Lecture 6: Vector Autoregression*

In this section, we will extend our discussion to vector valued time series. We will be mostly interested in vector autoregression (VAR), which is much easier to be estimated in applications. We will fist introduce the properties and basic tools in analyzing stationary VAR process, and then we'll move on to estimation and inference of the VAR model.

1 Covariance-stationary VAR(p) process

1.1 Introduction to stationary vector ARMA processes

1.1.1 VAR processes

A VAR model applies when each variable in the system does not only depend on its own lags, but also the lags of other variables. A simple VAR example is:

$$\begin{aligned} x_{1t} &= \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \epsilon_{1t} \\ x_{2t} &= \phi_{21}x_{2,t-1} + \phi_{22}x_{2,t-2} + \epsilon_{2t} \end{aligned}$$

where $E(\epsilon_{1t}\epsilon_{2s}) = \sigma_{12}$ for $t = s$ and zero for $t \neq s$. We could rewrite it as

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{21} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},$$

or just

$$\mathbf{x}_t = \Phi_1\mathbf{x}_{t-1} + \Phi_2\mathbf{x}_{t-2} + \boldsymbol{\epsilon}_t \tag{1}$$

and $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$,$E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_s) = \mathbf{0}$ for $s \neq t$ and

$$E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

As you can see, in this example, the vector-valued random variable $\mathbf{x}_t$ follows a VAR(2) process. A general VAR(p) process with white noise can be written as

$$\begin{aligned} \mathbf{x}_t &= \Phi_1\mathbf{x}_{t-1} + \Phi_2\mathbf{x}_{t-2} + \ldots + \boldsymbol{\epsilon}_t \\ &= \sum_{j=1}^{p} \Phi_j\mathbf{x}_{t-j} + \boldsymbol{\epsilon}_t \end{aligned}$$

or, if we make use of the lag operator,

$$\Phi(L)\mathbf{x}_t = \boldsymbol{\epsilon}_t,$$

where

$$\Phi(L) = I_k - \Phi_1 L - \ldots - \Phi_p L^p.$$

The error terms follow a vector white noise, i.e., $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$,

$$E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s') = \begin{cases} \Omega & \text{for } t = s \\ \mathbf{0} & \text{otherwise} \end{cases}$$

with Ω a $(k \times k)$ symmetric positive definite matrix.

Recall that in studying the scalar AR(p) process,

$$\phi(L) x_t = \epsilon_t,$$

we have the results that the process $\{x_t\}$ is covariance-stationary as long as all the roots in (2)

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0 \tag{2}$$

lies out side of the unit circle. Similarly, for the VAR(p) process to be stationary, we must have that the roots in the equation

$$|I_k - \Phi_1 z - \ldots - \Phi_p z^p| = 0$$

all lies outside the unit circle.

1.1.2 Vector moving average processes

Recall that we could invert a scalar stationary AR(p) process, $\phi(L)x_t = \epsilon_t$ to a MA(∞) process, $x_t = \theta(L)\epsilon_t$, where $\theta(L) = \phi(L)^{-1}$. The same is true for a covariance-stationary VAR(p) process, $\Phi(L)x_t = \boldsymbol{\epsilon}_t$. We could invert it to

$$\mathbf{x}_t = \Psi(L)\boldsymbol{\epsilon}_t$$

where

$$\Psi(L) = \Phi(L)^{-1}$$

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

$$(I_k - \Phi_1 L - \Phi_2 L^2 - \ldots - \Phi_p L^p)(I_k + \Psi_1 L + \Psi_2 L^2 + \ldots) = I_k.$$

Equating the coefficients of L^j, we have $\Psi_0 = I_k$, $\Psi_1 = \Phi_1$, $\Psi_2 = \Phi_1\Psi_1 + \Phi_2$, and in general, we have

$$\Psi_s = \Phi_1 \Psi_{s-1} + \Phi_2 \Psi_{s-2} + \ldots + \Phi_p \Psi_{s-p}.$$

1.2 Transforming to a state space representation

Sometime, it is more convenient to write a scalar valued time series, say an AR(p) process, in vector form. For example,

$$x_t = \sum_{j=1}^{p} \theta_j x_{t-j} + \epsilon_t.$$

where $\epsilon \sim N(0, \sigma^2)$. We could equivalently write it as

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

If we let $\boldsymbol{\xi}_t = (x_t, x_{t-1}, \dots, x_{t-p+1})'$, $\boldsymbol{\xi}_{t-1} = (x_{t-1}, x_{t-2}, \dots, x_{t-p})$, $\boldsymbol{\epsilon}_t = (\epsilon_t, 0, \dots, 0)$, and let F denote the parameter matrix, then we can write the process as:

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

where $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}_p)$. So we have rewrite an AR(p) scalar process as an vector autoregression of order one, denoted by VAR(1).

Similarly, we could also transform a VAR(p) process to a VAR(1) process. For the process

$$\mathbf{x}_t = \Phi_1 \mathbf{x}_{t-1} + \Phi_2 \mathbf{x}_{t-2} + \dots + \Phi_p \mathbf{x}_{t-p} + \boldsymbol{\epsilon}_t,$$

let

$$\boldsymbol{\xi}_t = \begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{pmatrix},$$

$$F = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_k & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_k & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & I_k & \mathbf{0} \end{pmatrix},$$

$$\mathbf{v}_t = \begin{pmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix},$$

Then we could rewrite the VAR(p) process in state space notations,

$$\boldsymbol{\xi}_t = F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t. \tag{3}$$

where $E(\boldsymbol{v}_t \boldsymbol{v}_s')$ equals Q for $t = s$ and equals zero otherwise, and

$$Q = \begin{bmatrix} \Omega & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}.$$

1.3 The autocovariance matrix

1.3.1 VAR process

For a covariance stationary k dimensional vector process $\{\mathbf{x}_t\}$, let $E(\mathbf{x}_t) = \boldsymbol{\mu}$, then the autocovariance is defined to be the following k by k matrix

$$\Gamma(h) = E[(\mathbf{x}_t - \boldsymbol{\mu})(\mathbf{x}_{t-h} - \boldsymbol{\mu})'].$$

For simplicity, assume that $\boldsymbol{\mu} = \mathbf{0}$. Then we have $\Gamma(h) = E(\mathbf{x}_t \mathbf{x}_{t-h}')$. Because of the lead-lag effect, we may not have $\Gamma(h) = \Gamma(-h)$, but we have $\Gamma(h)' = \Gamma(-h)$. To show this,

$$\Gamma(h) = E(\mathbf{x}_{t+h}\mathbf{x}_{t+h-h}') = E(\mathbf{x}_{t+h}\mathbf{x}_t'),$$

taking transpose

$$\Gamma(h)' = E(\mathbf{x}_t \mathbf{x}_{t+h}') = \Gamma(-h).$$

Similar as in the scalar case, we define the autocovariance generating function of the process $\mathbf{x}$ as

$$G_{\mathbf{x}}(z) = \sum_{h=-\infty}^{\infty} \Gamma(h) z^h$$

where z is again a complex scalar.

Let $\boldsymbol{\xi}_t$ as defined in (3). Assume that $\boldsymbol{\xi}$ and $\mathbf{x}$ are stationary, and let Σ denote the variance of $\boldsymbol{\xi}$,

$$\begin{aligned}
\Sigma &= E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') \\
&= E\left[\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_t' & \mathbf{x}_{t-1}' & \dots & \mathbf{x}_{t-p+1}' \end{pmatrix}\right] \\
&= \begin{bmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \Gamma(1)' & \Gamma(0) & \dots & \Gamma(p-2) \\ \vdots & \vdots & \dots & \vdots \\ \Gamma(p-1)' & \Gamma(p-2)' & \dots & \Gamma(0) \end{bmatrix}.
\end{aligned}$$

Postmultiplying (3) by its transpose and taking expectations gives

$$E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t') = E[(F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)(F\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)'] = FE(\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}')F' + E(\mathbf{v}_t \mathbf{v}_t'),$$

or

$$\Sigma = F\Sigma F' + Q. \tag{4}$$

To solve for Σ, we need to use the Kronecker product, and the following result: let A, B, C be matrices whose dimensions are such that the product ABC exists. Then

$$\text{vec}(ABC) = (C' \otimes A) \cdot \text{vec}(B).$$

where vec is the operator to stack each column of a matrix $(k \times k)$ into a k^2-dimensional vector, for example,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{vec}(A) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}.$$

Apply vec operator on both sides of (4), we get

$$\text{vec}(\Sigma) = (F \otimes F) \cdot \text{vec}(\Sigma) + \text{vec}(Q),$$

which gives

$$\text{vec}(\Sigma) = (I_m - F \otimes F)^{-1}\text{vec}(Q),$$

where $m = k^2p^2$. We can use this equation to solve for the first p order of autocovariance of $\mathbf{x}$, $\Gamma(0), \ldots, \Gamma(p)$. To derive the hth autocovariance of $\boldsymbol{\xi}$, denoted by $\Sigma(h)$, we can postmultiplying (3) by $\boldsymbol{\xi}'_{t-h}$ and take expectations,

$$E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t-h}) = FE(\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}'_{t-h}) + E(\mathbf{v}_t\boldsymbol{\xi}'_{t-h}),$$

then

$$\Sigma(h) = F\Sigma(h-1), \quad \text{or} \quad \Sigma(h) = F^h\Sigma.$$

Therefore we have the following relationship for $\Gamma(h)$

$$\Gamma(h) = \Phi_1\Gamma(h-1) + \Phi_2\Gamma(h-2) + \ldots + \Phi_p\Gamma(h-p).$$

1.3.2 Vector MA processes

We first consider the MA(q) process.

$$\mathbf{x}_t = \boldsymbol{\epsilon}_t + \Psi_1\boldsymbol{\epsilon}_{t-1} + \Psi_2\boldsymbol{\epsilon}_{t-2} + \ldots + \Psi_q\boldsymbol{\epsilon}_{t-q}.$$

Then the variance of $\mathbf{x}_t$ is

$$\begin{aligned} \Gamma(0) &= E(\mathbf{x}_t\mathbf{x}'_t) \\ &= E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t) + \Psi_1 E(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1})\Psi'_1 + \ldots + \Psi_q E(\boldsymbol{\epsilon}_{t-q}\boldsymbol{\epsilon}'_{t-q})\Psi'_q \\ &= \Omega + \Psi_1\Omega\Psi'_1 + \Psi_2\Omega\Psi'_2 + \ldots + \Psi_q\Omega\Psi'_q. \end{aligned}$$

and the autocovariances

$$\Gamma(h) = \begin{cases} \Psi_h\Omega + \Psi_{h+1}\Omega\Psi_1 + \Psi_{h+2}\Omega\Psi'_2 + \ldots + \Psi_q\Omega\Psi'_{q-j} & \text{for } h = 1, \ldots, q. \\ \Omega\Psi'_{-h} + \Psi_1\Omega\Psi'_{-h+1} + \Psi_2\Omega\Psi'_{-h+2} + \ldots + \Psi_{q+h}\Omega\Psi'_q & \text{for } h = -1, \ldots, -q. \\ \mathbf{0} \quad \text{for } |h| > q \end{cases}$$

As in the scalar case, any vector MA(q) process is stationary. Next consider the MA(∞) process

$$\mathbf{x}_t = \boldsymbol{\epsilon}_t + \Psi_1\boldsymbol{\epsilon}_{t-1} + \Psi_2\boldsymbol{\epsilon}_{t-2} + \ldots = \Psi(L)\boldsymbol{\epsilon}_t.$$

A sequence of matrices $\{\Psi_s\}_{-\infty}^{\infty}$ is absolutely summable if each of its element forms an absolutely summable scalar sequence, i.e.

$$\sum_{s=0}^{\infty} |\psi_{ij}^{(s)}| < \infty \quad \text{for } i, j = 1, 2, \dots n,$$

where $\psi_{ij}^{(s)}$ is the row i column j element (will use ijth for short) of Ψ_s. Some important results about MA(∞) process is summarized as follows:

Proposition 1 *Let $\mathbf{x}_t$ be a $k \times 1$ vector satisfying*

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j},$$

where $\boldsymbol{\epsilon}_t$ is vector white noise and Ψ_j is absolutely summable. Then

(a) The autocovaiance between the ith variable at time t and the jth variable s periods earlier, $E(x_{it}x_{j,t-s})$ exists and is given by the ijth element of

$$\Gamma(s) = \sum_{v=0}^{\infty} \Psi_{s+v} \Omega \Psi_v' \quad \text{for } s = 0, 1, 2, \dots;$$

(b) $\{\Gamma(h)\}_{h=0}^{\infty}$ is absolutely summable.

If $\{\boldsymbol{\epsilon}_t\}_{t=-\infty}^{\infty}$ is i.i.d. with $E|\epsilon_{i1,t}\epsilon_{i2,t}\epsilon_{i3,t}\epsilon_{i4,t}| < \infty$ for $i1, i2, i3, i4 = 1, 2, \dots, k$ then we also have

(c) $E|x_{i1,t1}x_{i2,t2}x_{i3,t3}x_{i4,t4}| < \infty$ for $i1, i2, i3, i4 = 1, 2, \dots, k$ and all $t1, t2, t3, t4$.

(d) $n^{-1}\sum_{t=1}^{n} x_{it}x_{j,t-s} \to_p E(x_{it}x_{j,t-s})$ for $i, j = 1, 2, \dots, k$ and for all s.

All of these results can be viewed as extensions from the scalar case to vector case, and its proof can be found on page 286-288 in Hamilton's book.

1.4 The Sample Mean of a Vector Process

Let $\mathbf{x}_t$ be a stationary process with $E(\mathbf{x}_t) = 0$ and $E(\mathbf{x}_t\mathbf{x}_{t-h}) = \Gamma(h)$, where $\Gamma(h)$ is absolutely summable. Then we consider the properties of the sample mean

$$\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t.$$

$$\begin{aligned} & E[(\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n') \\ = & \frac{1}{n^2}E[(\mathbf{x}_1 + \dots \mathbf{x}_n)(\mathbf{x}_1 + \dots \mathbf{x}_n)'] \\ = & \frac{1}{n^2}\sum_{i,j}^{n} E(\mathbf{x}_i\mathbf{x}_j') \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2} \sum_{h=-\infty}^{\infty} \Gamma(h) \\
&= \frac{1}{n} \left(\sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \Gamma(h) \right)
\end{aligned}
$$

Then

$$
\begin{aligned}
& nE[(\bar{\mathbf{x}}_n \bar{\mathbf{x}}_n')] \\
= & \left(\sum_{h=-n+1}^{n-1} \left(1 - \frac{|h|}{n}\right) \Gamma(h) \right) \\
= & \Gamma(0) + \left(1 - \frac{1}{n}\right)(\Gamma(1) + \Gamma(-1)) + \left(1 - \frac{2}{n}\right)(\Gamma(2) + \Gamma(-2)) + \dots \\
\to & \sum_{h=-\infty}^{\infty} \Gamma(h)
\end{aligned}
$$

This is very similar as what we did in the scalar case. Then we have the following proposition:

Proposition 2 *Let* $\mathbf{x}_t$ *be a zero mean stationary process with* $E(\mathbf{x}_t) = 0$ *and* $E(\mathbf{x}_t \mathbf{x}_{t-h}) = \Gamma(h)$*, where* $\Gamma(h)$ *is absolutely summable, then the sample mean satisfies*

(a) $\bar{\mathbf{x}}_n \to_p 0$

(b) $\lim_{n\to\infty}[nE(\bar{\mathbf{x}}_n \bar{\mathbf{x}}_n')] = \sum_{h=-\infty}^{\infty} \Gamma(h)$.

Let S denote the limit variance of $nE(\bar{\mathbf{x}}_n \bar{\mathbf{x}}_n')$. If the data are generated by a MA(q) process, then results (b) implies that

$$S = \sum_{h=-q}^{q} \Gamma(h).$$

Then a natural estimate for S is

$$\hat{S} = \hat{\Gamma}(h) + \sum_{h=1}^{q} (\Gamma(h) + \Gamma(h)'), \tag{5}$$

where

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=h+1}^{n} (\mathbf{x}_t - \bar{\mathbf{x}}_n)(\mathbf{x}_{t-h} - \bar{\mathbf{x}}_n)'.$$

$\hat{S}$ defined in (5) provides a consistent estimator for a large class of stationary processes. Even when the process has time-varying second moments, as long as

$$\frac{1}{n} \sum_{t=h+1}^{n} (\mathbf{x}_t - \bar{\mathbf{x}}_n)(\mathbf{x}_t - \bar{\mathbf{x}}_n)'$$

converges in probability to

$$\frac{1}{n}\sum_{t=h+1}^{n} E(\mathbf{x}_t\mathbf{x}_{t-h}),$$

$\hat{S}$ is a consistent estimate of $nE(\bar{\mathbf{x}}_n\bar{\mathbf{x}}_n')$. It is used not only for MA(q) process. Write the autocovariance as $E(\mathbf{x}_t\mathbf{x}_s)$, even it is nonzero for all t and s, if the matrix goes to zero sufficiently fast as $|t-s|\to\infty$, and q is growing with the sample n, then we still have $\hat{S}\to S$.

However, a problem with $\hat{S}$ is that it may not be positive semidefinite in small samples. Therefore, we can use the Newey and West estimate

$$\tilde{S} = \hat{\Gamma}_0 + \sum_{h=1}^{q}\left(1-\frac{h}{q+1}\right)(\Gamma(h)+\Gamma(h)'),$$

which is positive semidefinite and has the same consistency properties of $\hat{S}$ when $q,n\to\infty$ with $q/n^{1/4}\to 0$.

1.5 Impulse-response Function and Orthogonalization

1.5.1 Impulse-response function

Impulse-response function gives how a time series variable is affected given a shock at time t. Recall that for a scalar time series process, say, a AR(1) process $x_t = \phi x_{t-1} + \epsilon_t$ with $|\phi| < 1$, we can invert it to a MA process $x_t = (1 + \phi L + \phi^2 L + \ldots)\epsilon_t$, and the effects of ϵ on x are:

$$\begin{array}{lccccc} \epsilon: & 0 & 1 & 0 & 0 & \ldots \\ x: & 0 & 1 & \phi & \phi^2 & \ldots \end{array}$$

In other words, after we invert $\phi(L)x_t = \epsilon_t$ to $x_t = \theta(L)\epsilon_t$, the $\theta(L)$ function gives us how x response to a a unit shock from ϵ_t.

We could do similar thing on a VAR process. In our earlier example, we have a VAR(2) system,

$$\mathbf{x}_t = \Phi_1\mathbf{x}_{t-1} + \Phi_2\mathbf{x}_{t-2} + \boldsymbol{\epsilon}_t$$

and $\boldsymbol{\epsilon}_t \sim WN(0,\Omega)$ where

$$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$$

After we invert it to a MA(∞) representation

$$\mathbf{x}_t = \Psi(L)\boldsymbol{\epsilon}_t \tag{6}$$

where $\Psi(L) = (1-\Phi_1 L - \Phi_2 L^2)^{-1}$, we see that in this representation, the observations $\mathbf{x}_t$ is a linear combinations of shocks $\boldsymbol{\epsilon_t}$. However, suppose we are interested in another form of shocks, say

$$\mathbf{u}_t = Q\boldsymbol{\epsilon}_t$$

where Q is an arbitrary square matrix (in this example, it is 2 by 2), we have

$$\mathbf{x}_t = \Psi(L)Q^{-1}Q\boldsymbol{\epsilon}_t = A(L)\mathbf{u}_t \tag{7}$$

where we let $A(L) = \Psi(L)Q^{-1}$. Since this Q is arbitrary, you see that we can have many linear combinations of shocks, and response functions. Then which combinations shall we use?

1.5.2 Orthogonalization and model specification

In economic modeling, we calculate the impulse-response dynamics as we are interested how economic variables response to certain source of shocks. If the shocks are correlated, then it is hard to identify what is the response to a particular shock. From that view, we may want to choose the Q to make $\mathbf{u}_t = Q\boldsymbol{\epsilon}_t$ orthonormal, or uncorrelated across each other and with unit variance, i.e., $E(\mathbf{u}_t\mathbf{u}_t') = I$. To do so, we need a Q such that

$$Q^{-1}Q^{-1'} = \Omega,$$

then $E(\mathbf{u}_t\mathbf{u}_t') = E(Q\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'Q') = Q\Omega Q' = I_k$. So, we can use Choleski decomposition to find Q. However, Q is still not unique as you can form other Qs by multiplying an orthogonal matrix.

Sims (1980) proposes that we could specify the model by choosing a particular leading term in the coefficient, A_0. In (6), we see that $\Psi_0 = I_k$. However, in (7), $A_0 = Q^{-1}$ cannot be identity matrix unless Ω is diagonal. In our example, we would choose the Q which produces $\mathbf{A}_0 = Q^{-1}$ as a lower triangular matrix. That means after this transformation, shock u_{2t} has no effects on x_{1t}. The nice thing is that Choleski decomposition itself will produce a triangular matrix.

Example 1 Consider a AR(1) process of a 2-dimensional vector,

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

where

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

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,

$$\mathbf{x}_t = \Psi(L)\boldsymbol{\epsilon}_t.$$

We know that $\Psi_0 = I_2$, $\Psi_1 = \Phi_1$, etc. Then we find Q by Choleski decomposition of Ω, which gives

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

Then we can write

$$\mathbf{x}_t = \Psi(L)Q^{-1}Q\boldsymbol{\epsilon}_t = \Psi(L)Q^{-1}\mathbf{u}_t$$

where we define that $\mathbf{u}_t = Q\boldsymbol{\epsilon}_t$. Then we have

$$\mathbf{x}_t = \Psi_0 Q^{-1}\mathbf{u}_t + \Psi_1 Q^{-1}\mathbf{u}_{t-1} + \ldots.$$

or

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1.41 & 0 \\ 0.70 & 1.87 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} + \begin{bmatrix} 0.85 & 0.37 \\ 0.70 & 0.75 \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \ldots$$

In this example you see that we find a unique MA representation which is linear combination of uncorrelated error ($E(\mathbf{u}_t\mathbf{u}_t') = I_2$), and the second sources of shock does not have instantaneous effects on x_{1t}. We can then use this representation to compute the impulse-responses.

There are also other ways to specify the representation, depending on the problem of interest. For example, Quah (1988) suggests that find a Q so that the long-run response of one variable to another shocks is zero.

1.5.3 Variance decomposition

Now, let's consider how we could decompose the variance of the forecasting errors. $\mathbf{x}_t = \Psi(L)\boldsymbol{\epsilon}_t = A(L)\mathbf{u}_t$ where $A(L) = \Psi(L)Q$, $\mathbf{u}_t = Q\boldsymbol{\epsilon}_t$ and $E(\mathbf{u}_t\mathbf{u}_t') = I$. For simplicity, we let $(\mathbf{x}_t = (x_{1t}, x_{2t}')$. Suppose we do a one-period ahead forecasting, and let $\mathbf{y}_{t+1}$ denote the forecast error,

$$\mathbf{y}_{t+1} = \mathbf{x}_{t+1} - E_t(\mathbf{x}_{t+1}) = A_0\mathbf{u}_{t+1} = \begin{bmatrix} A_{11}^0 & A_{12}^0 \\ A_{21}^0 & A_{22}^0 \end{bmatrix} \begin{bmatrix} u_{1,t+1} \\ u_{2,t+1} \end{bmatrix}.$$

Since $E(u_{1t}u_{2t}) = 0$, $E(u_{it}^2) = 1$, the variance of the forecasting error is given by $E(\mathbf{y}_{t+1}\mathbf{y}_{t+1}') = A_0A_0'$. So the variance of forecasting error for x_{1t} is given by $(A_{11}^0)^2 + (A_{12}^0)^2$. We can interpret that $(A_{11}^0)^2$ is the amount of the one-step ahead forecasting error variance due to shock u_1, and $(A_{12}^0)^2$ is the amount due to shock u_2. Similarly the variance of forecasting error of x_{2t} is given by $(A_{21}^0)^2 + (A_{22}^0)^2$, and we can interpret them as amount due to shock u_1 and u_2 respectively. The variance for k-period ahead forecasting error can be computed in a similar way.

2 Estimation of VAR(p) process

2.1 Maximum Likelihood Estimation

Usually we use conditional likelihood in VAR estimation (recall that conditional likelihood functions are much easier to work with than unconditional likelihood functions).

Given a k-vector VAR(p) process,

$$\mathbf{y}_t = \mathbf{c} + \Phi_1\mathbf{y}_{t-1} + \Phi_2\mathbf{y}_{t-2} + \ldots + \boldsymbol{\epsilon}_t,$$

we could rewrite it more concisely as

$$\mathbf{y}_t = \Pi'\mathbf{x}_t + \boldsymbol{\epsilon}_t.$$

where

$$\Pi = \begin{pmatrix} \mathbf{c}' \\ \Phi_1' \\ \Phi_2' \\ \vdots \\ \Phi_p' \end{pmatrix} \quad \text{and} \quad \mathbf{x}_t = \begin{pmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix}$$

If we assume that $\boldsymbol{\epsilon} \sim i.i.d.N(0, \Omega)$, then we could use MLE to estimate the parameters in $\theta = (\mathbf{c}, \Pi, \Omega)$. Following the same way in the scalar case, assume that we have observed $(\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0)$, then the likelihood function for the $\mathbf{y}_t$ is

$$L(\mathbf{y}_t, \mathbf{x}_t; \theta) = (2\pi)^{-k/2}|\Omega^{-1}|^{1/2}\exp[(-1/2)(\mathbf{y}_t - \Pi'\mathbf{x}_t)'\Omega^{-1}(\mathbf{y}_t - \Pi'\mathbf{x}_t)]$$

The log likelihood function of observations $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is (constant omitted)

$$l(\mathbf{y}, \mathbf{x}; \theta) = (n/2)\log|\Omega^{-1}| - (1/2)\sum_{t=1}^{n}\left[(\mathbf{y}_t - \Pi'\mathbf{x}_t)'\Omega^{-1}(\mathbf{y}_t - \Pi'\mathbf{x}_t)\right]. \quad (8)$$

Taking first derivative with respect to Π and Ω, we have that

$$\hat{\Pi}'_n = \left[\sum_{t=1}^{n}\mathbf{y}_t\mathbf{x}'_t\right]\left[\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}'_t\right]^{-1}.$$

The jth row of $\hat{\Pi}'_n$ is

$$\hat{\boldsymbol{\pi}}'_j = \left[\sum_{t=1}^{n}y_{jt}\mathbf{x}'_t\right]\left[\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}'_t\right]^{-1}.$$

which is the estimated coefficient vector from an OLS regression of y_{jt} on $\mathbf{x}_t$. So the MLE estimates of the coefficients for the jth equation of a VAR are found by an OLS regression of y_{jt} on a constant term and p lags of all of the variables in the system.

The MLE estimate of Ω is

$$\hat{\Omega}_n = (1/n)\sum_{t=1}^{n}\hat{\boldsymbol{\epsilon}}_t\hat{\boldsymbol{\epsilon}}'_t$$

where

$$\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_t - \hat{\Pi}'_n\mathbf{x}_t$$

The details on the derivations can be found on page 292-296 on Hamilton book. The MLE estimates $\hat{\Pi}$ and $\hat{\Omega}$ are consistent even if the true innovations are non-Gaussian. In the next subsection, we will consider regression with non-Gaussian errors, and we will use the LS approach to derive for the asymptotics.

2.2 LS estimation and asymptotics

The asymptotic distribution of $\hat{\Pi}$ is summarized in the following proposition

Proposition 3

$$\mathbf{y}_t = \mathbf{c} + \Phi_1\mathbf{y}_{t-1} + \Phi_2\mathbf{y}_{t-2} + \ldots + \Phi_p\mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t,$$

$\boldsymbol{\epsilon}_t = i.i.d.(0, \Omega)$, and $E(\epsilon_{it}\epsilon_{jt}\epsilon_{lt}\epsilon_{mt}) < \infty$ for all i, j, l, and m and where roots of

$$|I_k - \Phi_1 z - \ldots - \Phi_p z^p| = 0$$

lie outside the unit circle. Let $m = kp + 1$ and let

$$\mathbf{x}'_t = [\ 1 \quad \mathbf{y}'_{t-1} \quad \mathbf{y}'_{t-2} \quad \ldots \quad \mathbf{y}'_{t-p}\ ],$$

So $\mathbf{x}_t$ is a m-dimensional vector. Let $\hat{\boldsymbol{\pi}}_n = \mathrm{vec}(\hat{\Pi}_n)$ denote the $km \times 1$ vector of coefficients resulting from OLS regression of each of the elements of $\mathbf{y}_t$ on $\mathbf{x}_t$ for a sample of size n:

$$\hat{\boldsymbol{\pi}}'_n = [\ \hat{\boldsymbol{\pi}}'_{1,n} \quad \hat{\boldsymbol{\pi}}'_{2,n} \quad \ldots \quad \hat{\boldsymbol{\pi}}'_{k,n}\ ]$$

where

$$\hat{\boldsymbol{\pi}}_{i,n} = \left[\sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t'\right]^{-1} \left[\sum_{t=1}^{n} \mathbf{x}_t' y_{it}\right],$$

and let $\hat{\boldsymbol{\pi}}_0$ *denote the* km *by 1 vector of the true parameter. Finally, let*

$$\hat{\Omega}_n = n^{-1} \sum_{t=1}^{n} \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t',$$

where

$$\begin{aligned} \hat{\boldsymbol{\epsilon}}_t' &= \left[\ \hat{\epsilon}_{1t} \quad \hat{\epsilon}_{2t} \quad \ldots \quad \hat{\epsilon}_{kt}\ \right] \\ \hat{\epsilon}_{it} &= y_{it} - \mathbf{x}_t' \hat{\boldsymbol{\pi}}_{i,n} \end{aligned}$$

Then

(a) $n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t' \to_p Q$ *where* $Q = E(\mathbf{x}_t \mathbf{x}_t')$;

(b) $\hat{\boldsymbol{\pi}}_n \to_p \boldsymbol{\pi}$;

(c) $\hat{\Omega}_n \to_p \Omega$;

(d) $\sqrt{n}(\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}) \to_d N(0, \Omega \otimes Q^{-1})$.

Result (a) is a vector version of that sample second moment converges to the population moment, and it follows that the coefficients are absolutely summable and it has finite fourth moment. Result (b) and (c) are similar to the derivations for single OLS regression in case 3 in lecture 5. To show result (d), let

$$Q_n = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{x}_t',$$

then we could write

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_{i,n} - \boldsymbol{\pi}_i) = Q_n^{-1} \left[n^{-1/2} \sum_{t=1}^{n} \mathbf{x}_t \epsilon_{it}\right]$$

and

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}) = \begin{bmatrix} Q_n^{-1} n^{-1/2} \sum_{t=1}^{n} \mathbf{x}_t \epsilon_{1t} \\ Q_n^{-1} n^{-1/2} \sum_{t=1}^{n} \mathbf{x}_t \epsilon_{2t} \\ \vdots \\ Q_n^{-1} n^{-1/2} \sum_{t=1}^{n} \mathbf{x}_t \epsilon_{kt} \end{bmatrix}. \tag{9}$$

Define $\boldsymbol{\xi}_t$ to be a $km \times 1$ vector

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{x}_t \epsilon_{1t} \\ \mathbf{x}_t \epsilon_{2t} \\ \vdots \\ \mathbf{x}_t \epsilon_{kt} \end{bmatrix}.$$

Note that $\boldsymbol{\xi}_t$ is a mds with finite fourth moments and variance

$$\begin{aligned} E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') &= \begin{bmatrix} E(\epsilon_{1t}^2) & E(\epsilon_{1t}\epsilon_{2t}) & \dots & E(\epsilon_{1t}\epsilon_{kt}) \\ E(\epsilon_{2t}\epsilon_{1t}) & E(\epsilon_{2t}^2) & \dots & E(\epsilon_{2t}\epsilon_{kt}) \\ \vdots & \vdots & \dots & \vdots \\ E(\epsilon_{kt}\epsilon_{1t}) & E(\epsilon_{kt}\epsilon_{2t}) & \dots & E(\epsilon_{kt}^2) \end{bmatrix} \otimes E(\mathbf{x}_t\mathbf{x}_t') \\ &= \Omega \otimes Q \end{aligned}$$

We can also show that

$$n^{-1}\sum_{t=1}^{n} \boldsymbol{\xi}_t\boldsymbol{\xi}_t' \to_p \Omega \otimes Q.$$

Apply the CLT for vector mds, we have

$$n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\xi}_t \to_d N(0, \Omega \otimes Q). \tag{10}$$

Now rewrite (9) as

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}) &= \begin{bmatrix} Q_n^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & Q_n^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & Q_n^{-1} \end{bmatrix} \begin{bmatrix} n^{-1/2}\sum_{t=1}^n \mathbf{x}_t\epsilon_{1t} \\ n^{-1/2}\sum_{t=1}^n \mathbf{x}_t\epsilon_{2t} \\ \vdots \\ n^{-1/2}\sum_{t=1}^n \mathbf{x}_t\epsilon_{kt} \end{bmatrix} \\ &= (I_k \otimes Q_n^{-1}) n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\xi}_t \end{aligned}$$

By result (a) we have $Q_n^{-1} \to_p Q^{-1}$. Thus

$$n^{1/2}(\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}) \to_p (I_k \otimes Q^{-1}) n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\xi}_t.$$

From (10) we know that this has a distribution that is Gaussian with mean $\mathbf{0}$ and variance

$$(I_k \otimes Q^{-1})(\Omega \otimes Q)(I_k \otimes Q^{-1}) = (I_k \Omega I_k) \otimes (Q^{-1}QQ^{-1}) = \Omega \otimes Q^{-1}.$$

Hence we got result (d). Each of $\hat{\boldsymbol{\pi}}_i$ has the distribution

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_{i,n} - \boldsymbol{\pi}_i) \to_d N(0, \sigma_i^2 Q^{-1}).$$

Given that the estimators are asymptotically normal, we can use it to test linear or nonlinear restrictions on the coefficients with the Wald statistics.

We know that vec is an operator to stack each column of a $k \times k$ matrix into one $k^2 \times 1$ vector. A similar operator, vech, is to stack all elements under the principal diagonal (so it transforms a $k \times k$ matrix into one $k(k+1)/2 \times 1$ vector). For example,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{vech}(A) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix}.$$

We will apply this operator on the variance matrix, which is symmetric. The joint distribution of $\hat{\boldsymbol{\pi}}_n$ and $\hat{\Omega}_n$ is given in the following proposition.

Proposition 4

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \ldots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t,$$

$\boldsymbol{\epsilon}_t = i.i.d. N(0, \Omega)$, and where roots of

$$|I_k - \Phi_1 z - \ldots - \Phi_p z^p| = 0$$

lie outside the unit circle. Let $\hat{\boldsymbol{\pi}}_n$, $\hat{\Omega}_n$, and Q be as defined in proposition 3, then

$$\begin{bmatrix} n^{1/2}[\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}] \\ n^{1/2}[\text{vech}(\hat{\Omega}_n) - \text{vech}(\Omega)] \end{bmatrix} \to_d N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Omega \otimes Q^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix} \right).$$

Let σ_{ij} denote the ijth element of Ω then the element of Σ_{22} corresponding to the covariance between $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{lm}$ is given by $(\sigma_{il}\sigma_{jm} + \sigma_{im}\sigma_{jl})$ for all $i, j, l, m = 1, \ldots k$.

The detailed proof can be found on page 341-342 in Hamilton book. Basically there are three steps: first, we show that $\hat{\Omega}_n = n^{-1} \sum_{t=1}^n \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t{}'$ has the same asymptotic distribution as $\hat{\Omega}_n^* = n^{-1} \sum_{t=1}^n \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'$. In the second step, write

$$\begin{bmatrix} n^{1/2}[\hat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}] \\ n^{1/2}[\text{vech}(\hat{\Omega}_n) - \text{vech}(\Omega)] \end{bmatrix} \to_d \begin{bmatrix} (I_k \otimes Q^{-1}) n^{-1/2} \sum_{t=1}^n \boldsymbol{\xi}_t \\ n^{-1/2} \sum_{t=1}^n \lambda_t \end{bmatrix}$$

where

$$\lambda_t = \text{vech} \begin{bmatrix} \epsilon_{1t}^2 - \sigma_{11} & \ldots & \epsilon_{1t}\epsilon_{kt} - \sigma_{1k} \\ \vdots & \ldots & \vdots \\ \epsilon_{kt}\epsilon_{k1} - \sigma_{k1} & \ldots & \epsilon_{kt}^2 - \sigma_{kk} \end{bmatrix}.$$

Now, $(\boldsymbol{\xi}_t', \lambda_t')$ is an mds and we apply the CLT for mds to get (with a few more computations)

$$\begin{bmatrix} n^{-1/2} \sum_{t=1}^n \boldsymbol{\xi}_t \\ n^{-1/2} \sum_{t=1}^n \lambda_t \end{bmatrix} \to_d N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Omega \otimes Q^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix} \right).$$

The final step in the proof is to show that $E(\lambda_t \lambda_t')$ is given by the matrix Σ_{22} as described in the proposition, which can be proved with a constructed error sequence which is uncorrelated Gaussian with zero mean and unit variance (see Hamilton's book for details).

With the asymptotic variance of $\hat{\Omega}_n$, we can then test if two errors are correlated. For example, for $k = 2$,

$$\sqrt{n} \begin{bmatrix} \hat{\sigma}_{11,n} - \sigma_{11} \\ \hat{\sigma}_{12,n} - \sigma_{12} \\ \hat{\sigma}_{22,n} - \sigma_{22} \end{bmatrix} \to_d N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & \sigma_{12}^2 \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2 \end{bmatrix} \right).$$

Then a Wald test of the null hypothesis that there is no covariance between ϵ_{1t} and ϵ_{2t} is given by

$$\frac{\sqrt{n}\hat{\sigma}_{12}}{(\hat{\sigma}_{11}\hat{\sigma}_{22} + \sigma_{12}^2)^{1/2}} \approx N(0, 1).$$

The matrix Σ_{22} can be expressed more compactly using the *duplication matrix.* Duplication matrix $\mathbf{D}_k$ is a matrix of size $k^2 \times k(k+1)/2$ matrix that transforms vech(Ω) into vec(Ω), i.e.

$$\mathbf{D}_k \text{vech}(\Omega) = \text{vec}(\Omega).$$

For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}.$$

Define

$$\mathbf{D}_k^+ \equiv (\mathbf{D}_k' \mathbf{D}_k)^{-1} \mathbf{D}_k$$

Note that $\mathbf{D}_k^+ \mathbf{D}_k = I_{k(k+1)/2}$. $\mathbf{D}_k^+$ is like the 'reverse' of $\mathbf{D}_k$ as it transform vec(Ω) into vech(Ω),

$$\text{vech}(\Omega) = \mathbf{D}_k^+ \text{vec}(\Omega).$$

For example, when $k = 2$, we have

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}.$$

With $\mathbf{D}_k$ and $\mathbf{D}_k^+$ we can write

$$\Sigma_{22} = 2\mathbf{D}_k^+ (\Omega \otimes \Omega)(\mathbf{D}_k^+)'.$$

3 Granger Causality

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 tells the 'co-occurrence' of x and y, not that x causes y. In other words, usually the regression only tells 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.

One good thing of time series vector autoregression is that we could test 'causality' in some sense. This test is first proposed by Granger (1969), and therefore we refer it Granger causality.

We will restrict our discussion to a system of two variables, x and y. y is said to *Granger-cause* x if current or lagged values of y helps to predict future values of x. On the other hand, y fails to Granger-cause x if for all $s > 0$, the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, \ldots)$ is the same as that is based on $(y_t, y_{t-1}, \ldots)$ and $(x_t, x_{t-1}, \ldots)$. If we restrict ourselves to linear functions, x fails to Granger-cause x if

$$\text{MSE}[\hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots)] = \text{MSE}[\hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots, y_t, y_{t-1}, \ldots)].$$

Equivalently, we can say that x is exogenous in the time series sense with respect to y, or y is not linearly informative about future x.

In the VAR equation, the example we proposed above implies a lower triangular coefficient matrix:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & 0 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} \phi_{11}^p & 0 \\ \phi_{21}^p & \phi_{22}^p \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{11}$$

Or if we use MA representations,

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}(L) & 0 \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \tag{12}$$

where

$$\phi_{ij}(L) = \phi_{ij}^0 + \phi_{ij}^1 L + \phi_{ij}^2 L + \ldots$$

with $\phi_{11}^0 = \phi_{22}^0 = 1$ and $\phi_{21}^0 = 0$. Another implication of Granger causality is stressed by Sims (1972).

Proposition 5 *Consider a linear projection of y_t on past, present and future x's,*

$$y_t = c + \sum_{j=0}^{\infty} b_j x_{t-j} + \sum_{j=1}^{\infty} d_j x_{t+j} + \eta_t, \tag{13}$$

where $E(\eta_t x_\tau) = 0$ for all t and τ. Then y fails to Granger-cause x iff $d_j = 0$ for $j = 1, 2, \ldots$.

Econometric tests of whether the series y Granger causes x can be based on any of the three implications (11), (12), or (13). The simplest test is to estimate the regression which is based on (11),

$$x_t = c_1 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{p} \beta_i y_{t-j} + u_t$$

using OLS and then conduct a F-test of the null hypothesis

$$H_0 : \ \beta_1 = \beta_2 = \ldots = \beta_p = 0.$$

Note: we have to be aware of that *Granger causality* does not equal to what we usually mean by *causality*. For instance, even if x_1 does not cause x_2, it may still help to predict x_2, and thus Granger-causes x_2 if changes in x_1 precedes that of x_2 for some reason. A naive example is that we observe that a dragonfly flies much lower before a rain storm, due to the lower air pressure. We know that dragonflies do not cause a rain storm, but it does help to predict a rain storm, thus Granger-causes a rain storm.

Reading: Hamilton Ch. 10, 11, 14.