Lecture 11. Covariance-stationary Vector Time Series

Muyi LI

WISE, Xiamen University

2020 Spring

Outline

- Stationarity and Cross-correlation Matrices
- Vector Autoregressive Model: Estimation, Diagnostic Checking, Forecasting
- Impulse Response Function and Variance Decomposition
- Structural Analysis

Cross-Correlation Matrices

- For a $d-$ dimensinal time series (weakly stationary) $\mathbf{y_t} = (y_{1t}, \dots, y_{dt})'$, we define its mean vector as

$$\mu = E(\mathbf{y}_t) = [E(y_{1t}), E(y_{2t}), \cdots E(y_{dt})]' := [\mu_1, \mu_2, \cdots, \mu_d]',$$

- The covariance matrix as

$$\mathbf{\Gamma}_0 = E[(\mathbf{y}_t - \mu)(\mathbf{y}_t - \mu)'] = \begin{bmatrix} \Gamma_{11}(0) & \Gamma_{12}(0) & \cdots & \Gamma_{1d}(0) \\ \Gamma_{21}(0) & \Gamma_{22}(0) & \cdots & \Gamma_{2d}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{d1}(0) & \Gamma_{d2}(0) & \cdots & \Gamma_{dd}(0) \end{bmatrix},$$

where

$$\Gamma_{ij}(0) = E[(y_{it} - \mu_i)(y_{jt} - \mu_j)], \quad i, j = 1, \dots, d$$

.

Cross-Correlation Matrices

- Let $\boldsymbol{D} = \mathrm{diag}\{\sqrt{\Gamma_{11}(0)}, \ldots, \sqrt{\Gamma_{dd}(0)}\}$.
- The concurrent, or lag-zero cross-correlation matrix is defined as

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

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

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

-

$$\rho_{ij}(0) = \mathrm{corr}(y_{it}, y_{jt}) = \mathrm{corr}(y_{jt}, y_{it}) = \rho_{ji}(0),$$
$$-1 \leq \rho_{ij}(0) \leq 1,$$
$$\text{and} \quad \rho_{ii}(0) = 1$$

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

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

$$\begin{aligned}\boldsymbol{\Gamma}_\ell &\equiv [\Gamma_{ij}(\ell)] = E[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-\ell} - \boldsymbol{\mu})'] \\ &= \begin{bmatrix} \mathrm{Cov}(y_{1t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{1t}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{1t}, y_{d,t-\ell}) \\ \mathrm{Cov}(y_{2t}, y_{1,t-\ell}) & \mathrm{Cov}(y_{2t}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{2t}, y_{d,t-\ell}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(y_{dt}, y_{1,t-\ell}) & \mathrm{Cov}(y_{dt}, y_{2,t-\ell}) & \cdots & \mathrm{Cov}(y_{dt}, y_{d,t-\ell}) \end{bmatrix} \\ &= \begin{bmatrix} \Gamma_{11}(\ell) & \Gamma_{12}(\ell) & \cdots & \Gamma_{1d}(\ell) \\ \Gamma_{21}(\ell) & \Gamma_{22}(\ell) & \cdots & \Gamma_{n2}(\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{d1}(\ell) & \Gamma_{d2}(\ell) & \cdots & \Gamma_{dd}(\ell) \end{bmatrix}. \end{aligned} \tag{1}$$

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

$$\boldsymbol{\rho}_\ell \equiv [\rho_{ij}(\ell)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell\boldsymbol{D}^{-1} = \begin{bmatrix} \rho_{11}(\ell) & \rho_{12}(\ell) & \cdots & \rho_{1d}(\ell) \\ \rho_{21}(\ell) & \rho_{22}(\ell) & \cdots & \rho_{2d}(\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1}(\ell) & \rho_{d2}(\ell) & \cdots & \rho_{dd}(\ell) \end{bmatrix}, \tag{2}$$

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 $\ell \geq 0$)

Information of CCM

- The diagonal elements $\rho_{ii}(\ell)$, $\ell = 0, 1, \ldots$ are the autocorrelation function (ACF) of y_{it}, $i = 1, \cdots, d$.
- 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 correlation (linear dependence) of y_{it} on the past value $y_{j,t-\ell}$.

Linear Relationship

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 $v > 0$.

Sample Cross-Correlation Matrices

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

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

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

Multivariate Portmanteau Tests

- Extend the univariate Ljung−Box test (1978) to the multivariate case(Hosking (1980)).
- $$H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = \mathbf{0} \leftrightarrow H_A : \boldsymbol{\rho}_i \neq \mathbf{0} \quad \text{for some} i \in \{1, \ldots, m\}.$$

  H_0 is true implies that there are neither auto- nor cross-correlations in the vector series $\boldsymbol{y}_t$.
- The test statistic:

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

  where T is the sample size, d 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.

Vector Autoregressive Model with order 1

- VAR(1) (reduced form):

$$\begin{aligned} \boldsymbol{y}_t &= \phi_0 + \boldsymbol{\Phi} \boldsymbol{y}_{t-1} + \varepsilon_t, \\ E(\varepsilon_t) = 0, \quad \text{and} \quad E(\varepsilon_t \varepsilon_s) &= \begin{cases} \boldsymbol{\Sigma}, & \text{if } t = s; \\ \mathbf{0}, & \text{otherwise,} \end{cases} \end{aligned} \tag{7}$$

  where ϕ_0 is a d-dimensional vector, $\boldsymbol{\Phi}$ is a $d \times d$ matrix.
- In empirical applications, it is often assumed that ε_t is i.i.d. multivariate normal $(\mathbf{0}, \boldsymbol{\Sigma})$.

- A VAR(1) model consists of the following two equations:

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

- 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 $y_{2,t-1}$, only depends on $y_{1,t-1}$.
- An unidirectional relationship from y_{1t} to y_{2t} if $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$.
- y_{1t} and y_{2t} are uncoupled, if $\Phi_{12} = \Phi_{21} = 0$.
- A feedback relationship between two series if $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$.

Stationary Condition

- Similar to the univariate AR process, for the VAR(1) process

$$\boldsymbol{y}_t = \phi_0 + \Phi \boldsymbol{y}_{t-1} + \varepsilon_t$$

- For stationarity, the roots in the characteristic equation

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

all lie outside the unit circle.

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

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

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

Structural Forms

- The previous VAR(1) model is called a reduced-form because it does not show explicitly the concurrent dependence between the component series.
- Next, we deduce its structural form by a simple linear transformation (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 $\boldsymbol{A}$.

Structural Forms

- The Cholesky decomposition shows that Σ which is the covariance matrix of ε_t (positive-definite matrix)can be diagonalized as

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

- The VAR(1) model

$$\boldsymbol{y}_t = \phi_0 + \Phi\boldsymbol{y}_{t-1} + \varepsilon_t$$

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

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

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

Structural form

- Multiplying $\boldsymbol{L}^{-1}$, obtain

$$\boldsymbol{L}^{-1}\boldsymbol{y}_t = \boldsymbol{L}^{-1}\phi_0 + \boldsymbol{L}^{-1}\boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{L}^{-1}\varepsilon_t = \phi_0^* + \boldsymbol{\Phi}^*\boldsymbol{y}_{t-1} + \boldsymbol{\eta}_t, \quad (8)$$

- Because of the lower triangle structure, the dth row of L^{-1} is in the form $(w_{d1}, w_{d2}, \ldots, w_{d,d-1}, 1)$. Consequently, the dth equation is

$$y_{dt} + \sum_{i=1}^{d-1} w_{di} y_{it} = \phi_{d,0}^* + \sum_{i=1}^{d} \Phi_{di}^* y_{i,t-1} + \eta_{dt}, \quad (9)$$

  where $\phi_{d,0}^*$ is the d^{th} element of ϕ_0^*, Φ_{di}^* is the $(d,i)^{th}$ element of $\boldsymbol{\Phi}^*$.

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

Structural Form

- For any other component, we can rearrange the VAR(1) model so that y_{it} becomes the last element of $\boldsymbol{y}_t$. The prior transformation can be applied to obtain a structural equation for y_{it}.
- Hence, the reduced-form model is equivalent to the structural form.
- In time series analysis, the reduced form is commonly used for two reasons. The first reason is easy in estimating. The second and main reason is that the concurrent correlations cannot be used in forecasting.

Vector AR(p) Models

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

$$\boldsymbol{y}_t = \phi_0 + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \varepsilon_t, \quad p > 0, \tag{10}$$

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

where ϕ_0 and ε_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

$$\begin{aligned} (\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{y}_t &= \phi_0 + \varepsilon_t, \\ \Longrightarrow \qquad \boldsymbol{\Phi}(B)\boldsymbol{y}_t &= \phi_0 + \varepsilon_t, \end{aligned}$$

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

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

$$\boldsymbol{y}_t = \boldsymbol{c} + \Phi_1 \boldsymbol{y}_{t-1} + \cdots + \Phi_p \boldsymbol{y}_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{NID}(\mathbf{0}, \Sigma). \tag{11}$$

We could rewrite it more concisely as

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

$$\text{where} \quad \underset{d\times(dp+1)}{\Pi'} \equiv \begin{bmatrix} \boldsymbol{c} & \Phi_1 & \Phi_2 & \cdots & \Phi_p \end{bmatrix} \quad \text{and} \quad \underset{(dp+1)\times 1}{\boldsymbol{x}_t} \equiv \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}.$$

Estimation

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 $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{1-p}$ is observed)*

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

Estimation

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

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

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

Estimation

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

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12}^2 & \cdots & \hat{\sigma}_{1d}^2 \\ \hat{\sigma}_{21}^2 & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{d1}^2 & \hat{\sigma}_{d2}^2 & \cdots & \hat{\sigma}_d^2 \end{bmatrix},$$

where

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

- 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 $i, j = 1, 2, \ldots, d$.

Order Selection

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

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

where $\widehat{\Sigma}(p) = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$ 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):

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

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

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

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

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

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 $j < 0$.

Forecast interval

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

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

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 $\widehat{\boldsymbol{\Sigma}}(h)$.

Impulse Response Function

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

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots, \tag{12}$$

  where $\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]^{-1}\phi_0$ 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. (12) is not straightforward.

Impulse Response Function

- By prior discussion, one can use the Cholesky decomposition($\Sigma = \boldsymbol{LGL}'$ and $\eta_t = L^{-1}\varepsilon_t$) to transform the innovations to be uncorrelated.
- $$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\mu} + LL^{-1}\varepsilon_t + \boldsymbol{\Psi}_1 LL^{-1}\varepsilon_{t-1} + \boldsymbol{\Psi}_2 LL^{-1}\varepsilon_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \boldsymbol{\Psi}_0^*\boldsymbol{\eta}_t + \boldsymbol{\Psi}_1^*\boldsymbol{\eta}_{t-1} + \boldsymbol{\Psi}_2^*\boldsymbol{\eta}_{t-2} + \cdots , \end{aligned} \tag{13}$$

  where $\boldsymbol{\Psi}_0^* = L$ and $\boldsymbol{\Psi}_i^* = \boldsymbol{\Psi}_i L$.
- 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 orthogonal impulse response function of y_i with respect to η_j. With d variables there are d^2

Forecast Error Variance Decomposition

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

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

Forecast Error Variance Decomposition

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

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

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

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

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

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

Structural Analysis: 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' 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.

Definition

- **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.

Structural Analysis

- 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:
$y_{1,t}$ 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 $E[y_{2,t+h}|I_{2,t}] = E[y_{2,t+h}|I_t]$.

Equivalent definition (model based)

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

$$\boldsymbol{y}_t = \mu + \Psi(B)\varepsilon_t = \mu + \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k}, \quad \Psi_0 = I_n \tag{14}$$

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

- 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 (14) is invertible, then

$$\boldsymbol{y}_t = \boldsymbol{c} + \Pi(B)\boldsymbol{y}_t + \varepsilon_t = \boldsymbol{c} + \sum_{k=1}^{\infty} \Pi_k \boldsymbol{y}_{t-k} + \varepsilon_t \tag{15}$$

- 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.
- 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.

Structural Analysis

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 $y_{3,2} \neq 0$.

Structural Analysis

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. See Lutkepohl (2005) and Dufor and Renault (1998) for detailed discussion.

Granger Causal analysis for bivariate VAR

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

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{16}$$

$$= \begin{bmatrix} \Psi_{11}(B) & \Psi_{12}(B) \\ \Psi_{21}(B) & \Psi_{22}(B) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{17}$$

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 $k = 1, \ldots, p$.

In other words, this corresponds to the restrictions that all cross-lags coefficients are all zeros which can be tested by Wald statistics.

Granger Causal analysis for bivariate VAR

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

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

$$H_0 : \phi_{21}^{(1)} = \phi_{21}^{(2)} = \cdots = \phi_{21}^{(p)} = 0, \tag{18}$$
$$H_A : \phi_{21}^{(1)} \neq 0 \quad \text{or} \quad \phi_{21}^{(2)} \neq 0 \quad \text{or} \quad \cdots \quad \phi_{21}^{(p)} \neq 0.$$

Granger causality (GC) test

- **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 $pF \xrightarrow{d} \chi_p^2$. So we have an asymmetrically equivalent test given by

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

Granger causality (GC) test

- **Wald test:** The Granger causal hypothesis given by (18) can be easily tested using the Wald statistic,

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

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

$$\boldsymbol{R} = \begin{bmatrix} \pi_1' & \phi_{20} & \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{21}^{(2)} & \phi_{22}^{(2)} & & \phi_{21}^{(p)} & \phi_{22}^{(p)} \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{1\times(n*p+1)} & . & . & . & . & . & \cdots & . & . \\ \mathbf{0}_{1\times(n*p+1)} & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

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