta}'\boldsymbol{y}_t$ has mean zero (*equilibrium term has zero mean*). (Almostly impossible!)

2. Restricted constant: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'\boldsymbol{\mu}_0$ is an $m$-dimensional nonzero constant vector. This implies that $\boldsymbol{\alpha}'_\perp\boldsymbol{\mu}_0 = 0$.

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}_0) + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

so that the components of $\boldsymbol{y}_t$ are $I(1)$ without drift, but $\boldsymbol{w}_t$ have a nonzero mean $-\boldsymbol{\beta}_0$. (Sometimes useful!)

3. Unrestricted constant: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$, which is nonzero. Constant can appear in both long-run and short-run, and the ECM becomes

$$\Delta \boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\beta}_0) + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}'_\perp\boldsymbol{\mu}_0$. Here the component series of $\boldsymbol{y}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ (*allowing for trends in the data series*) and $\boldsymbol{w}_t$ may have a nonzero mean. (Useful!)

4. Restricted trend and unrestricted constant: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}_1 t$, where $\boldsymbol{\beta}_1$ is a nonzero vector. Get constant in data and trend in cointegrating relations, the ECM becomes

$$\begin{aligned}\Delta \boldsymbol{y}_t &= \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\gamma}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 t) + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t\end{aligned}$$

so that the components of $\boldsymbol{y}_t$ are $I(1)$ with drift $\boldsymbol{\mu}_0$ and $\boldsymbol{w}_t$ has a linear time trend related to $\boldsymbol{\beta}_1 t$. (Useful!)

5. Unrestricted constant and trend: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, where $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are nonzero. In this case, the ECM becomes

$$\begin{aligned}\Delta \boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 t + \boldsymbol{\alpha}(\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 t) \\ + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t\end{aligned}$$

where $\boldsymbol{\gamma}_1 = \boldsymbol{\beta}_\perp(\boldsymbol{\alpha}'_\perp \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp \boldsymbol{\mu}_1$ and $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' \boldsymbol{\mu}_1$.

The components of $\boldsymbol{y}_t$ are $I(1)$ and have a quadratic time trend and $\boldsymbol{w}_t$ have a linear trend. (Almostly impossible!)

## 4.2 Maximum Likelihood Estimation

Suppose that the data are $\{\boldsymbol{y}_t, t = 1, \ldots, T\}$.

We write $\boldsymbol{\mu}_t = \boldsymbol{\mu}\boldsymbol{d}_t$, where $\boldsymbol{d}_t = [1, t]'$, and it is understood that $\boldsymbol{\mu}_t$ depends on the previous specification of the previous subsection.

For a given $m$, which is the rank of $\boldsymbol{\Pi}$, the ECM model becomes

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{9}$$

where $t = p+1, \ldots, T$.

Note that there are so many parameters and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identifiable.

A key step in the estimation is to *concentrate the likelihood function with respect to the deterministic term and the stationary effects*.

Consider the following two multivariate linear regressions:

$$\Delta \boldsymbol{y}_t = \gamma_0 \boldsymbol{d}_t + \boldsymbol{\Omega}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Omega}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t, \tag{10}$$

$$\boldsymbol{y}_{t-1} = \gamma_1 \boldsymbol{d}_t + \boldsymbol{\Xi}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Xi}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{v}_t. \tag{11}$$

Let $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$ be the residuals of Eqs. (10) and (11), respectively. Then,

$$\hat{\boldsymbol{u}}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \hat{\boldsymbol{v}}_t + \boldsymbol{\varepsilon}_t$$

and the *concentrated log-likelihood* is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = C - \frac{T-p}{2} \log |\boldsymbol{\Sigma}|$$

Assume that $\boldsymbol{\beta}$ is known, then

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}) = \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}, \quad \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10}$$

where $\boldsymbol{S}_{00} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'$, $\boldsymbol{S}_{01} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\boldsymbol{u}}_t\hat{\boldsymbol{v}}_t'$, $\boldsymbol{S}_{11} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\boldsymbol{v}}_t\hat{\boldsymbol{v}}_t'$.

The maximization of $L(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is now equivalent to the minimization of

$$|\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta})| = |\boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{S}_{10}| = \frac{|\boldsymbol{S}_{00}||\boldsymbol{\beta}'(\boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01})\boldsymbol{\beta}|}{|\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta}|}.$$

Here, we use the following result:

$$\begin{vmatrix} A & B \\ B' & C \end{vmatrix} = |A| \cdot |C - B'A^{-1}B| = |C| \cdot |A - BC^{-1}B'|,$$

where $A$ and $C$ are non-singular square matrices.

Next, compute the eigenvalues and eigenvectors of $\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}$ with respect to $\boldsymbol{S}_{11}$ (under the restriction $\boldsymbol{\beta}'\boldsymbol{S}_{11}\boldsymbol{\beta} = \boldsymbol{I}_m$). This amounts to solving the eigenvalue problem

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

Denote the eigenvalue and eigenvector pairs by $(\hat{\lambda}_i, \boldsymbol{e}_i)$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ and $\boldsymbol{e} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]$ is the matrix of eigenvectors.

The unnormalized maximum likelihood estimate (MLE) of the cointegrating vector $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m]$, from which we can obtain a MLE for $\boldsymbol{\beta}$ that satisfies the identifying constraint and normalization condition. Here, we also have the following result

$$\boldsymbol{\beta}'\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{\beta} = \mathrm{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m).$$

We can now express the determinant of the residual covariance matrix as:

$$|\widehat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})| = |S_{00}||\boldsymbol{I} - \mathrm{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m)| = |S_{00}| \prod_{i=1}^{m}(1 - \lambda_i).$$

Denote the resulting estimate by $\widehat{\boldsymbol{\beta}}_c$ with the subscript $c$ signifying constraints. The MLE of other parameters can then be obtained by the multivariate linear regression

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu} \boldsymbol{d}_t + \boldsymbol{\alpha} \widehat{\boldsymbol{\beta}}'_c \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t.$$

Under Gaussian innovations and the model is true, the estimates of the $\boldsymbol{\Phi}_j^*$ matrices are asymptotically normal and efficient.

The maximized value of the likelihood function based on $m$ cointegrating vectors is

$$\ell_{\max}^{-2/(T-p)} \propto |\boldsymbol{S}_{00}| \prod_{i=1}^{m} (1 - \hat{\lambda}_i).$$

This value is used in the maximum likelihood ratio test for testing $\mathrm{Rank}(\boldsymbol{\Pi}) = m$.

## 4.3 Johansen Cointegration Test

For a specified deterministic term $\boldsymbol{\mu}_t$, we now discuss the maximum likelihood test for testing the rank of the $\mathbf{\Pi}$ matrix in Eq. (4).

Let us consider

$$\ell_{\max} \propto \left[ |\boldsymbol{S}_{00}| \prod_{i=1}^{m} (1 - \hat{\lambda}_i) \right]^{-(T-p)/2}.$$

This likelihood function implies that

- $\hat{\lambda}_{m+1} = \hat{\lambda}_{m+2} = \cdots = \hat{\lambda}_n = 0$;
- Testing $\mathrm{Rank}(\mathbf{\Pi}) = m$ is equivalent to test $\hat{\lambda}_{m+1} = \cdots = \hat{\lambda}_n = 0$;

Let $H(m)$ be the null hypothesis that the rank of $\mathbf{\Pi}$ is $m$, i.e., $\mathrm{Rank}(\mathbf{\Pi}) = m$. The hypotheses of interest are

$$H(0) \subset \cdots \subset H(m) \subset \cdots \subset H(n).$$

where

- $H(0)$: $\mathrm{Rank}(\mathbf{\Pi}) = 0$ or all $\hat{\lambda}_i = 0$ for $i = 1, \ldots, n$, so that $\mathbf{\Pi} = 0$ and there is no cointegration;

- $H(m)$: $\mathrm{Rank}(\mathbf{\Pi}) = m$ or $\hat{\lambda}_{m+1} = \cdots = \hat{\lambda}_n = 0$, i.e. $n - m$ unit roots, $m$ cointegration relations; $\boldsymbol{y}_t$ is non-stationary;

- $H(n)$: $\mathrm{Rank}(\mathbf{\Pi}) = n$ or $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n \neq 0$, i.e. no unit roots; $\boldsymbol{y}_t$ is stationary.

## Trace test statistic

$$H_0 : \underbrace{\mathrm{Rank}(\mathbf{\Pi}) = m}_{\text{at most } m \to H(m)}, \quad \text{versus} \quad H_a : \underbrace{\mathrm{Rank}(\mathbf{\Pi}) > m}_{H(m+1)\cup\cdots\cup H(n) = H(n)} .$$

Johansen (1988) proposes the likelihood ratio (LR) statistic

$$\begin{aligned} LR_{\mathrm{tr}} &= -2\ln Q(H_m/H_n) = -2\ln[\ell_{\max}(m)/\ell_{\max}(n)] \\ &= (T-p)\ln\left\{\frac{|\boldsymbol{S}_{00}|(1-\hat{\lambda}_1)(1-\hat{\lambda}_2)\cdots(1-\hat{\lambda}_m)}{|\boldsymbol{S}_{00}|(1-\hat{\lambda}_1)(1-\hat{\lambda}_2)\cdots(1-\hat{\lambda}_m)\cdots(1-\hat{\lambda}_n)}\right\} \\ &= -(T-p)\sum_{i=m+1}^{n}\ln(1-\hat{\lambda}_i). \end{aligned}$$

If $\mathrm{Rank}(\mathbf{\Pi}) = m$, then $\hat{\lambda}_i$ should be small for $i > m$ and hence $LR_{\mathrm{tr}}(m)$ should be small. This is referred to as the trace statistic.

Due to the presence of unit roots, the asymptotic distribution of $LR_{\mathrm{tr}}(m)$ is not chi-squared, but a function of standard Brownian motions. Thus, critical values of $LR_{\mathrm{tr}}(m)$ must be obtained via simulation.

## Maximum eigenvalue test statistic

Johansen (1988) also considers a sequential procedure to determine the number of cointegrating vectors. Specifically,

$$H_0 : \underbrace{\mathrm{Rank}(\mathbf{\Pi}) = m}_{H(m)} \quad \text{versus} \quad H_a : \underbrace{\mathrm{Rank}(\mathbf{\Pi}) = m + 1}_{H(m+1)}.$$

The LR test statistic, called the maximum eigenvalue statistic, is

$$\begin{aligned} LR_{\max}(m) &= -2 \ln Q(H_m/H_{m+1}) = -2 \ln[\ell_{\max}(m)/\ell_{\max}(m+1)] \\ &= (T-p) \ln \left\{ \frac{|\boldsymbol{S}_{00}|(1-\hat{\lambda}_1)(1-\hat{\lambda}_2)\cdots(1-\hat{\lambda}_m)}{|\boldsymbol{S}_{00}|(1-\hat{\lambda}_1)(1-\hat{\lambda}_2)\cdots(1-\hat{\lambda}_m)(1-\hat{\lambda}_{m+1})} \right\} \\ &= -(T-p) \ln(1-\hat{\lambda}_{m+1}). \end{aligned}$$

Again, critical values of the maximum eigenvalue statistic $LR_{\max}(m)$ are nonstandard and must be evaluated via simulation.