Chapter Four

Multivariate Analysis and Vector Autoregressive Models

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

(For Undergraduate Program, 2026-2027, Spring Semester)

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

- **Outline:**
  - The Basics
  - Vector Autoregressive Process
  - Estimation and Model Specification
  - Structural Analysis
    - Granger causal testing
    - Impulse response anslysis
    - Variance decomposition
    - Connectedness
  - Structural VAR model
  - Cointegrated VAR model

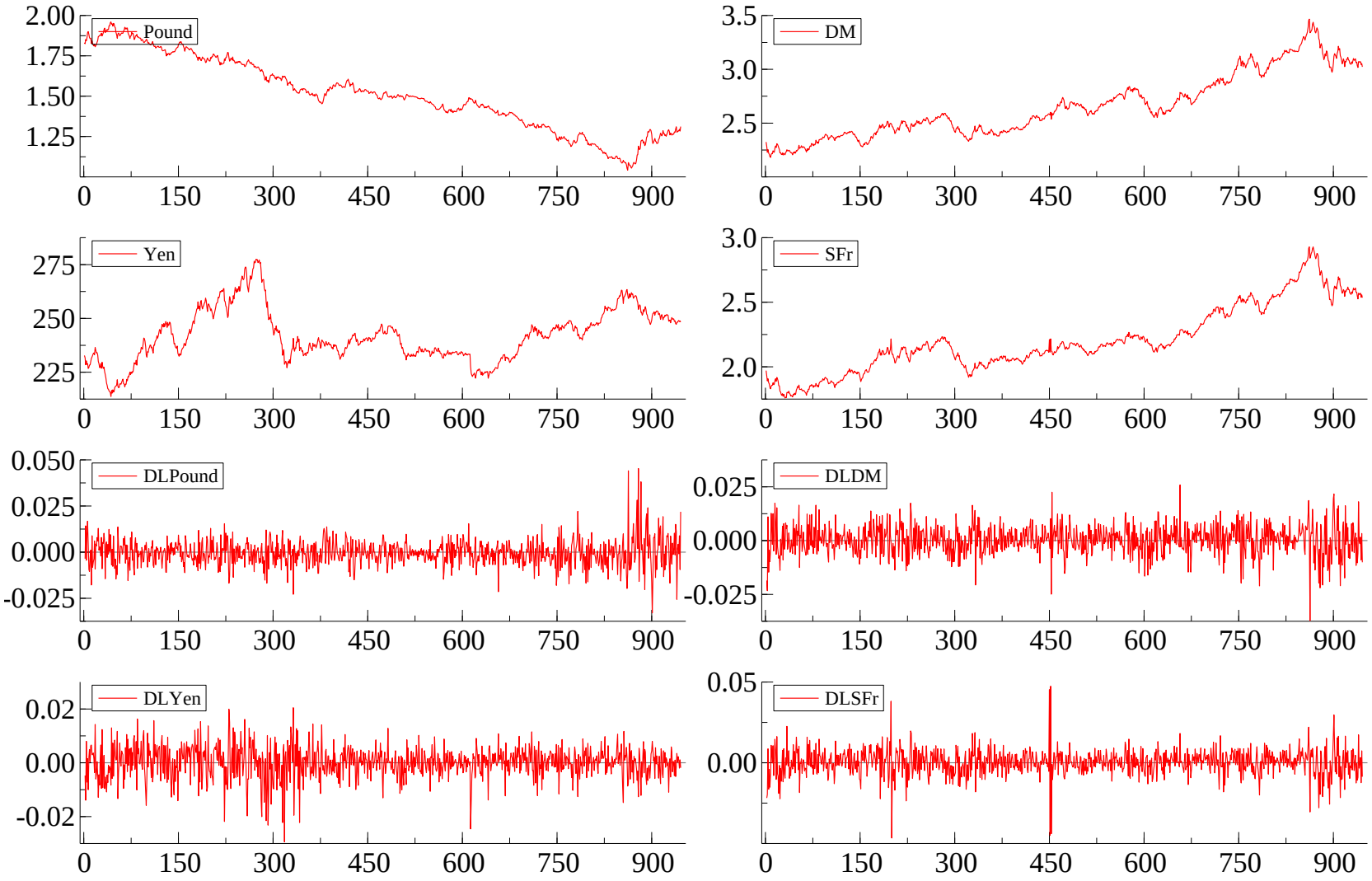

Figure 1 Exchange rate of Pound, DM, Yen, and SFr.

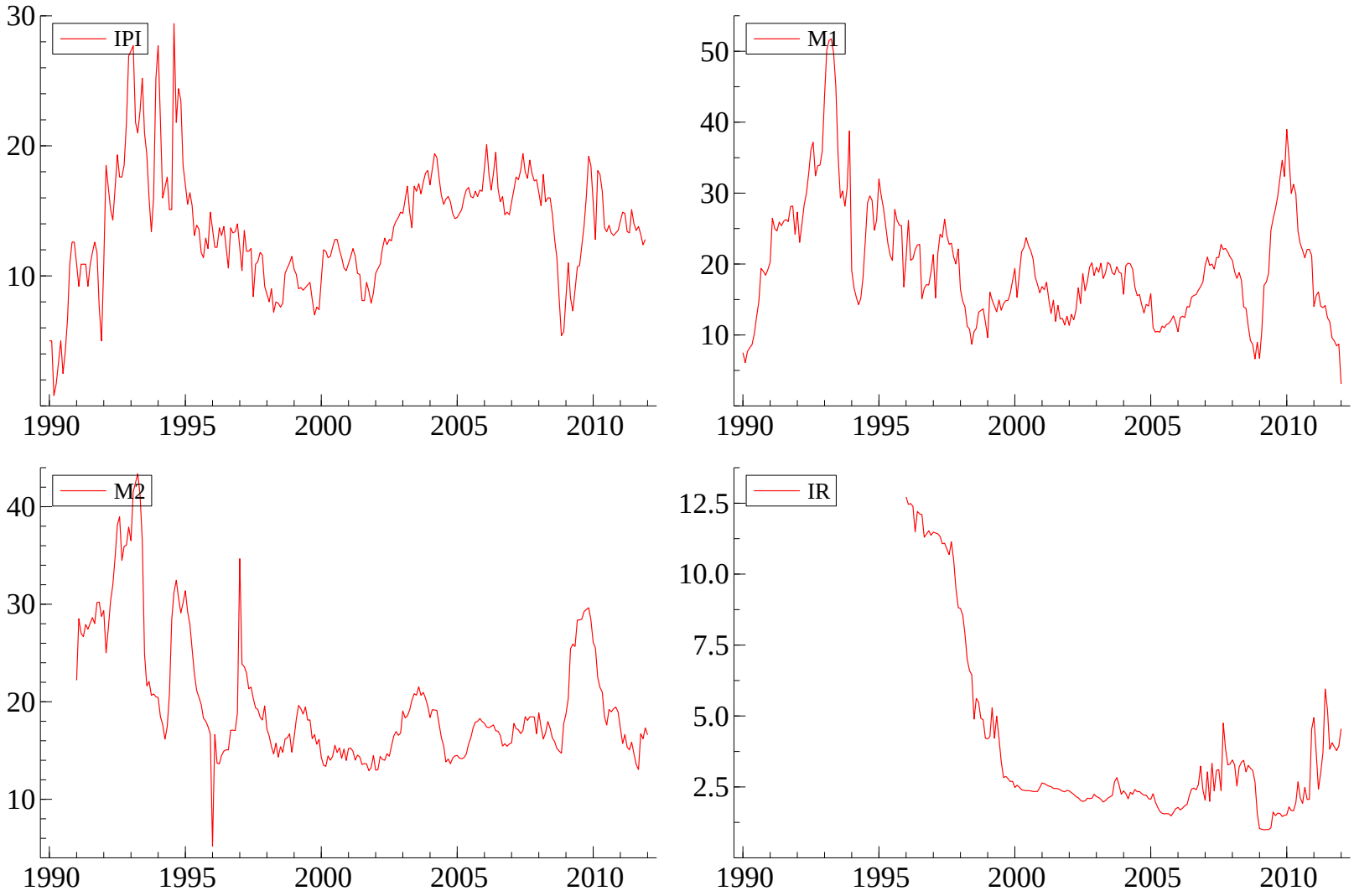

Figure 2 Money supply and output in China.

1. The Basics

1.1 Weak Stationarity

- Denote $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})'$;
- The series $\boldsymbol{y}_t$ is *weakly stationary* if its first and second moments are time-invariant.
- For a *weakly stationary* time series $\boldsymbol{y}_t$, we define its *mean vector* as

$$\boldsymbol{\mu} = E(\boldsymbol{y}_t) = \begin{bmatrix} E(y_{1t}) \\ E(y_{2t}) \\ \vdots \\ E(y_{nt}) \end{bmatrix} := \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix},$$

and the *covariance matrix* as

$$\boldsymbol{\Gamma}_0 = E[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_t - \boldsymbol{\mu})'] = \begin{bmatrix} \Gamma_{11}(0) & \Gamma_{12}(0) & \cdots & \Gamma_{1n}(0) \\ \Gamma_{21}(0) & \Gamma_{22}(0) & \cdots & \Gamma_{n2}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{n1}(0) & \Gamma_{n2}(0) & \cdots & \Gamma_{nn}(0) \end{bmatrix},$$

where

$$\Gamma_{ii}(0) = E[(y_{it} - \mu_i)^2],$$

and

$$\Gamma_{ij}(0) = E[(y_{it} - \mu_i)(y_{jt} - \mu_j)],$$

$i, j = 1, \ldots, n$.

- We write $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and $\boldsymbol{\Gamma}_0 = [\Gamma_{ij}(0)]$.

1.2 Cross-Correlation Matrices

- Let $\boldsymbol{D} = \text{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{nn}(0)}\} = \text{diag}\{\text{std}(y_{1t}), \ldots, \text{std}(y_{nt})\}$.

- The *concurrent*, or lag-zero, *cross-correlation matrix* is defined as

$$\boldsymbol{\rho}_0 \equiv [\rho_{ij}(0)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1} = \begin{bmatrix} \rho_{11}(0) & \rho_{12}(0) & \cdots & \rho_{1n}(0) \\ \rho_{21}(0) & \rho_{22}(0) & \cdots & \rho_{2n}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}(0) & \rho_{n2}(0) & \cdots & \rho_{nn}(0) \end{bmatrix}.$$

where the *correlation coefficient* between y_{it} and y_{jt} is

$$\rho_{ij}(0) = \text{corr}(y_{it}, y_{jt}) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\text{Cov}(y_{it}, y_{jt})}{\text{std}(y_{it})\text{std}(y_{jt})}.$$

– It is easy to see that

$$\rho_{ij}(0) = \text{corr}(y_{it}, y_{jt}) = \text{corr}(y_{jt}, y_{it}) = \rho_{ji}(0),$$

$$-1 \leq \rho_{ij}(0) \leq 1,$$

and $\quad \rho_{ii}(0) = 1$

for $1 \leq i, j \leq n$.

– Thus, $\boldsymbol{\rho}_0 = \boldsymbol{\rho}(0)$ is a symmetric matrix with unit *diagonal* elements,

$$\boldsymbol{\rho}_0 = \boldsymbol{\rho}(0) = \begin{bmatrix} 1 & \rho_{21}(0) & \cdots & \rho_{n1}(0) \\ \rho_{21}(0) & 1 & \cdots & \rho_{n2}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}(0) & \rho_{n2}(0) & \cdots & 1 \end{bmatrix}.$$

1.3 Lead−lag relationships

- We next propose the *lagged cross-correlation matrices*, which are used to measure the strength of *linear dependence* between time series.

- Under weak stationarity of $\{\boldsymbol{y}_t\}$, the lag-ℓ *cross-covariance matrix* is defined as

$$
\begin{aligned}
\boldsymbol{\Gamma}_\ell \equiv [\Gamma_{ij}(\ell)] &= E[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-\ell} - \boldsymbol{\mu})'] \\
&= \begin{bmatrix}
\mathrm{Cov}(y_{1t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{1t}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{1t}, y_{n,t-\ell}) \\
\mathrm{Cov}(y_{2t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{2t}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{2t}, y_{n,t-\ell}) \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{Cov}(y_{nt}, y_{1,t-\ell}) & \mathrm{Cov}(y_{nt}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{nt}, y_{n,t-\ell})
\end{bmatrix} \\
&= \begin{bmatrix}
\Gamma_{11}(\ell) & \Gamma_{12}(\ell) & \cdots & \Gamma_{1n}(\ell) \\
\Gamma_{21}(\ell) & \Gamma_{22}(\ell) & \cdots & \Gamma_{n2}(\ell) \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{n1}(\ell) & \Gamma_{n2}(\ell) & \cdots & \Gamma_{nn}(\ell)
\end{bmatrix}. \qquad (1)
\end{aligned}
$$

- The lag-ℓ *cross-correlation matrix* (CCM) of $\boldsymbol{y}_t$ is defined as

$$\boldsymbol{\rho}_\ell \equiv [\rho_{ij}(\ell)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell \boldsymbol{D}^{-1} = \begin{bmatrix} \rho_{11}(\ell) & \rho_{12}(\ell) & \cdots & \rho_{1n}(\ell) \\ \rho_{21}(\ell) & \rho_{22}(\ell) & \cdots & \rho_{2n}(\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}(\ell) & \rho_{n2}(\ell) & \cdots & \rho_{nn}(\ell) \end{bmatrix}, \quad (2)$$

  where

$$\rho_{ij}(\ell) = \text{corr}(y_{it}, y_{j,t-\ell}) = \frac{\text{Cov}(y_{it}, y_{j,t-\ell})}{\text{std}(y_{it})\text{std}(y_{jt})} = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}. \quad (3)$$

  – The diagonal element $\rho_{ii}(\ell)$ is simply the lag-ℓ *autocorrelation coefficient* of y_{it}. In general, $\rho_{ii}(\ell) \neq 1$.
  – When $\ell > 0$, the correlation coefficient $\rho_{ij}(\ell)$ measures the *linear dependence* of y_{it} on $y_{j,t-\ell}$. If $\rho_{ij}(\ell) \neq 0$ and $\ell > 0$, we say that the series y_{jt} leads the series y_{it} at lag ℓ.
  – Similarly, $\rho_{ji}(\ell)$ measures the *linear dependence* of y_{jt} and $y_{i,t-\ell}$, and we say that the series y_{it} leads the series y_{jt} at lag ℓ if $\rho_{ji}(\ell) \neq 0$ and $\ell > 0$.

Important properties

- $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$ for $\ell > 0$ are in general *not symmetric*.

- Using $\mathrm{Cov}(x, y) = \mathrm{Cov}(y, x)$ and the weak stationarity assumption,

$$\begin{aligned}\Gamma_{ij}(\ell) &= \mathrm{Cov}(y_{it}, y_{j,t-\ell}) \\ &= \mathrm{Cov}(y_{j,t-\ell}, y_{it}) = \mathrm{Cov}(y_{jt}, y_{i,t+\ell}) = \mathrm{Cov}(y_{jt}, y_{i,t-(-\ell)}) \\ &= \Gamma_{ji}(-\ell).\end{aligned}$$

  but $\neq \Gamma_{ji}(\ell)$, and so $\boldsymbol{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$ are not symmetric.

- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Gamma}'_{-\ell}$ and $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$.
  (Because $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$, it suffices in practice to consider the cross-correlation matrices $\boldsymbol{\rho}_\ell$ for $\ell \geq 0$)

1.4 Linear Dependence

Considered jointly, the cross-correlation matrices $\{\boldsymbol{\rho}_\ell \mid \ell = 0, 1, \ldots\}$ of a weakly stationary vector time series contain the following information:

- The diagonal elements $\rho_{ii}(\ell),\ \ell = 0, 1, \ldots$ are the autocorrelation function (ACF) of y_{it}.

- The *off-diagonal* element $\rho_{ij}(0)$ measures the *concurrent* linear relationship between y_{it} and y_{jt}.

- For $\ell > 0$, the off-diagonal element $\rho_{ij}(\ell)$ measures the *linear dependence* of y_{it} on the past value $y_{j,t-\ell}$.

In general, the linear relationship between two time series $\{y_{it}\}$ and $\{y_{jt}\}$ can be summarized as follows:

1. y_{it} and y_{jt} have *no linear relationship* if $\rho_{ij}(\ell) = \rho_{ji}(\ell) = 0$ for all $\ell \geq 0$.

2. y_{it} and y_{jt} are *concurrently correlated* if $\rho_{ij}(0) \neq 0$.

3. y_{it} and y_{jt} have *no lead–lag relationship* if $\rho_{ij}(\ell) = 0$ and $\rho_{ji}(\ell) = 0$ for all $\ell > 0$.

4. There is a *unidirectional relationship* from y_{it} to y_{jt} if $\rho_{ij}(\ell) = 0$ for all $\ell > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$. In this case, y_{it} does not depend on any past value of y_{jt}, but y_{jt} depends on some past values of y_{it}.

5. There is a *feedback relationship* between y_{it} and y_{jt} if $\rho_{ij}(\ell) \neq 0$ for some $\ell > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

1.5 Sample Cross-Correlation Matrices

Given the data $\{\boldsymbol{y}_t\}_{t=1}^T$, the *cross-covariance matrix* $\boldsymbol{\Gamma}_\ell$ can be estimated by

$$\hat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T}\sum_{t=\ell+1}^{T}(\boldsymbol{y}_t - \bar{\boldsymbol{y}})(\boldsymbol{y}_{t-\ell} - \bar{\boldsymbol{y}})', \quad \ell \geq 0, \tag{4}$$

where $\bar{\boldsymbol{y}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t$ is the vector of *sample means*.

The *cross-correlation matrix* $\boldsymbol{\rho}_\ell$ is estimated by

$$\hat{\boldsymbol{\rho}}_\ell = \hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\Gamma}}_\ell\hat{\boldsymbol{D}}^{-1}, \quad \ell \geq 0, \tag{5}$$

where $\hat{\boldsymbol{D}} = \mathrm{diag}\{\widehat{\Gamma}_{11}^{1/2}(0), \ldots, \widehat{\Gamma}_{nn}^{1/2}(0)\} = \mathrm{diag}\{\widehat{\mathrm{std}}(y_{1t}), \ldots, \widehat{\mathrm{std}}(y_{nt})\}$ is the $n \times n$ *diagonal* matrix of the *sample standard deviations* of the component series.

CCF test between two series

Consider the linear dependence of $\{y_t\}$ on $\{x_t\}$ and test for $\rho_{xy}(\ell) = 0$, $\ell > 0$.

Under the hypothesis that there is *no relation* at time ℓ and that at least one of the two series are independent and identically distributed, we also have the result:

$$\sqrt{n}\hat{\rho}_{xy}(\ell) \overset{A}{\sim} N(0, 1).$$

1.6 Multivariate Portmanteau Tests

The univariate Ljung−Box statistic $Q(m)$ has been generalized to the multivariate case by Hosking (1980, 1981) and Li and McLeod (1981).

The null hypothesis is $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0}$, and the alternative hypothesis $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $i \in \{1, \ldots, m\}$. Thus, the statistic is used to test that there are *no auto- and cross-correlations* in the vector series $\boldsymbol{y}_t$.

The test statistic assumes the form

$$Q_n(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \mathrm{tr}(\hat{\boldsymbol{\Gamma}}_\ell' \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{\Gamma}}_0^{-1}) \quad \sim \chi^2(n^2 m), \tag{6}$$

where T is the sample size, n is the dimension of $\boldsymbol{y}_t$, and $\mathrm{tr}(A)$ is the trace of the matrix A, which is the sum of the diagonal elements of A.

1.7 Multivariate Wold Representation

Any $(n \times 1)$ covariance stationary multivariate time series $\boldsymbol{y}_t$ has a Wold or linear process representation of the form

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}\end{aligned}$$

where $\boldsymbol{\Psi}_0 = I_n$, $\boldsymbol{\varepsilon}_t \sim WN(0, \boldsymbol{\Sigma})$, and $\boldsymbol{\Psi}_k$ is an $n \times n$ matrix with (i, j)th element $\psi_{ij}^{(k)}$.

$$\boldsymbol{\Psi}_k = [\psi_{ij}^{(k)}] = \begin{bmatrix} \psi_{11}^{(k)} & \cdots & \psi_{1n}^{(k)} \\ \vdots & \ddots & \vdots \\ \psi_{n1}^{(k)} & \cdots & \psi_{nn}^{(k)} \end{bmatrix}.$$

In lag operator notation, the Wold form is

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\Psi}(L) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k L^k.$$

The moments of $\boldsymbol{y}_t$ are given by

$$E(\boldsymbol{y}_t) = \boldsymbol{\mu}, \quad \mathrm{Var}(\boldsymbol{y}_t) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_k'.$$

2. Vector Autoregressive Process

2.1 Vector AR(1) Model

A multivariate time series $\boldsymbol{y}_t$ is a VAR(1) process, if it follows the model

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{7}$$

$$E(\boldsymbol{\varepsilon}_t) = 0, \quad \text{and} \quad E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s) = \begin{cases} \boldsymbol{\Sigma}, & \text{if } t = s; \\ \mathbf{0}, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\phi}_0$ is a n-dimensional vector, $\boldsymbol{\Phi}$ is a $n \times n$ matrix. In *empirical applications*, it is often assumed that $\boldsymbol{\varepsilon}_t$ is *multivariate normal*, i.e.

$$\boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}).$$

For instance, a VAR(1) model consists of the following two equations:

$$\begin{aligned} y_{1t} &= \phi_{10} + \Phi_{11} y_{1,t-1} + \Phi_{12} y_{2,t-1} + \varepsilon_{1t}, \\ y_{2t} &= \phi_{20} + \Phi_{21} y_{1,t-1} + \Phi_{22} y_{2,t-1} + \varepsilon_{2t}, \end{aligned}$$

where Φ_{ij} is the $(i,j)^{th}$ element of $\mathbf{\Phi}$ and ϕ_{i0} is the i^{th} element of $\boldsymbol{\phi}_0$.

- First equation: Φ_{12} is the *conditional effect* of $y_{2,t-1}$ on y_{1t} given $y_{1,t-1}$.

If $\Phi_{12} = 0$, then y_{1t} does not depend on $y_{2,t-1}$, but depends on $y_{1,t-1}$.

- Second equation: Similar to the first equation.

- Two equations jointly:

(a) *An unidirectional relationship* from y_{1t} to y_{2t} if $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$.

(b) y_{1t} and y_{2t} are *uncoupled*, if $\Phi_{12} = \Phi_{21} = 0$.

(c) *A feedback relationship* between two series if $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$.

(1) Reduced and Structural Forms

The VAR(1) model (7) is a *reduced-form*. We can rewrite it equivalently as a *structural form* by using Cholesky decomposition.

Cholesky decomposition: For a *symmetric matrix* $\boldsymbol{A}$, there exists a *lower triangular matrix* $\boldsymbol{L}$ with diagonal elements being 1 and a *diagonal matrix* $\boldsymbol{G}$ such that $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$.

If $\boldsymbol{A}$ is positive definite, then the diagonal elements of $\boldsymbol{G}$ are positive. In this case, we have

$$\boldsymbol{A} = \boldsymbol{L}\sqrt{\boldsymbol{G}}\sqrt{\boldsymbol{G}}\boldsymbol{L}' = (\boldsymbol{L}\sqrt{\boldsymbol{G}})(\boldsymbol{L}\sqrt{\boldsymbol{G}})',$$

where $\boldsymbol{L}\sqrt{\boldsymbol{G}}$ is again a lower triangular matrix. Such a decomposition is called the *Cholesky decomposition* of $\boldsymbol{A}$.

Example 1. If we take the symmetric matrix $\boldsymbol{A} = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, the Cholesky decomposition of $\boldsymbol{A}$ is

$$\boldsymbol{C} = \boldsymbol{L}\sqrt{\boldsymbol{G}} = \begin{bmatrix} 2 & 0 \\ 0.5 & 1.658 \end{bmatrix},$$

and then $\boldsymbol{G} = \begin{bmatrix} 4 & 0 \\ 0 & 2.75 \end{bmatrix}$, $\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 0.25 & 1 \end{bmatrix}$.

(This use the program example in Oxmetrics function `decldl()`)

The Choleski decomposition shows that a *positive-definite matrix* $\boldsymbol{A}$ can be diagonalized as

$$\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}')^{-1} = \boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{-1})' = \boldsymbol{G}.$$

Structural form

Again, the VAR(1) model

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{7}$$

- Define $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{nt})' = \boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_t$. Then $\mathrm{E}(\boldsymbol{\eta}_t) = \boldsymbol{L}^{-1}\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$,

$$\mathrm{Cov}(\boldsymbol{\eta}_t) = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^{-1})' = \boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}')^{-1} = \boldsymbol{G}.$$

  Since $\boldsymbol{G}$ is a diagonal matrix, the components of $\boldsymbol{\eta}_t$ are *uncorrelated*.

- Multiplying $\boldsymbol{L}^{-1}$ from the left to model (7), we obtain

$$\boldsymbol{L}^{-1}\boldsymbol{y}_t = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0 + \boldsymbol{L}^{-1}\boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{\phi}_0^* + \boldsymbol{\Phi}^*\boldsymbol{y}_{t-1} + \boldsymbol{\eta}_t, \tag{8}$$

  where $\boldsymbol{\phi}_0^* = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0$ is a $n \times 1$ vector and $\boldsymbol{\Phi}^* = \boldsymbol{L}^{-1}\boldsymbol{\Phi}$ is a $n \times n$ matrix.

- Because of the special matrix structure, the nth row of L^{-1} is in the form $(w_{n1}, w_{n2}, \ldots, w_{n,n-1}, 1)$. Consequently, the nth equation is

$$y_{nt} + \sum_{i=1}^{n-1} w_{ni} y_{it} = \phi_{n,0}^* + \sum_{i=1}^{n} \Phi_{ni}^* y_{i,t-1} + \eta_{nt}, \tag{9}$$

  where $\phi_{n,0}^*$ is the n^{th} element of $\boldsymbol{\phi}_0^*$, Φ_{ni}^* is the $(n,i)^{th}$ element of $\boldsymbol{\Phi}^*$.

- Because η_{nt} is uncorrelated with η_{it} for $1 \leq i < n$, Eq. (9) shows explicitly the concurrent linear dependence of y_{nt} on y_{it}, where $1 \leq i \leq n-1$. This equation is referred to as a *structural equation* for y_{nt}.

- The *reduced-form* model (7) is equivalent to the structural form (8).

- See Example 8.3. in Tsay's Textbook, page 400-401.

(2) Stationarity Condition and Moments

Again, the VAR(1) model

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{7}$$

Stationarity: Similar to the scalar AR process, for the VAR(1) process to be *weakly stationary*, we must have that the roots in the *characteristic equation*

$$\det(\boldsymbol{I} - \boldsymbol{\Phi}z) = |\boldsymbol{I} - \boldsymbol{\Phi}z| = 0, \tag{10}$$

all *lie outside* the unit circle.

In other ways, it is equivalent to say that the roots in the polynomial equation

$$\det(\lambda\boldsymbol{I} - \boldsymbol{\Phi}) = |\lambda\boldsymbol{I} - \boldsymbol{\Phi}| = 0, \tag{11}$$

all *lie inside* the unit circle, where $\lambda = z^{-1}$.

Spectral radius: Let $\lambda_1, \ldots, \lambda_n$ be the *(real or complex) eigenvalues* of the matrix $\mathbf{\Phi}$. Then its spectral radius of $\mathbf{\Phi}$, that is $\rho(\mathbf{\Phi})$, is defined as:

$$\rho(\mathbf{\Phi}) \equiv \max\{|\lambda_1|, \ldots, |\lambda_n|\} \tag{12}$$

Hence, the stationary condition is further equivalent to $\rho(\mathbf{\Phi}) < 1$.

The condition (10) provides an easy tool for checking the stability of a VAR process.

Example 2. Consider the three-dimensional VAR(1) process

$$\boldsymbol{y}_t = \boldsymbol{c} + \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{13}$$

For this process the characteristic equation is

$$\begin{aligned} \det\left(I_3 - \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{bmatrix} z \right) &= \det\left(\begin{bmatrix} 1-.5z & 0 & 0 \\ -.1z & 1-.1z & -.3z \\ 0 & .2z & 1-.3z \end{bmatrix} \right) \\ &= (1-.5z)(1-.4z-.03z^2). \end{aligned}$$

The roots of this polynomial are easily seen to be

$$z_1 = 2, \qquad z_2 = 2.1525, \qquad z_3 = -15.4858.$$

They are obviously all greater than 1 in absolute value. Therefore the process is stationary.

Moments

Again, the VAR(1) model

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{7}$$

Assume that the VAR(1) model in Eq. (7) is *weakly stationary*.

- Taking *expectation* of the model and using $\mathrm{E}(\boldsymbol{\varepsilon}_t) = 0$, we obtain

$$\mathrm{E}(\boldsymbol{y}_t) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\mathrm{E}(\boldsymbol{y}_{t-1}).$$

- Since $\mathrm{E}(\boldsymbol{y}_t)$ is time-invariant, we have *(Note: $\rho(\boldsymbol{\Phi}) < 1$)*

$$\boldsymbol{\mu} \equiv \mathrm{E}(\boldsymbol{y}_t) = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_0.$$

- Using $\phi_0 = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}$, the VAR(1) model in Eq. (7) can be written as

$$(\boldsymbol{y}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t.$$

- Let $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{\mu}$ be the *mean-corrected* time series. Then the VAR(1) model becomes

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{14}$$

- By repeated substitutions, we can rewrite Eq. (14) as

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{\varepsilon}_{t-2} + \boldsymbol{\Phi}^3\boldsymbol{\varepsilon}_{t-3} + \cdots .$$

  This expression shows several characteristics of a VAR(1) process:

  – $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{y}_{t-1}) = \boldsymbol{0}$.

– $\mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$.

– $\boldsymbol{y}_t$ depends on the past innovation $\boldsymbol{\varepsilon}_{t-j}$ with coefficient matrix $\boldsymbol{\Phi}^j$.

– The covariance matrix of $\boldsymbol{y}_t$ is

$$\mathrm{Cov}(\boldsymbol{y}_t) = \boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}' + \boldsymbol{\Phi}^2\boldsymbol{\Sigma}(\boldsymbol{\Phi}^2)' + \cdots = \sum_{i=0}^{\infty} \boldsymbol{\Phi}^i \boldsymbol{\Sigma} (\boldsymbol{\Phi}^i)'$$

or

$$\mathrm{vec}(\mathrm{Cov}(\boldsymbol{y}_t)) = \mathrm{vec}(\boldsymbol{\Phi}\mathrm{Cov}(\boldsymbol{y}_{t-1})\boldsymbol{\Phi}' + \boldsymbol{\Sigma})$$

$$\mathrm{vec}(\boldsymbol{\Gamma}_0) = (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi}) \cdot \mathrm{vec}(\boldsymbol{\Gamma}_0) + \mathrm{vec}(\boldsymbol{\Sigma}),$$

$$\mathrm{vec}(\boldsymbol{\Gamma}_0) = [\boldsymbol{I} - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})]^{-1}\mathrm{vec}(\boldsymbol{\Sigma}).$$

Proposition 1. *Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ be matrices whose dimensions are such that the product $\boldsymbol{ABC}$ exists. Then*

$$\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A}) \cdot \mathrm{vec}(\boldsymbol{B}),$$

where the symbol $\otimes$ denotes the Kronecker product.

Here for $\mathrm{vec}()$ and $\otimes$, we have

$$\mathrm{vec}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 2 \\ 4 \\ 6 \end{bmatrix}, \quad \underset{m\times n}{A} \otimes \underset{p\times q}{B} = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{bmatrix}_{(mp\times nq)}$$

Example 3. Using, for instance, $\boldsymbol{\Sigma} = \begin{bmatrix} 2.25 & 0 & 0 \\ 0 & 1.0 & .5 \\ 0 & .5 & .74 \end{bmatrix}$, we get for the example process (13), then

$$\text{vec}(\boldsymbol{\Gamma}_0) = [I_9 - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})]^{-1}\text{vec}(\boldsymbol{\Sigma})$$

$$= \begin{bmatrix} .75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.05 & .95 & -.15 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.10 & .85 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.05 & 0 & 0 & .95 & 0 & 0 & -.15 & 0 & 0 \\ -.01 & -.01 & -.03 & -.01 & 0.99 & -.03 & -.03 & -.03 & -.09 \\ 0 & -.02 & -.03 & 0 & -.02 & .97 & 0 & -.06 & -.09 \\ 0 & 0 & 0 & -.01 & 0 & 0 & .85 & 0 & 0 \\ 0 & 0 & 0 & -.02 & -.02 & -.06 & -.03 & .97 & -.09 \\ 0 & 0 & 0 & 0 & -.04 & -.06 & 0 & -.06 & .91 \end{bmatrix}^{-1} \begin{bmatrix} 2.25 \\ 0 \\ 0 \\ 0 \\ 1.0 \\ .5 \\ 0 \\ .5 \\ .74 \end{bmatrix} = \begin{bmatrix} 3.000 \\ .161 \\ .019 \\ .161 \\ 1.172 \\ .674 \\ .019 \\ .674 \\ .954 \end{bmatrix}$$

It follows that

$$\boldsymbol{\Gamma}_0 = \begin{bmatrix} 3.000 & .161 & .019 \\ .161 & 1.172 & .674 \\ .019 & .674 & .954 \end{bmatrix}.$$

Again, the mean-corrected VAR(1) model

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{14}$$

- **Moment equation:** Postmultiplying $\tilde{\boldsymbol{y}}'_{t-\ell}$ to Eq. (14), taking expectation, and using the result $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{y}_{t-j}) = \mathrm{E}(\boldsymbol{\varepsilon}_t\tilde{\boldsymbol{y}}'_{t-j}) = \mathbf{0}$ for $j > 0$, we obtain

$$\mathrm{E}(\tilde{\boldsymbol{y}}_t\tilde{\boldsymbol{y}}'_{t-\ell}) = \boldsymbol{\Phi}\mathrm{E}(\tilde{\boldsymbol{y}}_{t-1}\tilde{\boldsymbol{y}}'_{t-\ell}), \quad \ell > 0.$$

Therefore,

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}, \quad \ell > 0. \tag{15}$$

where $\boldsymbol{\Gamma}_j$ is the lag-j *cross-covariance matrix* of $\boldsymbol{y}_t$.

- By repeated substitutions, Eq. (15) shows that

$$\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}^\ell \boldsymbol{\Gamma}_0, \quad \text{for } \ell > 0.$$

- Pre- and post-multiplying Eq. (15) by $\boldsymbol{D}^{-1}$, we obtain

$$\begin{aligned} \boldsymbol{\rho}_\ell &= \boldsymbol{D}^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1} \\ &= \boldsymbol{D}^{-1}\boldsymbol{\Phi}\boldsymbol{D} \cdot \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1} \\ \boldsymbol{\rho}_\ell &= \boldsymbol{\Upsilon}\boldsymbol{\rho}_{\ell-1}, \end{aligned}$$

  where $\boldsymbol{\Upsilon} = \boldsymbol{D}^{-1}\boldsymbol{\Phi}\boldsymbol{D}$.

- Consequently, the *correlation coefficient matrix* (CCM) of a VAR(1) model satisfies

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}^\ell \boldsymbol{\rho}_0, \quad \text{for } \ell > 0.$$

2.2 Vector AR(p) Models

The time series $\boldsymbol{y}_t$ follows a VAR(p) model if it satisfies

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad p > 0, \tag{16}$$

$$E(\boldsymbol{\varepsilon}_t) = 0, \quad \text{and} \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_\tau) = \begin{cases} \boldsymbol{\Sigma}, & \text{if } t = \tau; \\ \mathbf{0}, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\phi}_0$ and $\boldsymbol{\varepsilon}_t$ are defined as before, and $\boldsymbol{\Phi}_j$ are $n \times n$ matrices.

Using the *back-shift operator* B, the VAR(p) model canbe written as

$$\begin{aligned} (\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{y}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\varepsilon}_t, \\ \Longrightarrow \qquad \boldsymbol{\Phi}(B)\boldsymbol{y}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\varepsilon}_t, \end{aligned}$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ is a *matrix polynomial*.

(1) Moment properties

If $\boldsymbol{y}_t$ is *weakly stationary*, then we have

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$$

provided that the inverse exists.

Let $\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{\mu}$. The VAR($p$) model becomes

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{\Phi}_1\tilde{\boldsymbol{y}}_{t-1} + \cdots + \boldsymbol{\Phi}_p\tilde{\boldsymbol{y}}_{t-p} + \boldsymbol{\varepsilon}_t. \tag{17}$$

Under weak stationarity of $\{\boldsymbol{y}_t\}$, we have the following Wold representation

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{\mu}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots$$

Using this equation and the same techniques as those for VAR(1) models, we obtain that:

- $\mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{\varepsilon}_t$;
- $\mathrm{Cov}(\boldsymbol{y}_{t-\ell}, \boldsymbol{\varepsilon}_t) = \mathbf{0}$ for $\ell > 0$;
- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{\ell-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}_{\ell-p}$ for $\ell > 0$, since

$$\mathrm{E}(\tilde{\boldsymbol{y}}_t \tilde{\boldsymbol{y}}'_{t-\ell}) = \boldsymbol{\Phi}_1 \mathrm{E}(\tilde{\boldsymbol{y}}_{t-1} \tilde{\boldsymbol{y}}'_{t-\ell}) + \cdots + \boldsymbol{\Phi}_1 \mathrm{E}(\tilde{\boldsymbol{y}}_{t-p} \tilde{\boldsymbol{y}}'_{t-\ell}), \quad \ell > 0.$$

The last property is called the *moment equations* of a VAR(p) model. It is a multivariate version of the *Yule−Walker equation* of AR(p) model.

In terms of CCM, the moment equations become

$$\begin{aligned}\boldsymbol{\rho}_\ell &= \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell\boldsymbol{D}^{-1} \\ &= \boldsymbol{D}^{-1}\boldsymbol{\Phi}_1\boldsymbol{\Gamma}_{\ell-1}\boldsymbol{D}^{-1} + \cdots + \boldsymbol{D}^{-1}\boldsymbol{\Phi}_p\boldsymbol{\Gamma}_{\ell-p}\boldsymbol{D}^{-1} \\ &= \boldsymbol{\Upsilon}_1\boldsymbol{\rho}_{\ell-1} + \cdots + \boldsymbol{\Upsilon}_p\boldsymbol{\rho}_{\ell-p}\end{aligned}$$

for $\ell > 0$, where $\boldsymbol{\Upsilon}_i = \boldsymbol{D}^{-1}\boldsymbol{\Phi}_i\boldsymbol{D}$.

(2) State-space representation

Sometimes, it is more covenient to write a scalar valued time series, say an AR(p) process, in vector form.

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\boldsymbol{\xi}_t = F\boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_t$$

So we have rewrite an AR(p) scalar process as an VAR(1).

Similarly, we could also transform a VAR(p) process to a *"companion form (友矩阵形式)" VAR(1) process* with np-dimensional

$$\boldsymbol{\xi}_t = \boldsymbol{\Phi}^* \boldsymbol{\xi}_{t-1} + \boldsymbol{u}_t, \tag{18}$$

where

$$\boldsymbol{\xi}_t = \begin{bmatrix} \tilde{\boldsymbol{y}}_t \\ \tilde{\boldsymbol{y}}_{t-1} \\ \tilde{\boldsymbol{y}}_{t-2} \\ \vdots \\ \tilde{\boldsymbol{y}}_{t-p+1} \end{bmatrix}, \quad \underset{np\times np}{\boldsymbol{\Phi}^*} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \boldsymbol{I}_n & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_n & \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{u}_t = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}.$$

Furthermore, $\boldsymbol{u}_t \sim N_{np}(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$.

(3) Stationarity condition

Using the stationarity condition (10) for VAR(1) process, we can easily get the following general result for VAR(p) process.

The VAR(p) is covariance-stationary, if all the values of z satisfying

$$\det(\boldsymbol{I}_{np\times np} - \boldsymbol{\Phi}^*{}_{np\times np} \cdot z) = |\boldsymbol{I}_n - \boldsymbol{\Phi}_1 z - \boldsymbol{\Phi}_2 z^2 - \cdots - \boldsymbol{\Phi}_p z^p| = 0$$

lie outside the unit circle, or equivalently, if the eigenvalues of $\boldsymbol{\Phi}^*{}_{np\times np}$ lie inside the unit circle. The eigenvalues of $\boldsymbol{\Phi}^*{}_{np\times np}$ satisfy

$$\det(\lambda \boldsymbol{I}_{np\times np} - \boldsymbol{\Phi}^*{}_{np\times np}) = |\boldsymbol{I}_n \lambda^p - \boldsymbol{\Phi}_1 \lambda^{p-1} - \boldsymbol{\Phi}_2 \lambda^{p-2} - \cdots - \boldsymbol{\Phi}_p| = 0.$$

Hence, the stationary condition is also required such that $\rho(\boldsymbol{\Phi}^*) < 1$.

Example 4. Consider the bivariate (two-dimensional) VAR(2) process

$$\boldsymbol{y}_t = \boldsymbol{c} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \boldsymbol{y}_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \boldsymbol{y}_{t-2} + \boldsymbol{\varepsilon}_t.$$

Its characteristic polynomial is

$$\det\left(I_2 - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} z^2 \right) = 1 - z + .21z^2 - .025z^3.$$

The roots of this polynomial (using `http://www.99cankao.com/algebra/cubic-equation.php`) are

$$z_1 = 1.3, \qquad z_2 = 3.55 + 4.26i, \qquad z_3 = 3.55 - 4.26i$$

Here $i := \sqrt{-1}$ denotes the imaginary unit. Note that the modulus of z_2 and z_3 is $|z_2| = |z_3| = \sqrt{3.55^2 + 4.26^2} = 5.545$. Thus, the process satisfies the stability condition because all roots are outside the unit circle.

(4) Vector moving average representation

Recall that we could invert a scalar stationary AR(p) process, $\phi(L)y_t = \varepsilon_t$ to a MA(∞) process, $y_t = \psi(L)\varepsilon_t$, where $\psi(L) = \phi(L)^{-1}$.

Invertibility: The same is true for a covariance-stationary VAR(p) process, $\mathbf{\Phi}(B)\boldsymbol{y}_t = \boldsymbol{\varepsilon}_t$ (for simplicity set $\boldsymbol{c} = \mathbf{0}$). We could invert it to

$$\boldsymbol{y}_t = \mathbf{\Psi}(B)\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots$$

where

$$\mathbf{\Psi}(B) = \mathbf{\Phi}(B)^{-1}.$$

The coefficients of $\boldsymbol{\Psi}$ can be solved in the same way as in the scalar case, i.e., if $\boldsymbol{\Phi}^{-1}(B) = \boldsymbol{\Psi}(B)$, then $\boldsymbol{\Phi}(B)\boldsymbol{\Psi}(B) = \boldsymbol{I}_n$:

$$(\boldsymbol{I}_n - \boldsymbol{\Phi}_1 B - \boldsymbol{\Phi}_2 B - \cdots - \boldsymbol{\Phi}_p B^p)(\boldsymbol{I}_n + \boldsymbol{\Psi}_1 B + \boldsymbol{\Psi}_2 B^2 + \cdots) = \boldsymbol{I}_n.$$

Equating the coefficients of B^j, we have

$$\begin{aligned}
\boldsymbol{\Psi}_0 &= \boldsymbol{I}_n \\
\boldsymbol{\Psi}_1 &= \boldsymbol{\Phi}_1 \\
\boldsymbol{\Psi}_2 &= \boldsymbol{\Phi}_1\boldsymbol{\Psi}_1 + \boldsymbol{\Phi}_2 \\
&\cdots \\
\text{in general} \quad \boldsymbol{\Psi}_s &= \boldsymbol{\Phi}_1\boldsymbol{\Psi}_{s-1} + \boldsymbol{\Phi}_2\boldsymbol{\Psi}_{s-2} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{\Psi}_{s-p},
\end{aligned}$$

with $\boldsymbol{\Psi}_j = \mathbf{0}$ for $j < 0$.

3. Estimation and Model Specification

3.1 VAR Estimation

Consider the following n-dimensional VAR(p) model:

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}). \tag{19}$$

We could rewrite it more concisely as

$$\boldsymbol{y}_t = \boldsymbol{\Pi}' \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t,$$

where $\underset{n\times(np+1)}{\boldsymbol{\Pi}'} \equiv \begin{bmatrix} \boldsymbol{c} & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_p \end{bmatrix}$ and $\underset{(np+1)\times 1}{\boldsymbol{x}_t} \equiv \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}$.

Let the vector of parameters

$$\boldsymbol{\theta} \equiv (\boldsymbol{c} \quad \boldsymbol{\Phi}_1 \quad \boldsymbol{\Phi}_2 \quad \cdots \quad \boldsymbol{\Phi}_p \quad \boldsymbol{\Sigma}) .$$

The conditional log-likelihood function

$$f(\mathbf{y}_t, \boldsymbol{x}_t; \boldsymbol{\theta}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{y}_t - \boldsymbol{\Pi}' \boldsymbol{x}_t)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_t - \boldsymbol{\Pi}' \boldsymbol{x}_t) \right\},$$

Then the log-likelihood function is *(given that $y_0, \ldots, y_{1-p}$ is observed)*

$$\begin{aligned} \ln L(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \ln f(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{y}_{-p+1}; \boldsymbol{\theta}) \\ &= -\frac{Tn}{2} \ln(2\pi) - \frac{T}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{t=1}^{T} \left[(\boldsymbol{y}_t - \boldsymbol{\Pi}' \boldsymbol{x}_t)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_t - \boldsymbol{\Pi}' \boldsymbol{x}_t) \right] . \end{aligned}$$

Taking first derivative with respect to $\mathbf{\Pi}$ and $\mathbf{\Sigma}$, we have that

$$\widehat{\mathbf{\Pi}}' = \left(\sum_{t=1}^{T} \boldsymbol{y}_t \boldsymbol{x}_t'\right) \left(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1}.$$

Let $\widehat{\mathbf{\Pi}} = \begin{bmatrix} \hat{\boldsymbol{\pi}}_1 & \cdots & \hat{\boldsymbol{\pi}}_n \end{bmatrix}$ and so $\widehat{\mathbf{\Pi}}' = \begin{bmatrix} \hat{\boldsymbol{\pi}}_1' \\ \cdots \\ \hat{\boldsymbol{\pi}}_n' \end{bmatrix}$. The jth row of $\widehat{\mathbf{\Pi}}'$ is

$$\hat{\boldsymbol{\pi}}_j' = \left(\sum_{t=1}^{T} y_{jt} \boldsymbol{x}_t'\right) \left(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \Leftrightarrow \quad \hat{\boldsymbol{\pi}}_j = \left(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{x}_t y_{jt}\right),$$

which is the estimated coefficient vector from an OLS regression of y_{jt} on $\boldsymbol{x}_t = (1, y_{1,t-1}, \ldots, y_{n,t-1}, y_{1,t-2}, \ldots, y_{1,t-p}, \ldots, y_{n,t-p})'$.

The MLE estimate of $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12}^2 & \cdots & \hat{\sigma}_{1n}^2 \\ \hat{\sigma}_{21}^2 & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{n1}^2 & \hat{\sigma}_{n2}^2 & \cdots & \hat{\sigma}_n^2 \end{bmatrix},$$

where

$$\hat{\boldsymbol{\varepsilon}}_t' = \boldsymbol{y}_t - \widehat{\boldsymbol{\Pi}}' \boldsymbol{x}_t.$$

In addition, the (i,i)th element and the (i,j)th element of $\boldsymbol{\Sigma}$ are

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 \quad \text{and} \quad \hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt},$$

for $i, j = 1, 2, \ldots, n$.

Asymptotic distribution of $\widehat{\mathbf{\Pi}}$:

- MLE estimates of $\widehat{\mathbf{\Pi}}$ and $\widehat{\mathbf{\Sigma}}$ are consistent even if the true innovations are non-Gaussian.
- Standard errors for $\widehat{\mathbf{\Pi}}$ can be based on the usual OLS formulas.
- Standard OLS t and F statistics applied to the coefficients of any single equation of the VAR are asymptotically valid.
- $\widehat{\boldsymbol{\pi}}_i \overset{d}{\sim} \mathcal{N}\left(\boldsymbol{\pi}_i, \hat{\sigma}_i^2 \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right]^{-1}\right)$, where $\widehat{\boldsymbol{\pi}}_i$ is the $(np+1) \times 1$ vector of coefficients in the ith equation of the VAR.

- Alternatively, $\sqrt{T}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{Q}^{-1}\right)$, where

$$\underset{nk\times 1}{\widehat{\boldsymbol{\pi}}} = \operatorname{vec}\left(\underset{k\times n}{\widehat{\boldsymbol{\Pi}}}\right) = \begin{bmatrix} \widehat{\boldsymbol{\pi}}_1 \\ \widehat{\boldsymbol{\pi}}_2 \\ \vdots \\ \widehat{\boldsymbol{\pi}}_n \end{bmatrix}$$

and $k = np + 1$ (p lags plus a constant) and

$$\boldsymbol{Q} = E(\boldsymbol{x}_t \boldsymbol{x}_t').$$

Let

$$\boldsymbol{Q}_T = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'.$$

3.2 Testing Linear Constraints

A general linear hypotheses of the form $\boldsymbol{R} \cdot \mathrm{vec}(\boldsymbol{\Pi}) = \boldsymbol{r}$ involving coefficients across different equations of the VAR may be tested using the Wald statistic

$$Wald = \left(\boldsymbol{R} \cdot \mathrm{vec}(\widehat{\boldsymbol{\Pi}}) - \boldsymbol{r}\right)' \left[\boldsymbol{R}\left(\widehat{\boldsymbol{\Sigma}} \otimes \boldsymbol{Q}_T^{-1}\right)\boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R} \cdot \mathrm{vec}(\widehat{\boldsymbol{\Pi}}) - \boldsymbol{r}\right)$$

Under the null, this statistic has a limiting $\chi^2(q)$ distribution where $q = \mathrm{Rank}(\boldsymbol{R})$ gives the number of linear restricts, i.e.,

$$Wald \overset{H_0}{\sim} \chi^2(q).$$

The Wald test for Granger Causality is also know as the Wald-type Granger Causality test.

3.3 Lag Length Selection

The lag length for the VAR(p) model may be determined using model selection criteria. The general approach is fit VAR(p) models with orders $p = 0, 1, \ldots, p_{\max}$ and choose the value of p which minimizes some model selection criteria.

Model selection criteria for VAR(p) models have the form

$$IC(p) = \ln |\widehat{\mathbf{\Sigma}}(p)| + c_T \cdot \psi(n, p)$$

where $\widehat{\mathbf{\Sigma}}(p) = T^{-1} \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$ is the residual covariance matrix without a degrees of freedom correction from a VAR(p) model, c_T is a sequence indexed by the sample size T, and $\psi(n, p)$ is a penalty function which penalizes large VAR(p) models.

The three most common information criteria are the Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ):

$$AIC(p) = \ln|\widehat{\mathbf{\Sigma}}(p)| + \frac{2}{T}pn^2,$$

$$BIC(p) = \ln|\widehat{\mathbf{\Sigma}}(p)| + \frac{\ln(T)}{T}pn^2,$$

$$HQ(p) = \ln|\widehat{\mathbf{\Sigma}}(p)| + \frac{2\ln\ln(T)}{T}pn^2.$$

3.4 VAR Forecasting

The best linear predictor, in terms of minimum mean squared error (MSE), of $\boldsymbol{y}_{t+1}$ or 1-step forecast based on information available at time t is

$$\hat{\boldsymbol{y}}_{t+1|t} = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_t + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p+1}.$$

Forecasts for longer horizons h (h-step forecasts) may be obtained using the chain-rule of forecasting as

$$\hat{\boldsymbol{y}}_{t+h|t} = \boldsymbol{c} + \boldsymbol{\Phi}_1 \hat{\boldsymbol{y}}_{t+h-1|t} + \cdots + \boldsymbol{\Phi}_p \hat{\boldsymbol{y}}_{t+h-p|t},$$

where $\hat{\boldsymbol{y}}_{t+j|t} = \boldsymbol{y}_{t+j}$ if $j < 0$.

The h-step forecast errors may be expressed as

$$\hat{\boldsymbol{e}}_{t+h|t} = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t} = \sum_{s=0}^{h-1} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t+h-s}$$

where the matrices $\boldsymbol{\Psi}_s$ are determined by recursive substitution

$$\boldsymbol{\Psi}_s = \boldsymbol{\Phi}_1 \boldsymbol{\Psi}_{s-1} + \boldsymbol{\Phi}_2 \boldsymbol{\Psi}_{s-2} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{\Psi}_{s-p},$$

with $\boldsymbol{\Psi}_0 = \boldsymbol{I}_n$ and $\boldsymbol{\Psi}_j = \boldsymbol{0}$ for $j < 0$.

The forecasts are unbiased since all of the forecast errors have expectation zero and the MSE matrix for $\hat{\boldsymbol{y}}_{t+h|t}$ is

$$\boldsymbol{\Sigma}(h) = \mathsf{MSE}(\hat{\boldsymbol{e}}_{t+h|t}) = \sum_{s=0}^{h-1} \boldsymbol{\Psi}_s \boldsymbol{\Sigma} \boldsymbol{\Psi}_s'.$$

Asymptotic $(1-\alpha)\cdot 100\%$ confidence intervals for the individual elements of $\hat{\boldsymbol{y}}_{t+h|t}$ are then computed as

$$\left[\hat{y}_{k,t+h|t} - z_{\alpha/2}\hat{\sigma}_k(h), \hat{y}_{k,t+h|t} + z_{\alpha/2}\hat{\sigma}_k(h)\right],$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}_k(h)$ denotes the square root of the diagonal element of $\widehat{\boldsymbol{\Sigma}}(h)$.

4. Structural Analysis

4.1 Granger Causality (Granger, 1969)

- In most regressions in econometrics, it is very hard to discuss causality.

- For instance, the significance of the coefficient β in the regression

$$y_i = \beta x_i + \epsilon_i,$$

  only tell the 'co-occurrence（共生性）' of x and y, not that x cause y.

- In other words, ususally the regression only tell us there is some 'relationship' between x and y, and does not tell the nature of the relationship, such as whether x causes y or y causes x.

(1) Definition

Definition 1 (General definition)**.** X_t *is said not to Granger-cause* Y_t *if for all* $h > 0$

$$F(Y_{t+h} \mid \Omega_t) = F(y_{t+h} \mid \Omega_t - X_t)$$

where F *denotes the conditional distribution, and* $\Omega_t - X_t$ *is all the information in the universe except series* X_t*. In plain words,* X_t *is said not to Granger cause* Y_t *if* X *cannot help predict future* Y*.*

Remarks:

- The whole distribution F is generally difficult to handle empirically and we turn to conditional expectation and variance.

- It is defined for all $h > 0$ and not only for $h = 1$. Causality at different h does not imply each other. They are neither sufficient nor necessary.

Let

$$I_{1,t} = \{y_{1,t}, y_{1,t-1}, \ldots\}, \quad I_{2,t} = \{y_{2,t}, y_{2,t-1}, \ldots\},$$

and

$$I_t = \{I_{1,t}\} \cup \{I_{2,t}\}.$$

A redefined definition become as below:

Definition 2. $y_{1,t}$ *does not Granger cause* $y_{2,t+h}$ *with respect to information* $I_{2,t}$ *if*

$$E[y_{2,t+h}|I_{2,t}] = E[y_{2,t+h}|I_t].$$

We say that y_{1t} does not Granger cause y_{2t}, or $y_{1t} \nrightarrow y_{2t}$, if for all $h > 0$ we have $E[y_{2,t+h}|I_{2,t}] = E[y_{2,t+h}|I_t]$.

(2) Equivalent definition (model based)

For a n-dimension stationary process, $\boldsymbol{y}_t$, there is a canonical MA representation

$$\boldsymbol{y}_t = \mu + \Psi(B)\boldsymbol{\varepsilon}_t = \mu + \sum_{k=0}^{\infty} \Psi_k \boldsymbol{\varepsilon}_{t-k}, \quad \Psi_0 = I_n \tag{20}$$

$$\Psi_k = \begin{bmatrix} \Psi_{11}^{(k)} & \Psi_{12}^{(k)} & \cdots & \Psi_{1n}^{(k)} \\ \Psi_{21}^{(k)} & \Psi_{22}^{(k)} & \cdots & \Psi_{2n}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{n1}^{(k)} & \Psi_{n2}^{(k)} & \cdots & \Psi_{nn}^{(k)} \end{bmatrix}$$

Remark 1 (Impulse response (IR) analysis)**.** A necessary and sufficient condition for variable i not Granger-cause variable j is that $\Psi_{ji}^{(k)} = 0$, for $k = 1, 2, \ldots$.

If the process (20) is invertible, then

$$\boldsymbol{y}_t = \boldsymbol{c} + \Pi(B)\boldsymbol{y}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{c} + \sum_{k=1}^{\infty} \Pi_k \boldsymbol{y}_{t-k} + \boldsymbol{\varepsilon}_t \tag{21}$$

Remark 2 (VAR analysis)**.** If there are only two variables, or two-group of variables, j and i, then a necessary and sufficient condition for variable i not to Granger-cause variable j is that $\Pi_{ji}^{(k)} = 0$, for $k = 1, 2, \ldots$.

The condition is good for all forecast horizon h.

Remark 3. Note that for a VAR(1) process with dimension equal or greater than 3, $\Pi_{ji}^{(k)} = 0$, $k = 1$, is sufficient for non-causality at $h = 1$ but insufficient for $h > 1$.

Variable i might affect variable j in two or more period in the future via the effect through other variables.

For example,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-2} \\ y_{3t-3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}.$$

Then,

$$\boldsymbol{y}_0 = \begin{bmatrix} \varepsilon_{10} \\ \varepsilon_{20} \\ \varepsilon_{30} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \quad \boldsymbol{y}_1 = \Phi_1 \boldsymbol{y}_0 = \begin{bmatrix} .5 \\ .1 \\ 0 \end{bmatrix}; \quad \boldsymbol{y}_2 = \Phi_1^2 \boldsymbol{y}_0 = \begin{bmatrix} .25 \\ .06 \\ .02 \end{bmatrix}.$$

For y_1, the value of $y_{3,1}$ the condition $\phi_{30} = 0$ is sufficient, while for y_2 we see that $y_{3,2} \neq 0$.

To summarize,

1. For bivariate or two groups of variables, IR analysis is equivalent to applying Granger-causality test to VAR model;

2. For testing the impact of one variable on the other with a high dimensional (≥ 2) system, IR analysis can not be substituded by the Granger-causality test to the VAR model. The test has to be based upon IR.

(3) Granger Causal analysis for bivariate VAR

For a bivariate VAR(p) system, $\boldsymbol{y}_t = (y_{1t}, y_{2t})'$, defined by

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{22}$$

$$= \begin{bmatrix} \Psi_{11}(B) & \Psi_{12}(B) \\ \Psi_{21}(B) & \Psi_{22}(B) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{23}$$

where $\Phi_{12}(B) = \Phi_{12}^{(1)}B + \cdots + \Phi_{12}^{(p)}B^p$ and $\Psi_{12}(B) = \Psi_{12}^{(1)}B + \Psi_{12}^{(2)}B^2 + \cdots$

- y_{2t} does not Granger-cause y_{1t} if $\Psi_{12}(B) = 0$ or $\Psi_{12}^{(k)} = 0$, for $k = 1, 2, \ldots$.

- This condition is equivalent to $\Phi_{12}^{(k)} = 0$, for $k = 1, \ldots, p$.

Four possible causal directions (cases) between two variables (groups) are listed as follows:

Case 1: $y_{1t} \rightarrow y_{2t}$ but $y_{2t} \nrightarrow y_{1t}$. In this case, we have a one-way causality running from y_{1t} to y_{2t}.

Case 2: $y_{2t} \rightarrow y_{1t}$ but $y_{1t} \nrightarrow y_{2t}$. In this case, we have a one-way causality running from $y_{2,t}$ to y_{1t}.

Case 3: $y_{1t} \leftrightarrow y_{2t}$. Here we obtain a feedback between y_{1t} and y_{2t}.

Case 4: $y_{1t} \perp y_{1t}$. Here we obtain no causal relationship between y_{1t} and y_{2t}.

Granger causality (GC) test

Consider the following bivariate VAR(p) process:

$$y_{1t} = \phi_{10} + \sum_{i=1}^{p} \phi_{11}^{(i)} y_{1,t-i} + \sum_{i=1}^{p} \phi_{12}^{(i)} y_{2,t-i} + \varepsilon_{1t}$$

$$y_{2t} = \phi_{20} + \sum_{i=1}^{p} \phi_{21}^{(i)} y_{1,t-i} + \sum_{i=1}^{p} \phi_{22}^{(i)} y_{2,t-i} + \varepsilon_{2t}$$

where ε_{it} $(i = 1, 2)$ is a disturbance term.

To analyze whether $y_{1t} \to y_{2t}$, or y_{1t} Granger-causes y_{2t}, we carry out the following testing:

$$H_0 : \phi_{21}^{(1)} = \phi_{21}^{(2)} = \cdots = \phi_{21}^{(p)} = 0, \tag{24}$$

$$H_A : \phi_{21}^{(1)} \neq 0 \quad \text{or} \quad \phi_{21}^{(2)} \neq 0 \quad \text{or} \quad \cdots \quad \phi_{21}^{(p)} \neq 0.$$

- **F-test:** This can be tested using the F test or asymptotic chi-square test. F-statistic is shown as follows:

$$S_1 = \frac{(RSS - USS)/p}{USS/(T - 2p - 1)} \sim F(p, T - 2p - 1),$$

where T is the sample size, RSS is the restricted residual sum of Squares and USS is the unrestricted residual sum of Squares.

It is also shown that $pF \xrightarrow{d} \chi^2(p)$. So we have an asymmetrically equivalent test given by

$$S_2 = \frac{T(RSS - USS)}{USS} \sim \chi^2(p).$$

- **Wald test:** The Granger causal hypothesis given by (24) can be easily tested using the Wald statistic,

$$Wald = \left(\boldsymbol{R} \cdot \text{vec}(\widehat{\boldsymbol{\Pi}}) - \boldsymbol{r}\right)' \left[\boldsymbol{R}\left(\widehat{\boldsymbol{\Sigma}} \otimes \boldsymbol{Q}_T^{-1}\right)\boldsymbol{R}'\right]^{-1} \left(\boldsymbol{R} \cdot \text{vec}(\widehat{\boldsymbol{\Pi}}) - \boldsymbol{r}\right),$$

where $\boldsymbol{R}$ for the hypothesis that y_{1t} does not Granger cause y_{2t} is

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{\pi}_1' & \phi_{20} & \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{21}^{(2)} & \phi_{22}^{(2)} & & \phi_{21}^{(p)} & \phi_{22}^{(p)} \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{1\times(n*p+1)} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

and $\boldsymbol{r} = \mathbf{0}_{p\times 1}$.

4.2 Impulse Response Function

A VAR(p) model can be written as a linear function of the past innovations (Wold representation), that is,

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots, \tag{25}$$

where $\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$ and $\boldsymbol{\Psi}(B) = \boldsymbol{\Phi}(B)^{-1}$.

- The coefficient matrix $\boldsymbol{\Psi}_i$ is the impact of the past innovation $\boldsymbol{\varepsilon}_{t-i}$ on $\boldsymbol{y}_t$, or equivalently, $\boldsymbol{\Psi}_i$ is the effect of $\boldsymbol{\varepsilon}_t$ on the future observation $\boldsymbol{y}_{t+i}$.
- $\boldsymbol{\Psi}_i$ is often referred to as the *impulse response function* of $\boldsymbol{y}_t$ on $\boldsymbol{\varepsilon}_t$.
- Since the components of $\boldsymbol{\varepsilon}_t$ are often correlated, the interpretation of elements in $\boldsymbol{\Psi}_i$ of Eq. (25) is not straightforward.

To aid interpretation, one can use the *Cholesky decomposition* to transform the innovations so that the resulting components are *uncorrelated*.

- Let $\boldsymbol{\eta}_t = Q^{-1}\boldsymbol{\varepsilon}_t$, then $\mathrm{Cov}(\boldsymbol{\eta}_t) = I$, so that the elements η_{it} and η_{jt} are uncorrelated. Rewrite Eq. (25) as

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\mu} + QQ^{-1}\boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 QQ^{-1}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 QQ^{-1}\boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \boldsymbol{\Psi}_0^*\boldsymbol{\eta}_t + \boldsymbol{\Psi}_1^*\boldsymbol{\eta}_{t-1} + \boldsymbol{\Psi}_2^*\boldsymbol{\eta}_{t-2} + \cdots, \end{aligned} \tag{26}$$

  where $\boldsymbol{\Psi}_0^* = Q$ and $\boldsymbol{\Psi}_i^* = \boldsymbol{\Psi}_i Q$.

  - The coefficient matrices $\boldsymbol{\Psi}_i^*$ are called the impulse response function of $\boldsymbol{y}_t$ with orthogonal innovations $\boldsymbol{\eta}_t$.
  - Specifically, the (i, j)-th element of $\boldsymbol{\Psi}_\ell^*$ is the impact of $\eta_{j,t}$ on the

future observation $y_{i,t+\ell}$, that is,

$$\frac{\partial y_{i,t+\ell}}{\partial \eta_{j,t}} = \frac{\partial y_{i,t}}{\partial \eta_{j,t-\ell}} = \psi^*_{ij}(\ell).$$

– A plot of $\psi^*_{ij}(\ell)$ against ℓ is called the *one s.t.d. orthogonal impulse response function* of y_i with respect to η_j. With n variables there are n^2 possible impulse response functions.

- If using $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{L}'$ and let $\boldsymbol{\eta}_t = \boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_t$, then $\mathrm{Cov}(\boldsymbol{\eta}_t) = \boldsymbol{G}$. We have

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \boldsymbol{\Psi}^*_0\boldsymbol{\eta}_t + \boldsymbol{\Psi}^*_1\boldsymbol{\eta}_{t-1} + \boldsymbol{\Psi}^*_2\boldsymbol{\eta}_{t-2} + \cdots,\end{aligned}$$

Similary, a plot of $\psi^*_{ij}(\ell)$ against ℓ is called the *one unit impulse response function*.

Example 5. Consider a VAR(1) process of a 2-dimensional vector,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$\boldsymbol{\Sigma} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}.$$

First we verify that this process is stationary, as

$$\left| \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} \right| = 0$$

gives $\lambda_1 = 0.94$ and $\lambda_2 = -0.04$, both lies inside the unit circle. Invert it to a moving average process,

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t.$$

We know that $\boldsymbol{\Psi}_0 = I_2$, $\boldsymbol{\Psi}_1 = \boldsymbol{\Phi}_1$, $\boldsymbol{\Psi}_2 = \boldsymbol{\Phi}_1^2$, etc.

Then we find Q by Choleski decomposition of $\boldsymbol{\Sigma}$, which gives

$$Q = \begin{bmatrix} 1.41 & 0 \\ 0.70 & 1.87 \end{bmatrix} \quad \text{and} \quad Q^{-1} = \begin{bmatrix} 0.70 & 0 \\ -0.27 & 0.53 \end{bmatrix}.$$

The we can write

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)QQ^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{\Psi}(L)Q\boldsymbol{\eta}_t,$$

where we define that $\boldsymbol{\eta}_t = Q^{-1}\boldsymbol{\varepsilon}_t$. Then we have

$$\boldsymbol{y}_t = \boldsymbol{\Psi}_0 Q\boldsymbol{\eta}_t + \boldsymbol{\Psi}_1 Q\boldsymbol{\eta}_{t-1} + \cdots,$$

or

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.41 & 0 \\ 0.70 & 1.87 \end{bmatrix} \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix} + \begin{bmatrix} 0.85 & 0.37 \\ 0.70 & 0.75 \end{bmatrix} \begin{bmatrix} \eta_{1,t-1} \\ \eta_{2,t-1} \end{bmatrix} + \cdots.$$

4.3 Forecast Error Variance Decomposition (方差贡献)

To answer question: what portion of the variance of the forecast error in predicting $y_{i,t+h}$ is due to the structural shock ε_{jt} (or equivalently η_{jt})?

Using the orthogonal shocks $\boldsymbol{\eta}_t = Q^{-1}\boldsymbol{\varepsilon}_t$, the h-step ahead forecast error vector

$$\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h|t} = \sum_{s=0}^{h-1} \boldsymbol{\Psi}_s^* \boldsymbol{\eta}_{t+h-s}.$$

For a particular variable $y_{i,t+h}$, this forecast error has the form

$$\begin{aligned} y_{i,t+h} - \hat{y}_{i,t+h|t} &= \sum_{s=0}^{h-1}\sum_{j=1}^{n} \psi_{ij}^s \eta_{j,t+h-s} = \sum_{j=1}^{n}\sum_{s=0}^{h-1} \psi_{ij}^s \eta_{j,t+h-s} \\ &= \sum_{s=0}^{h-1} \psi_{i1}^s \eta_{1,t+h-s} + \cdots + \sum_{s=0}^{h-1} \psi_{in}^s \eta_{n,t+h-s}. \end{aligned}$$

Since the structural errors are orthogonal, the variance of the h-step forecast error is

$$\mathrm{Var}(y_{i,t+h} - \hat{y}_{i,t+h|t}) = \sigma_1^2 \sum_{s=0}^{h-1} (\psi_{i1}^s)^2 + \cdots + \sigma_n^2 \sum_{s=0}^{h-1} (\psi_{in}^s)^2,$$

where $\sigma_j^2 = \mathrm{Var}(\eta_{jt}) \equiv 1$.

The portion of $\mathrm{Var}(y_{i,t+h} - \hat{y}_{i,t+h|t})$ due to shock ε_j or η_j is then

$$FEVD_{i,j}(h) = \frac{\sigma_j^2 \sum_{s=0}^{h-1} (\psi_{ij}^s)^2}{\sigma_1^2 \sum_{s=0}^{h-1} (\psi_{i1}^s)^2 + \cdots + \sigma_n^2 \sum_{s=0}^{h-1} (\psi_{in}^s)^2}, \quad i, j = 1, \ldots, n.$$

In a VAR with n variables there will be n^2 $FEVD_{i,j}(h)$ values.

4.4 Generalized impulse responses

Reconsider the VAR(p) process

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, \Omega), \\ &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \boldsymbol{\varepsilon}_{t-s},\end{aligned}$$

with $\boldsymbol{\Psi}_i = \boldsymbol{\Phi}_1 \boldsymbol{\Psi}_{i-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{\Psi}_{i-p}$, $\boldsymbol{\Psi}_0 = I_N$ and $\boldsymbol{\Psi}_i = 0$ for $i < 0$.

Following Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998), the generalized impulse response function (GIRF) is given by

$$GI_y(h, \delta, I_{t-1}) = E[\boldsymbol{y}_{t+h} | \boldsymbol{\varepsilon}_t = \delta, I_{t-1}] - E[\boldsymbol{y}_{t+h} | I_{t-1}] = \boldsymbol{\Psi}_h \delta,$$

where δ is a choice vector, which is central for any GIRF.

As an alternative to shocking all elements of $\boldsymbol{\varepsilon}_t$ one may consider just shocking one element such that $\varepsilon_{jt} = \delta_j$. Then, the GIRF is defined as

$$GI_y(h, \delta_j, I_{t-1}) = E[\boldsymbol{y}_{t+h}|\varepsilon_{jt} = \delta_j, I_{t-1}] - E[\boldsymbol{y}_{t+h}|I_{t-1}] = \boldsymbol{\Psi}_h E[\boldsymbol{\varepsilon}_t|\varepsilon_{jt} = \delta_j].$$

Letting $\delta_j = \sqrt{\omega_{jj}}$, the standard deviation of ε_{jt}, it follows that:

$$E[\boldsymbol{\varepsilon}_t|\varepsilon_{jt} = \sqrt{\omega_{jj}}] = \Omega e_j \omega_{jj}^{-1/2},$$

where Ω is the covariance matrix of $\boldsymbol{\varepsilon}_t$ and e_j is the j-th column of I_N.

Then, for the VAR model we can find that

$$GI_y(h, \delta_j, I_{t-1}) = \boldsymbol{\Psi}_h \Omega e_j \omega_{jj}^{-1/2},$$

which measures the response in $\boldsymbol{y}_{t+h}$ from a one standard deviation shock to ε_{jt}, where the correlation between ε_{jt} and ε_{it} is considered.

Defining the diagonal $N \times N$ matrix Σ as:

$$\Sigma = \text{diag} \begin{bmatrix} (e_1' \Omega e_1)^{-1/2} \\ (e_2' \Omega e_2)^{-1/2} \\ \vdots \\ (e_N' \Omega e_N)^{-1/2} \end{bmatrix} = \text{diag}\{\omega_{11}^{-1/2}, \ldots, \omega_{NN}^{-1/2}\}$$

we may express the generalized impulse responses in matrix form as:

$$GI_y(h, \sqrt{\omega_{11}}, \ldots, \sqrt{\omega_{NN}}, I_{t-1}) = \mathbf{\Psi}_h \Omega \Sigma,$$

where column j is given by $GI_y(h, \sqrt{\omega_{jj}}, I_{t-1})$.

4.5 Connectedness

Diebold and Yilmaz (2012), "Better to give than to receive: Predictive directional measurement of volatility spillovers", *International Journal of Forecasting, 28: 57-66*.

Variable j's contribution to variable i's H-step-ahead generalized forecast error variance, $\theta_{ij}(H)$, for $H = 1, 2, \ldots$, is defined as:

$$\theta_{ij}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' \Psi_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e_i' \Psi_h \Sigma \Psi_h' e_i)}$$

- The sum of the elements of each row of the variance decomposition table is not necessarily equal to 1: $\sum_{j=1}^{N} \theta_{ij}(H) \neq 1$.
- Normalize each entry of the GVD matrix by: $d_{ij}^H = \frac{\theta_{ij}(H)}{\sum_{j=1}^{N} \theta_{ij}(H)}$.

- Now by construction $\sum_{j=1}^{N} d_{ij}^{H} = 1$, and $\sum_{i,j=1}^{N} d_{ij}^{H} = N$.

	x_1	x_2	...	x_N	From Others
x_1	d_{11}^H	d_{12}^H	$\cdots$	d_{1N}^H	$\Sigma_{j=1}^{N} d_{1j}^H, j \neq 1$
x_2	d_{21}^H	d_{22}^H	$\cdots$	d_{2N}^H	$\Sigma_{j=1}^{N} d_{2j}^H, j \neq 2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_N	d_{N1}^H	d_{N2}^H	$\cdots$	d_{NN}^H	$\Sigma_{j=1}^{N} d_{Nj}^H, j \neq N$
To Others	$\Sigma_{i=1}^{N} d_{i1}^H$ $i \neq 1$	$\Sigma_{i=1}^{N} d_{i2}^H$ $i \neq 2$	$\cdots$	$\Sigma_{i=1}^{N} d_{iN}^H$ $i \neq N$	$\frac{1}{N}\Sigma_{i,j=1}^{N} d_{ij}^H$ $i \neq j$

Table 1: Connectedness Table Schematic. See Text for details.

- *Total connectedness:* Using the normalized entries of the GVD matrix, it is constructed as:

$$C(H) = \frac{\sum_{i,j=1, i\neq j}^{N} d_{ij}^{H}}{\sum_{i,j=1}^{N} d_{ij}^{H}} \times 100 = \frac{\sum_{i,j=1, i\neq j}^{N} d_{ij}^{H}}{N} \times 100$$

- *Directional connectedness:*

  – The gross directional connectedness received by i from all other j is measured as

$$C_{i\leftarrow\bullet} = \frac{\sum_{j=1, j\neq i}^{N} d_{ij}^{H}}{\sum_{i,j=1}^{N} d_{ij}^{H}} \times 100 = \frac{\sum_{j=1, j\neq i}^{N} d_{ij}^{H}}{N} \times 100$$

  – In a similar fashion, the directional connectedness transmitted by i

to all other j is measured as:

$$C_{\bullet \leftarrow i} = \frac{\sum_{j=1, j\neq i}^{N} d_{ji}^{H}}{\sum_{i,j=1}^{N} d_{ij}^{H}} \times 100 = \frac{\sum_{j=1, j\neq i}^{N} d_{ji}^{H}}{N} \times 100$$

- *Net spillovers:* We obtain the net spillover from market i to all other markets j as

$$C_i(H) = C_{\bullet \leftarrow i} - C_{i \leftarrow \bullet}$$

- *Net pairwise spillovers:* The net pairwise spillover between markets i and j is defined as

$$\begin{aligned} C_{ij}(H) &= \left(\frac{d_{ji}^{H}}{\sum_{i,k=1}^{N} d_{ik}^{H}} - \frac{d_{ij}^{H}}{\sum_{j,k=1}^{N} d_{jk}^{H}} \right) \times 100 \\ &= \left(\frac{d_{ji}^{H} - d_{ij}^{H}}{N} \right) \times 100 \end{aligned}$$

5 Structural VAR Models

5.1 Structural Representation

Consider the following Structural VAR model

$$\begin{aligned} y_{1t} &= \gamma_{10} - b_{12} y_{2t} + \gamma_{11} y_{1t-1} + \gamma_{12} y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \gamma_{20} - b_{21} y_{1t} + \gamma_{21} y_{1t-1} + \gamma_{22} y_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{27}$$

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim iidN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right).$$

Remark 4.

1. ε_{1t} and ε_{2t} are called *structural errors*
2. In general, $\mathrm{Cov}(y_{2t}, \varepsilon_{1t}) \neq 0$ and $\mathrm{Cov}(y_{1t}, \varepsilon_{2t}) \neq 0$
3. All variables are *endogenous* – OLS is not appropriate!

In matrix form, the model becomes

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

or

$$\boldsymbol{B}\boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

$$\mathrm{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{D} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

In *lag operator* notation, the SVAR is

$$\boldsymbol{B}(L)\boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{B}(L) = \boldsymbol{B} - \boldsymbol{\Gamma}_1 L$.

Reduced Form Representation

Solve for $\boldsymbol{y}_t$ in terms of $\boldsymbol{y}_{t-1}$ and $\boldsymbol{\varepsilon}_t$:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{B}^{-1}\boldsymbol{\gamma}_0 + \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_1\boldsymbol{y}_{t-1} + \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t \\ &= \boldsymbol{a}_0 + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \end{aligned}$$

where $\boldsymbol{a}_0 = \boldsymbol{B}^{-1}\boldsymbol{\gamma}_0$, $\boldsymbol{A}_1 = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_1$, and $\boldsymbol{u}_t = \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$, or

$$\boldsymbol{A}(L)\boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{u}_t$$

where $\boldsymbol{A}(L) = \boldsymbol{I} - \boldsymbol{A}_1 L$.

Given that

$$\boldsymbol{B}^{-1} = \frac{1}{\Delta}\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}, \quad \Delta = \det(\boldsymbol{B}) = 1 - b_{12}b_{21},$$

we have

$$\boldsymbol{a}_0 = \boldsymbol{B}^{-1}\boldsymbol{\gamma}_0 = \frac{1}{\Delta}\begin{bmatrix} \gamma_{10} - b_{12}\gamma_{20} \\ \gamma_{20} - b_{21}\gamma_{10} \end{bmatrix},$$

$$\boldsymbol{A}_1 = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_1 = \frac{1}{\Delta}\begin{bmatrix} \gamma_{11} - b_{12}\gamma_{21} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} - b_{21}\gamma_{11} & \gamma_{22} - b_{21}\gamma_{12} \end{bmatrix},$$

$$\boldsymbol{u}_t = \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t = \frac{1}{\Delta}\begin{bmatrix} \varepsilon_{1t} - b_{12}\varepsilon_{2t} \\ \varepsilon_{2t} - b_{21}\varepsilon_{1t} \end{bmatrix}.$$

The reduced form errors $\boldsymbol{u}_t$ are *linear combinations* of the structural errors $\boldsymbol{\varepsilon}_t$ and have covariance matrix

$$E[\boldsymbol{u}_t\boldsymbol{u}_t'] = \boldsymbol{B}^{-1}E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t']\boldsymbol{B}^{-1'} = \boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1'} = \boldsymbol{\Sigma}.$$

Specifically, the elements in $\boldsymbol{\Sigma}$ are

$$\begin{bmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{bmatrix} = \frac{1}{\Delta^2}\begin{bmatrix} \sigma_1^2 + b_{12}^2\sigma_2^2 & -(b_{21}\sigma_1^2 + b_{12}\sigma_2^2) \\ -(b_{21}\sigma_1^2 + b_{12}\sigma_2^2) & \sigma_2^2 + b_{21}^2\sigma_1^2 \end{bmatrix}.$$

Note that $\boldsymbol{\Sigma}$ is diagonal only if $b_{12} = b_{21} = 0$.

Identification Issues

Without some restrictions, the parameters in the SVAR are *not identified*. That is, given values of the *reduced form parameters* $\boldsymbol{a}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{\Sigma}$, it is not possible to uniquely solve for the *structural parameters* $\boldsymbol{B}$, $\boldsymbol{\gamma}_0$, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{D}$.

- 10 structural parameters and only 9 reduced form parameters
- Order condition requires *at least 1 restriction* on the SVAR parameters

Typical identifying restrictions include

- *Zero (exclusion) restrictions* on the elements of $\boldsymbol{B}$, e.g., $b_{12} = 0$.
- *Linear restrictions* on the elements of $\boldsymbol{B}$, e.g., $b_{12} + b_{21} = 1$.

5.2 MA Representations

The moving average (MA) of Wold representation of the reduced form VAR is found by multiplying both sides by $\boldsymbol{A}(L) = (I_2 - \boldsymbol{A}_1 L)^{-1}$ to give

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{u}_t,$$

where

$$\boldsymbol{\Psi}(L) = (I_2 - \boldsymbol{A}_1 L)^{-1} = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k L^k, \quad \boldsymbol{\Psi}_0 = I_2, \quad \boldsymbol{\Psi}_k = \boldsymbol{A}_1^k, \quad \boldsymbol{\mu} = (I_2 - \boldsymbol{A}_1)^{-1}\boldsymbol{a}_0.$$

In the Wold representation for $\boldsymbol{y}_t$, the first matrix in the moving average polynomial $\boldsymbol{\Psi}(L)$ is $\boldsymbol{\Psi}_0 = I_2$. In addition, the error terms $\boldsymbol{u}_t$ are generally contemporaneously correlated and have covariance matrix $\boldsymbol{\Sigma}$.

The structural moving average (SMA) representation of $\boldsymbol{y}_t$ is based on an infinite moving average of the structural innovations $\boldsymbol{\varepsilon}_t$. Substituting $\boldsymbol{u}_t = \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$ gives

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t, \tag{28}$$

where

$$\boldsymbol{\Theta}(L) = \sum_{k=0}^{\infty} \boldsymbol{\Theta}_k L^k = \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1} + \boldsymbol{\Psi}_1 \boldsymbol{B}^{-1} L + \cdots$$

That is, $\boldsymbol{\Theta}_k = \boldsymbol{\Psi}_k \boldsymbol{B}^{-1}$ for $k = 0, 1, \ldots$. In particular, notice that $\boldsymbol{\Theta}_0 = \boldsymbol{B}^{-1} \neq I_2$, and $\boldsymbol{\Theta}(L) = \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1} = (I_2 - \boldsymbol{A}_1 L)^{-1}\boldsymbol{B}^{-1}$.

It is instructive to look at the SMA representation for the bivariate system

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \theta_{11}^{(0)} & \theta_{12}^{(0)} \\ \theta_{21}^{(0)} & \theta_{22}^{(0)} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{bmatrix} + \cdots$$

which illustrates that the elements of the $\mathbf{\Theta}_k$ matrices, $\theta_{ij}^{(k)}$, give that dynamic multipliers or impulse responses of y_{1t} and y_{2t} to change in ε_{1t} and ε_{2t}.

Impulse Response Function

Consider the SMA representation (28) at time $t+s$

$$\begin{bmatrix} y_{1t+s} \\ y_{2t+s} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \theta_{11}^{(0)} & \theta_{12}^{(0)} \\ \theta_{21}^{(0)} & \theta_{22}^{(0)} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t+s} \\ \varepsilon_{2t+s} \end{bmatrix} + \cdots + \begin{bmatrix} \theta_{11}^{(s)} & \theta_{12}^{(s)} \\ \theta_{21}^{(s)} & \theta_{22}^{(s)} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \cdots .$$

The structural dynamic multipliers are

$$\begin{aligned} \frac{\partial y_{1t+s}}{\partial \varepsilon_{1t}} &= \theta_{11}^{(s)}, \quad \frac{\partial y_{1t+s}}{\partial \varepsilon_{2t}} = \theta_{12}^{(s)} \\ \frac{\partial y_{2t+s}}{\partial \varepsilon_{1t}} &= \theta_{21}^{(s)}, \quad \frac{\partial y_{2t+s}}{\partial \varepsilon_{2t}} = \theta_{22}^{(s)}. \end{aligned}$$

The structural impluse response functions (IRFs) are the plots of $\theta_{ij}^{(s)}$ vs. s for $i, j = 1, 2$.

Since y_t is assumed to be covariance stationary, we know that

$$\lim_{s\to\infty} \theta_{ij}^{(s)} = 0, \quad i,j = 1,2,$$

so that no structural shock has a long-run impact on the level of y. The long-run cumulative impact of the structural shocks is captured by the long-run impact matrix

$$\mathbf{\Theta}(1) = \begin{bmatrix} \theta_{11}(1) & \theta_{12}(1) \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix} = \begin{bmatrix} \sum_{s=0}^{\infty} \theta_{11}^{(s)} & \sum_{s=0}^{\infty} \theta_{12}^{(s)} \\ \sum_{s=0}^{\infty} \theta_{21}^{(s)} & \sum_{s=0}^{\infty} \theta_{22}^{(s)} \end{bmatrix}$$

where

$$\mathbf{\Theta}(L) = \begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix} = \begin{bmatrix} \sum_{s=0}^{\infty} \theta_{11}^{(s)} L^s & \sum_{s=0}^{\infty} \theta_{12}^{(s)} L^s \\ \sum_{s=0}^{\infty} \theta_{21}^{(s)} L^s & \sum_{s=0}^{\infty} \theta_{22}^{(s)} L^s \end{bmatrix}.$$

Identification issues

In some applications, identification of the parameters of the SVAR is achieved *through restrictions on the parameters of the SMA representation*.

For example, suppose that ε_{2t} has no contemporaneous impact on y_{1t}. Then $\theta_{12}^{(0)} = 0$ and so $\mathbf{\Theta}_0$ becomes triangular

$$\mathbf{\Theta}_0 = \begin{bmatrix} \theta_{11}^{(0)} & 0 \\ \theta_{21}^{(0)} & \theta_{22}^{(0)} \end{bmatrix}.$$

Since $\mathbf{\Theta}_0 = \boldsymbol{B}^{-1}$ we then have

$$\begin{bmatrix} \theta_{11}^{(0)} & 0 \\ \theta_{21}^{(0)} & \theta_{22}^{(0)} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$$

which implies that $b_{12} = 0$. Hence, assuming $\theta_{12}^{(0)} = 0$ in the SMA representation is equivalent to assuming $b_{12} = 0$ in the SVAR representation (27).

As another example, suppose ε_{2t} has no long-run cumulative impact on y_{1t}. Then $\theta_{12}(1) = \sum_{s=0}^{\infty} \theta_{12}^{(s)} = 0$ and the long-run impact matrix $\boldsymbol{\Theta}(1)$ becomes triangular:

$$\boldsymbol{\Theta}(1) = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}.$$

This type of long-run restriction places restrictions on the coefficients of the SVAR (27) since

$$\boldsymbol{\Theta}(1) = \boldsymbol{\Psi}(1)\boldsymbol{B}^{-1} = (\boldsymbol{I}_2 - \boldsymbol{A}_1)^{-1}\boldsymbol{B}^{-1}.$$

Estimation issues

In order to compute the structural IRFs, the parameters of the SMA representation (28) need to be estimated. Since $\boldsymbol{\Theta}(L) = \boldsymbol{\Psi}(L)\boldsymbol{B}^{-1}$ and $\boldsymbol{\Psi}(L) = \boldsymbol{A}(L)^{-1} = (\boldsymbol{I}_2 - \boldsymbol{A}_1 L)^{-1}$ the estimation of the elements in $\boldsymbol{\Theta}(L)$ can often be broken down into two steps.

First, $\boldsymbol{A}_1$ is estimated from the reduced for VAR model. Given $\widehat{\boldsymbol{A}}_1$, the matrices in $\boldsymbol{\Psi}(L)$ can be estimated using $\widehat{\boldsymbol{\Psi}}_k = \widehat{\boldsymbol{A}}_1^k$.

Second, $\boldsymbol{B}$ is estimated from the SVAR (27). Given $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\Psi}}_k$ the estimates of $\boldsymbol{\Theta}_k$, $k = 0, 1, \ldots$, are given by $\widehat{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Psi}}_k \widehat{\boldsymbol{B}}^{-1}$.

5.3 Identification Using Recursive Causal Ordering **

Consider the bivariate SVAR(1). From the previous discussion, we know that we need at least one restriction on the parameters of (27) for identification. Suppose $b_{12} = 0$,

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B}^{-1} = \boldsymbol{\Theta}_0 = \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix}.$$

- This assumption imposes the restriction that y_{2t} does not have a contemporaneous effect on y_{1t}.
- Since $b_{21} \neq 0$ a priori we allow for the possibility that y_{1t} has a contemporaneous effect on y_{2t}.
- Under this assumption the reduced form VAR errors $\boldsymbol{u}_t = \boldsymbol{B}^{-1}\boldsymbol{\varepsilon}_t$ become

$$\boldsymbol{u}_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} - b_{21}\varepsilon_{1t} \end{bmatrix}.$$

- The restriction $b_{12} = 0$ is sufficient to just identify b_{21} and, hence, just identify $\boldsymbol{B}$.

To establish this result, we show how b_{21} can be uniquely identified from the elements of the reduced form covariance matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{\Sigma} = \boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{B}^{-1'}$ and $\boldsymbol{B}$ is lower triangular, we have

$$\begin{bmatrix} w_1^2 & w_{12} \\ w_{12} & w_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & -b_{21} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & -b_{21}\sigma_1^2 \\ -b_{21}\sigma_1^2 & \sigma_2^2 + b_{21}^2\sigma_1^2 \end{bmatrix}.$$

Then, we can solve for b_{21} via

$$b_{21} = -\frac{-b_{21}\sigma_1^2}{\sigma_1^2} = -\frac{w_{12}}{w_1^2} := -\frac{\rho w_1 w_2}{w_1^2} = -\rho\frac{w_2}{w_1},$$

where $\rho = w_{12}/w_1 w_2$ is the correlation between u_1 and u_2. Notice that $b_{21} \neq 0$ provided $\rho \neq 0$.

Estimation Procedure

The SMA representation of the SVAR based on a recursive causal ordering may be estimated using the following procedure:

- Estimate the reduced form VAR by OLS equation by equation:

$$\boldsymbol{y}_t = \hat{\boldsymbol{a}}_0 + \widehat{\boldsymbol{A}}_1 \boldsymbol{y}_{t-1} + \hat{\boldsymbol{u}}_t$$

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_t'$$

- Estimate b_{21} and $\boldsymbol{B}$ from $\widehat{\boldsymbol{\Sigma}}$:

$$\hat{b}_{21} = -\frac{\hat{w}_{12}}{\hat{w}_1^2}, \quad \widehat{\boldsymbol{B}} = \begin{bmatrix} 1 & 0 \\ \hat{b}_{21} & 1 \end{bmatrix}.$$

- Estimate SMA from estimates of $\boldsymbol{a}_0$, $\boldsymbol{A}_1$ and $\boldsymbol{B}$:

$$\begin{aligned}
\boldsymbol{y}_t &= \hat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\Theta}}(L)\hat{\boldsymbol{\varepsilon}}_t \\
\hat{\boldsymbol{\mu}} &= \hat{\boldsymbol{a}}_0(\boldsymbol{I}_2 - \widehat{\boldsymbol{A}}_1)^{-1} \\
\widehat{\boldsymbol{\Theta}}_k &= \widehat{\boldsymbol{A}}_1^k \widehat{\boldsymbol{B}}^{-1}, \qquad k = 0, 1, \ldots \\
\widehat{\boldsymbol{D}} &= \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{B}}'.
\end{aligned}$$

Recovering the SMA representation using the Choleski Factorization

The SVAR representation based on a recursive causal ordering may be computed using the Choleski factorization of the reduced form covariance matrix $\boldsymbol{\Sigma}$. Recall, the Choleski factorization of the positive semi-definite matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{P}',$$

where

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix}$$

is a lower triangular matrix with $p_{ii} \geq 0$, $i = 1, 2$. A closely related factorization obtained from the Choleski factorization is the triangular factorization

$$\boldsymbol{\Sigma} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}', \tag{29}$$

where $\boldsymbol{T}$ is a lower triangular matrix with 1's along the diagonal and $\boldsymbol{\Lambda}$ is a diagonal matrix with non-negative elements:

$$\boldsymbol{T} = \begin{bmatrix} 1 & 0 \\ t_{21} & 1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \lambda_i \geq 0, \quad i = 1, 2.$$

Consider the reduced form VAR

$$\boldsymbol{y}_t = \boldsymbol{a}_0 + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \quad \boldsymbol{\Sigma} = E[\boldsymbol{u}_t \boldsymbol{u}_t'], \tag{30}$$

and perform the triangular factorization (29) on the covariance matrix $\boldsymbol{\Sigma}$. Now construct a pseudo SVAR model by premultiplying (30) by $\boldsymbol{T}^{-1}$:

$$\boldsymbol{T}^{-1}\boldsymbol{y}_t = \boldsymbol{T}^{-1}\boldsymbol{a}_0 + \boldsymbol{T}^{-1}\boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{T}^{-1}\boldsymbol{u}_t$$

or

$$\boldsymbol{B}\boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

where

$$\boldsymbol{B} = \boldsymbol{T}^{-1}, \quad \boldsymbol{\gamma}_0 = \boldsymbol{T}^{-1}\boldsymbol{a}_0, \quad \boldsymbol{\Gamma}_1 = \boldsymbol{T}^{-1}\boldsymbol{A}_1, \quad \boldsymbol{\varepsilon}_t = \boldsymbol{T}^{-1}\boldsymbol{u}_t.$$

Notice that the pseudo structural errors $\boldsymbol{\varepsilon}_t$ have a diagonal covariance matrix $\boldsymbol{\Lambda}$ since

$$\begin{aligned} E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] &= \boldsymbol{T}^{-1} E[\boldsymbol{u}_t \boldsymbol{u}_t'] \boldsymbol{T}^{-1'} \\ &= \boldsymbol{T}^{-1} \boldsymbol{\Sigma} \boldsymbol{T}^{-1'} \\ &= \boldsymbol{T}^{-1} \boldsymbol{T} \boldsymbol{\Lambda} \boldsymbol{T}' \boldsymbol{T}^{-1'} \\ &= \boldsymbol{\Lambda}. \end{aligned}$$

In the pseudo SVAR (30),

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix} = \boldsymbol{T}^{-1} = \begin{bmatrix} 1 & 0 \\ -t_{21} & 1 \end{bmatrix}$$

so that $b_{12} = 0$ and $b_{21} = -t_{21}$. It is straightforward to show that $t_{21} = +w_{12}/w_{11}$.

Variable Ordering

The identification of the SVAR using the triangular factorization depends on the ordering of the variables in $\boldsymbol{y}_t$. In the above analysis, it is assumed that $\boldsymbol{y}_t = (y_{1t}, y_{2t})'$ so that y_{1t} comes first in the ordering of the variables.

When the triangular factorization is conducted and the pseudo SVAR (30) is computed the structural $\boldsymbol{B}$ matrix has the form

$$\boldsymbol{B} = \boldsymbol{T}^{-1} = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix}$$

where $b_{12} = 0$.

If the ordering of the variables is reversed, $\boldsymbol{y}_t = (y_{2t}, y_{1t})'$, then the recursive causal ordering of the SVAR is reversed and the structural $\boldsymbol{B}$ matrix becomes

$$\boldsymbol{B} = \boldsymbol{T}^{-1} = \begin{bmatrix} 1 & 0 \\ b_{12} & 1 \end{bmatrix}$$

where $b_{21} = 0$.

5.4 Structual VAR(p) model

A SVAR(p) model is is defined as:

$$\boldsymbol{B}\boldsymbol{y}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p\boldsymbol{y}_{t-p} + \boldsymbol{L}\boldsymbol{\varepsilon}_t. \tag{31}$$

- the structural errors, $\boldsymbol{\varepsilon}_t$, are white noise
- the coefficient matrices $\boldsymbol{\Gamma}_i$ for $i = 1, \ldots, p$, are structural coefficients that differ in general from their reduced form counterparts.

Consider the resulting equation by left-multiplying Equation (31) with the inverse of $\boldsymbol{B}$:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{B}^{-1}\boldsymbol{\gamma}_0 + \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{B}^{-1}\boldsymbol{\Gamma}_p\boldsymbol{y}_{t-p} + \boldsymbol{B}^{-1}\boldsymbol{L}\boldsymbol{\varepsilon}_t \\ &\downarrow \\ \boldsymbol{y}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \end{aligned}$$

It should be noted that the *reduced form residuals* can be retrieved from a SVAR model by

$$\boldsymbol{u}_t = \boldsymbol{B}^{-1}\boldsymbol{L}\boldsymbol{\varepsilon}_t,$$

and its *variance-covariance matrix* by

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{B}^{-1}\boldsymbol{L}\boldsymbol{L}'\boldsymbol{B}^{-1'}.$$

Depending on the imposed restrictions, three types of SVAR models can be distinguished:

- First model: $\boldsymbol{L}$ is set to $\boldsymbol{I}_n$.
  (minimum number of restrictions for identification is $n(n-1)/2$).

- Second model: $\boldsymbol{B}$ is set to $\boldsymbol{I}_n$.
  (minimum number of restrictions to be imposed for identification is also $n(n-1)/2$).

- Third model: restrictions can be placed on both matrices.
  (minimum number of restrictions for identification is $n^2 + n(n-1)/2$).

The parameters are estimated by minimizing the negative of the concentrated log-likelihood function:

$$\begin{aligned}
\ln L_c(\boldsymbol{B}, \boldsymbol{L}) &= -\frac{nT}{2}\ln(2\pi) + \frac{T}{2}\ln|\boldsymbol{\Sigma}_u| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{u}_t'\boldsymbol{\Sigma}_u^{-1}\boldsymbol{u}_t \\
&= -\frac{nT}{2}\ln(2\pi) + \frac{T}{2}\ln|\boldsymbol{\Sigma}_u| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_u^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t'\right) \\
&= -\frac{nT}{2}\ln(2\pi) + \frac{T}{2}\ln|\boldsymbol{\Sigma}_u| - \frac{T}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_u^{-1}\widehat{\boldsymbol{\Sigma}}_u\right) \\
&= -\frac{nT}{2}\ln(2\pi) + \frac{T}{2}\ln|\boldsymbol{B}|^2 - \frac{T}{2}\ln|\boldsymbol{L}|^2 - \frac{T}{2}\mathrm{tr}(\boldsymbol{B}'\boldsymbol{L}^{-1\prime}\boldsymbol{L}^{-1}\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_u),
\end{aligned}$$

whereby $\widehat{\boldsymbol{\Sigma}}_u = 1/T\sum_{t=1}^{T}\hat{\boldsymbol{u}}_t\hat{\boldsymbol{u}}_t'$ signifies an estimate of the reduced form variance/covariance matrix for the error process.

Structual VAR(p) implementation in Eviews

Suppose that you have estimated a VAR for a three-variable system of the sort that we discussed last time. Say, you want to impose a set of recursive restrictions with the most exogenous variable ordered first and the least exogenous variable ordered last.

- Once you've obtained the output for the unrestricted (reduced-form) VAR, go to `Proc` and select `Estimate Structural Factorization`.

- Eviews designates the reduced-form shocks by `@e` and unobservable structural shocks by `@u`, immediately followed by the number of the equation.

- So you can describe the requisite set of restrictions in the `Text` box

by:

$$@e1 = c(1) * @u1$$

$$@e2 = c(2) * @e1 + c(3) * @u2$$

$$@e3 = c(4) * @e1 + c(5) * @e2 + c(6) * @u3$$

You will then be able to obtain the structural impulse responses that identify the unobservable shocks according to your set of restrictions.

6. Cointegrated VAR Models

6.1 Cointegration

(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 6 (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,

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

  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.

(2) Cointegration

Definition 3. *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 β 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 4 (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 $b > 0$.*

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.

(3) Superconsistency

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

Example 7. 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:

$$\widehat{\beta} = \beta + \frac{\sum_{t=1}^{T} x_t u_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0) u_t}{\sum_{t=1}^{T} (\alpha_x t + \sum_{k=1}^{t} \epsilon_k^x + x_0)^2} \tag{32}$$

In this case, the terms in t^2 will dominate as $T \to \infty$ 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 (32) 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 $\frac{1}{T^2}\sum_{t=1}^{T}\alpha_x t \to \frac{\alpha_x}{2}$ 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.

(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*.

(5) Error correction model (ECM)

定理 **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 β_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 $\beta(L) = \beta_0 + \beta_1 L + \cdots + \beta_q L^q$.

(6) Engle-Granger methodology

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

$$y_t = \alpha + \beta x_t + \varepsilon_t, \tag{33}$$

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 β.

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 β, 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 3 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)*

(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 (3×2) 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.

5.2 Vector error-correction representation

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

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{34}$$

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

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{35}$$

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

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^* \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^* \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{35}$$

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

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}', \tag{36}$$

   where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times m$ matrices with $\text{Rank}(\boldsymbol{\alpha}) = \text{Rank}(\boldsymbol{\beta}) = m$. The ECM of Eq. (35) becomes

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \mathbf{\Phi}_1^*\Delta \boldsymbol{y}_{t-1} + \cdots + \mathbf{\Phi}_p^*\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t. \tag{37}$$

   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 $\text{Rank}(\mathbf{\Pi}) = m$

- 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 8. 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, α_i and β_1 must satisfy the stationarity constraint $|1 + \alpha_1 + \alpha_2 \beta_1| < 1$.

5.3 Discussion of a Simple Model

(1) Granger representation or MA representation

Consider a cointegrated VAR model with no lagged differences:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{38}$$

- **Stationary representation:** Multiplying $\boldsymbol{\beta}'$ from the left to model (38), 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 (38),

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

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

- **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

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

- The impulse response function tend to C, i.e. $\boldsymbol{\Psi}_h \to C$, 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.

(2) Deterministic Terms

$$\Delta \boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{y}_{t-1} + \underline{\boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_t} \tag{39}$$

- Granger representation:

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

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

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:

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

6.4 Estimation and Cointegration Tests

(1) Specification of the Deterministic Function

Consider the following VECM

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^*\Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^*\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t. \tag{37}$$

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!)

(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

$$\Delta\boldsymbol{y}_t = \boldsymbol{\mu}\boldsymbol{d}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_1^*\Delta\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{p-1}^*\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \quad (40)$$

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:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\gamma}_0 \boldsymbol{d}_t + \boldsymbol{\Omega}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Omega}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{u}_t, \tag{41}$$

$$\boldsymbol{y}_{t-1} = \boldsymbol{\gamma}_1 \boldsymbol{d}_t + \boldsymbol{\Xi}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Xi}_{p-1} \Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{v}_t. \tag{42}$$

Let $\hat{\boldsymbol{u}}_t$ and $\hat{\boldsymbol{v}}_t$ be the residuals of Eqs. (41) and (42), 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

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

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

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

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 $\mathrm{Rank}(\boldsymbol{\Pi}) = m$.

(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 $\boldsymbol{\Pi}$ matrix in Eq. (35).

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 $\mathrm{Rank}(\boldsymbol{\Pi}) = m$ 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

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

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

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

If $\text{Rank}(\mathbf{\Pi}) = m$, then $\hat{\lambda}_i$ should be small for $i > m$ and hence $LR_{\text{tr}}(m)$ 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,

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

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.