# Ch. 18 Vector Time Series 

(May 4, 2017)


## 1 Introduction

In dealing with economic variables often the value of one variables is not only related to its predecessors in time but, in addition, it depends on past values of other variables. This naturally extends the concept of univariate stochastic process to vector time series analysis. This chapter describes the dynamic interactions among a set of variables collected in an $(k \times 1)$ vector $\mathbf{y}_{t}$.

## Definition.

Let $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ be a probability space and $\mathcal{T}$ an index set of real numbers and define the $k$-dimensional vector function $\mathbf{y}(\cdot, \cdot)$ by $\mathbf{y}(\cdot, \cdot): \mathcal{S} \times \mathcal{T} \rightarrow \mathbb{R}^{k}$. The ordered sequence of random vector $\{\mathbf{y}(\cdot, t), t \in \mathcal{T}\}$ is called a $k$-dimensional vector stochastic process.

### 1.1 First Two Moments of Stationary Vector Time Series

From now on in this chapter we follows convention to use $\mathbf{y}_{t}$ in stead of $\mathbf{y}(\cdot, t)$ to indicate that we are considering discrete vector time series. The first two moments of
a vector time series $\mathbf{y}_{t}$ are

$$
E\left(\mathbf{y}_{t}\right)=\boldsymbol{\mu}_{t}=\left[\begin{array}{c}
\mu_{1 t} \\
\mu_{2 t} \\
\cdot \\
\cdot \\
\cdot \\
\mu_{k t}
\end{array}\right], \text { and } \boldsymbol{\Gamma}_{t, j}=E\left[\left(\mathbf{y}_{t}-\boldsymbol{\mu}_{t}\right)\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}_{t-j}\right)^{\prime}\right] \quad \text { for all } t \in \mathcal{T}
$$

If neither $\boldsymbol{\mu}_{t}$ and $\boldsymbol{\Gamma}_{t, j}$ are function of $t$, that is, $\boldsymbol{\mu}_{t}=\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}_{t, j}=\boldsymbol{\Gamma}_{j}$, then we say that $\mathbf{y}_{t}$ is a covariance-stationary vector process.

Note that although $\gamma_{j}=\gamma_{-j}$ for a scalar stationary process, the same is not true of a vector process:

$$
\boldsymbol{\Gamma}_{j} \neq \boldsymbol{\Gamma}_{-j}
$$

Instead, the correct relation is

$$
\boldsymbol{\Gamma}_{j}^{\prime}=\boldsymbol{\Gamma}_{-j}
$$

since

$$
\begin{aligned}
\boldsymbol{\Gamma}_{j} & =E\left[\left(\mathbf{y}_{t+j}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{(t+j)-j}-\boldsymbol{\mu}\right)^{\prime}\right] \\
& =E\left[\left(\mathbf{y}_{t+j}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime}\right]
\end{aligned}
$$

and taking transpose,

$$
\begin{aligned}
\boldsymbol{\Gamma}_{j}^{\prime} & =E\left[\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t+j}-\boldsymbol{\mu}\right)^{\prime}\right] \\
& =E\left[\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t-(-j)}-\boldsymbol{\mu}\right)^{\prime}\right]=\boldsymbol{\Gamma}_{-j}
\end{aligned}
$$

### 1.2 Vector White Noise Process

Like the univariate time series analysis, the basic building element of a vector $A R M A$ model is the vector white noise process.

## $\mathfrak{D e f i n i t i o n .}$

A $k \times 1$ vector process $\left\{\varepsilon_{t}, t \in \mathcal{T}\right\}$ is said to be a white-noise process if
(i). $\quad E\left(\varepsilon_{t}\right)=\mathbf{0}$;
(ii). $E\left(\varepsilon_{t} \varepsilon_{\tau}^{\prime}\right)=\left\{\begin{array}{cc}\boldsymbol{\Omega} & \text { if } t=\tau, \\ \mathbf{0} & \text { if } t \neq \tau,\end{array}\right.$
where $\boldsymbol{\Omega}$ is an $(k \times k)$ symmetric positive definite matrix.

It is important to note that in general $\boldsymbol{\Omega}$ is not necessary a diagonal matrix, since it is the contemporaneous correlation among variables that called for the needs of vector time series analysis.

### 1.3 Vector MA(q) Process

A vector moving average process of order $q$ takes the form

$$
\mathbf{y}_{t}=\boldsymbol{\mu}+\varepsilon_{t}+\Theta_{1} \varepsilon_{t-1}+\Theta_{2} \varepsilon_{t-2}+\ldots+\Theta_{q} \varepsilon_{t-q},
$$

where $\boldsymbol{\varepsilon}_{t}$ is a vector white noise process and $\boldsymbol{\Theta}_{j}$ denotes an $(k \times k)$ matrix of $M A$ coefficients for $j=1,2, \ldots, q$. The mean of $\mathbf{y}_{t}$ is $\boldsymbol{\mu}$, and the variance is

$$
\begin{aligned}
\Gamma_{0}= & E\left[\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime}\right] \\
= & E\left[\varepsilon_{t} \varepsilon_{t}^{\prime}\right]+\Theta_{1} E\left[\varepsilon_{t-1} \varepsilon_{t-1}^{\prime}\right] \Theta_{1}^{\prime}+\Theta_{2} E\left[\varepsilon_{t-2} \varepsilon_{t-2}^{\prime}\right] \Theta_{2}^{\prime} \\
& +\ldots+\Theta_{q} E\left[\varepsilon_{t-q} \varepsilon_{t-q}^{\prime}\right] \Theta_{q}^{\prime} \\
= & \mathbf{\Omega}+\Theta_{1} \Omega \Theta_{1}^{\prime}+\Theta_{2} \Omega \Theta_{2}^{\prime}+\ldots+\Theta_{q} \Omega \Theta_{q}^{\prime},
\end{aligned}
$$

with autocovariance (compares with $\gamma_{j}$ of Ch. 14 on p.3) $E\left[\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)^{\prime}\right]$,

$$
\boldsymbol{\Gamma}_{j}=\left\{\begin{array}{cc}
\boldsymbol{\Theta}_{j} \boldsymbol{\Omega}+\boldsymbol{\Theta}_{j+1} \boldsymbol{\Omega} \boldsymbol{\Theta}_{1}^{\prime}+\boldsymbol{\Theta}_{j+2} \boldsymbol{\Omega} \boldsymbol{\Theta}_{2}^{\prime}+\ldots+\boldsymbol{\Theta}_{q} \boldsymbol{\Omega} \boldsymbol{\Theta}_{q-j}^{\prime} & \text { for } j=1,2, \ldots, q \\
\boldsymbol{\Omega} \boldsymbol{\Theta}_{-j}^{\prime}+\boldsymbol{\Theta}_{1} \boldsymbol{\Omega} \boldsymbol{\Theta}_{-j+1}^{\prime}+\boldsymbol{\Theta}_{2} \boldsymbol{\Omega} \boldsymbol{\Theta}_{-j+2}^{\prime}+\ldots+\boldsymbol{\Theta}_{q+j} \boldsymbol{\Omega} \boldsymbol{\Theta}_{q}^{\prime} & \text { for } j=-1,-2, \ldots,-q \\
\mathbf{0} & \text { for }|j|>q,
\end{array}\right.
$$

where $\boldsymbol{\Theta}_{0}=\mathbf{I}_{k}$. Thus any vector $M A(q)$ process is covariance-stationary.

### 1.4 Vector $M A(\infty)$ Process

The vector $M A(\infty)$ process is written

$$
\mathbf{y}_{t}=\boldsymbol{\mu}+\varepsilon_{t}+\mathbf{\Psi}_{1} \varepsilon_{t-1}+\mathbf{\Psi}_{2} \varepsilon_{t-2}+\ldots
$$

where $\boldsymbol{\varepsilon}_{t}$ is a vector white noise process and $\boldsymbol{\Psi}_{j}$ denotes an $(k \times k)$ matrix of $M A$ coefficients. Many of the results for scalar $M A(\infty)$ process with absolutely summable coefficients go through for vector processes as well.

## Definition.

For an $(n \times m)$ matrix $\mathbf{H}$, the sequence of matrices $\left\{\mathbf{H}_{s}\right\}_{s=0}^{\infty}$ is absolutely summable if each of its elements forms an absolutely summable scalar sequence.

## Example:

If $\psi_{i j}^{(s)}$ denotes the row $i$, column $j$ element of the moving average parameters matrix $\boldsymbol{\Psi}_{s}$ associated with lag $s$, then the sequence $\left\{\boldsymbol{\Psi}_{s}\right\}_{s=0}^{\infty}$ is absolutely if

$$
\sum_{s=0}^{\infty}\left|\psi_{i j}^{(s)}\right|<\infty \quad \text { for } i=1,2, \ldots, k \text { and } j=1,2, \ldots, k
$$

## Theorem.

Consider the infinite vector moving average process

$$
\mathbf{y}_{t}=\boldsymbol{\mu}+\varepsilon_{t}+\mathbf{\Psi}_{1} \varepsilon_{t-1}+\mathbf{\Psi}_{2} \varepsilon_{t-2}+\ldots
$$

where $\varepsilon_{t}$ is a vector white noise process and $\left\{\boldsymbol{\Psi}_{l}\right\}_{l=0}^{\infty}$ is absolutely summable. Let $y_{i t}$ denote the $i$ th element of $\mathbf{y}_{t}$, and let $\mu_{i}$ denote the $i$ th element of $\boldsymbol{\mu}$. Then
(a). the autocovariance between the $i$ th variable at time $t$ and the $j$ th variable $s$ period earlier, $E\left(y_{i t}-\mu_{i}\right)\left(y_{j, t-s}-\mu_{j}\right)$, exist and is given by the row $i$, column $j$ element of

$$
\boldsymbol{\Gamma}_{s}=\sum_{v=0}^{\infty} \boldsymbol{\Psi}_{s+v} \boldsymbol{\Omega} \boldsymbol{\Psi}_{v}^{\prime} \quad \text { for } s=0,1,2, \ldots
$$

(b). the sequence of matrices $\left\{\boldsymbol{\Gamma}_{s}\right\}_{s=0}^{\infty}$ is absolutely summable.

## $\mathfrak{P r o o f}$.:

(a). By definition

$$
\boldsymbol{\Gamma}_{s}=E\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t-s}-\boldsymbol{\mu}\right)^{\prime}
$$

or

$$
\begin{aligned}
\boldsymbol{\Gamma}_{s}= & E\left[\varepsilon_{t}+\boldsymbol{\Psi}_{1} \varepsilon_{t-1}+\mathbf{\Psi}_{2} \varepsilon_{t-2}+\ldots+\boldsymbol{\Psi}_{s} \varepsilon_{t-s}+\boldsymbol{\Psi}_{s+1} \varepsilon_{t-s-1}+\ldots .\right] \\
& {\left[\varepsilon_{t-s}+\mathbf{\Psi}_{1} \varepsilon_{t-s-1}+\mathbf{\Psi}_{2} \varepsilon_{t-s-2}+\ldots .\right]^{\prime} } \\
= & \boldsymbol{\Psi}_{s} \boldsymbol{\Omega} \boldsymbol{\Psi}_{0}^{\prime}+\boldsymbol{\Psi}_{s+1} \boldsymbol{\Omega} \mathbf{\Psi}_{1}^{\prime}+\mathbf{\Psi}_{s+2} \boldsymbol{\Omega} \mathbf{\Psi}_{2}^{\prime}+\ldots \\
= & \sum_{v=0}^{\infty} \boldsymbol{\Psi}_{s+v} \boldsymbol{\Omega} \boldsymbol{\Psi}_{v}^{\prime} \text { for } s=0,1,2, \ldots
\end{aligned}
$$

The row $i$, column $j$ element of $\boldsymbol{\Gamma}_{s}$ is therefore the autocovariance between the $i$ th variable at time $t$ and the $j$ th variable $s$ period earlier, $E\left(y_{i t}-\mu_{i}\right)\left(y_{j, t-s}-\mu_{j}\right)$.
(b).

## 2 Vector Autoregressive Process, $V A R$

A $k$-variate $p$ th order vector autoregression, denoted $\operatorname{VAR}(p)$ is written as;

$$
\begin{equation*}
\mathbf{y}_{t}=\mathbf{c}+\boldsymbol{\Phi}_{1} \mathbf{y}_{t-1}+\boldsymbol{\Phi}_{2} \mathbf{y}_{t-2}+\ldots+\boldsymbol{\Phi}_{p} \mathbf{y}_{t-p}+\boldsymbol{\varepsilon}_{t} \tag{18-1}
\end{equation*}
$$

where $\mathbf{c}$ denotes an $(k \times 1)$ vector of constants and $\boldsymbol{\Phi}_{j}$ an $(k \times k)$ matrix of autoregressive coefficients for $j=1,2, \ldots, p$ and $\varepsilon_{t}$ is a vector white noise process.

### 2.1 Population Characteristics

Let $c_{i}$ denotes the $i$ th element of the vector $\mathbf{c}$ and let $\phi_{i j}^{(s)}$ denote the row $i$, column $j$ element of the matrix $\boldsymbol{\Phi}_{s}$, then the first row of the vector system in (18-1) specifies that

$$
\begin{aligned}
y_{1 t}= & c_{1}+\phi_{11}^{(1)} y_{1, t-1}+\phi_{12}^{(1)} y_{2, t-1}+\ldots+\phi_{1 k}^{(1)} y_{k, t-1} \\
& +\phi_{11}^{(2)} y_{1, t-2}+\phi_{12}^{(2)} y_{2, t-2}+\ldots+\phi_{1 k}^{(2)} y_{k, t-2} \\
& +\ldots .+\phi_{11}^{(p)} y_{1, t-p}+\phi_{12}^{(p)} y_{2, t-p}+\ldots+\phi_{1 k}^{(p)} y_{k, t-p}+\varepsilon_{1 t} .
\end{aligned}
$$

Thus, a vector autoregression is a system in which each variable is regressed on a constant and $p$ of its own lags as well as on $p$ lags of each of the other $(k-1)$ variables in the $V A R$. Note that each regression has the same explanatory variables.

Using lag operator notation, i.e., $L \mathbf{y}_{t}=\mathbf{y}_{t-1},{ }^{1}(18-1)$ can be written in this form

$$
\left[\mathbf{I}_{k}-\boldsymbol{\Phi}_{1} L-\boldsymbol{\Phi}_{2} L^{2}-\ldots-\boldsymbol{\Phi}_{p} L^{p}\right] \mathbf{y}_{t}=\mathbf{c}+\boldsymbol{\varepsilon}_{t}
$$

or

$$
\begin{equation*}
\mathbf{\Phi}(L) \mathbf{y}_{t}=\mathbf{c}+\varepsilon_{t} \tag{18-2}
\end{equation*}
$$

Here $\boldsymbol{\Phi}(L)$ indicate an $k \times k$ matrix polynomial in the lag operator $L$. The row $i$, column $j$ elements of $\boldsymbol{\Phi}(L)$ is a scalar polynomial in $L$ :

$$
\boldsymbol{\Phi}(L)_{i j}=\left[\delta_{i j}-\phi_{i j}^{(1)} L^{1}-\phi_{i j}^{(2)} L^{2}-\ldots-\phi_{i j}^{(p)} L^{p}\right],
$$

where $\delta_{i j}$ is unity if $i=j$ and zero otherwise.

[^0]
## Example.

For $p=k=2$,

$$
\begin{aligned}
\boldsymbol{\Phi}(L) & =\mathbf{I}_{2}-\boldsymbol{\Phi}_{1} L-\mathbf{\Phi}_{2} L^{2} \\
& =\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]-\left[\begin{array}{ll}
\phi_{11}^{(1)} & \phi_{12}^{(1)} \\
\phi_{21}^{(1)} & \phi_{22}^{(1)}
\end{array}\right] L-\left[\begin{array}{ll}
\phi_{11}^{(2)} & \phi_{12}^{(2)} \\
\phi_{21}^{(2)} & \phi_{22}^{(2)}
\end{array}\right] L^{2} \\
& =\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]-\left[\begin{array}{ll}
\phi_{11}^{(1)} L & \phi_{12}^{(1)} L \\
\phi_{21}^{(1)} L & \phi_{22}^{(1)} L
\end{array}\right]-\left[\begin{array}{ll}
\phi_{11}^{(2)} L^{2} & \phi_{12}^{(2)} L^{2} \\
\phi_{21}^{(2)} L^{2} & \phi_{22}^{(2)} L^{2}
\end{array}\right] .
\end{aligned}
$$

If the $\operatorname{VAR}(p)$ process is stationary, we can take expectation of both side of (18-1) to calculate the mean $\boldsymbol{\mu}$ of the process:

$$
\boldsymbol{\mu}=\mathbf{c}+\boldsymbol{\Phi}_{1} \boldsymbol{\mu}+\boldsymbol{\Phi}_{2} \boldsymbol{\mu}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\mu}
$$

or

$$
\boldsymbol{\mu}=\left(\mathbf{I}_{k}-\mathbf{\Phi}_{1}-\mathbf{\Phi}_{2}-\ldots-\mathbf{\Phi}_{p}\right)^{-1} \mathbf{c} .
$$

Equation (18-1) can then be written in terms of deviations from the mean as

$$
\begin{equation*}
\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)=\boldsymbol{\Phi}_{1}\left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)+\boldsymbol{\Phi}_{2}\left(\mathbf{y}_{t-2}-\boldsymbol{\mu}\right)+\ldots+\boldsymbol{\Phi}_{p}\left(\mathbf{y}_{t-p}-\boldsymbol{\mu}\right)+\boldsymbol{\varepsilon}_{t} . \tag{18-3}
\end{equation*}
$$

### 2.1.1 Conditions for Stationarity

As in the case of the univariate $A R(p)$ process, it is helpful to rewrite (18-3) in terms of a $\operatorname{VAR}(1)$ process. Toward this end, define

$$
\boldsymbol{\xi}_{t}=\left[\begin{array}{c}
\mathbf{y}_{t}-\boldsymbol{\mu}  \tag{18-4}\\
\mathbf{y}_{t-1}-\boldsymbol{\mu} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{y}_{t-p+1}-\boldsymbol{\mu}
\end{array}\right]_{(k p \times 1)}
$$

$$
\mathbf{F}=\left[\begin{array}{ccccccc}
\boldsymbol{\Phi}_{1} & \boldsymbol{\Phi}_{2} & \boldsymbol{\Phi}_{3} & . & . & . & \boldsymbol{\Phi}_{p-1}  \tag{18-5}\\
\mathbf{I}_{k} & \mathbf{0} & \mathbf{0} & . & . & \boldsymbol{\Phi}_{p} \\
\mathbf{0} & \mathbf{I}_{k} & \mathbf{0} & . & . & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & \mathbf{0} & \mathbf{0} \\
. & . & . & . & . & . & . \\
. & . & . \\
\mathbf{0} & .0 & . & . & . & . & . \\
\mathbf{0} & . & . & \mathbf{I}_{k} & \mathbf{0}
\end{array}\right]_{(k p \times k p)}
$$

and

$$
\mathbf{v}_{t}=\left[\begin{array}{c}
\varepsilon_{t} \\
\mathbf{0} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{0}
\end{array}\right]_{(k p \times 1)}
$$

The $V A R(p)$ in (18-3) can then be rewritten as the following $\operatorname{VAR}(1)$ :

$$
\begin{equation*}
\boldsymbol{\xi}_{t}=\mathbf{F} \boldsymbol{\xi}_{t-1}+\mathbf{v}_{t} \tag{18-6}
\end{equation*}
$$

which implies

$$
\begin{equation*}
\boldsymbol{\xi}_{t+s}=\mathbf{v}_{t+s}+\mathbf{F} \mathbf{v}_{t+s-1}+\mathbf{F}^{2} \mathbf{v}_{t+s-2}+\ldots+\mathbf{F}^{s-1} \mathbf{v}_{t+1}+\mathbf{F}^{s} \boldsymbol{\xi}_{t} \tag{18-7}
\end{equation*}
$$

where

$$
E\left(\mathbf{v}_{t} \mathbf{v}_{s}^{\prime}\right)=\left\{\begin{array}{c}
\mathbf{Q} \quad \text { for } t=s, \\
\mathbf{0} \quad \text { otherwise },
\end{array}\right.
$$

and

$$
\mathbf{Q}=\left[\begin{array}{cccccc}
\Omega & 0 & . & . & . & 0 \\
0 & 0 & . & . & . & 0 \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & 0 & . & . & . & 0
\end{array}\right]
$$

In order for the process to be covariance-stationary, the consequence of any given $\varepsilon_{t}$ must eventually die out. If the eigenvalues of $\mathbf{F}$ all lie inside the unit circle, then the $V A R$ turns out to be covariance-stationary.

## $\mathfrak{P r o p o s i t i o n .}$

The eigenvalues of the matrix $\mathbf{F}$ in (18-5) also satisfy

$$
\begin{equation*}
\left|\mathbf{I}_{k} \lambda^{p}-\mathbf{\Phi}_{1} \lambda^{p-1}-\boldsymbol{\Phi}_{2} \lambda^{p-2}-\ldots-\boldsymbol{\Phi}_{p}\right|=0 . \tag{18-8}
\end{equation*}
$$

Hence, a $\operatorname{VAR}(p)$ is covariance-stationary as long as $|\lambda|<1$ for all the $k \times p$ eigenvalues of $\lambda$ satisfying (18-8). Equivalently, the $V A R$ is stationary if all values $z$ satisfying

$$
\left|\mathbf{I}_{k}-\boldsymbol{\Phi}_{1} z-\boldsymbol{\Phi}_{2} z^{2}-\ldots-\boldsymbol{\Phi}_{p} z^{p}\right|=0
$$

lie outside the unit circle.

### 2.1.2 Vector $M A(\infty)$ Representation

The first $k$ rows of the vector system represented in (18-7) constitute a vector system:

$$
\begin{aligned}
\mathbf{y}_{t+s}= & \boldsymbol{\mu}+\boldsymbol{\varepsilon}_{t+s}+\mathbf{\Psi}_{1} \boldsymbol{\varepsilon}_{t+s-1}+\mathbf{\Psi}_{2} \varepsilon_{t+s-2}+\ldots+\mathbf{\Psi}_{s-1} \varepsilon_{t+1} \\
& +\mathbf{F}_{11}^{s}\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)+\mathbf{F}_{12}^{s}\left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)+\ldots+\mathbf{F}_{1 p}^{s}\left(\mathbf{y}_{t-p+1}-\boldsymbol{\mu}\right)
\end{aligned}
$$

Here $\boldsymbol{\Psi}_{j}=\mathbf{F}_{11}^{(j)}$ and $\mathbf{F}_{11}^{(j)}$ denotes the upper left block of $\mathbf{F}^{j}$, where $\mathbf{F}^{j}$ is the matrix $\mathbf{F}$ raised to the $j$ th power.

If the eigenvalues of $\mathbf{F}$ all lie inside the unit circle, then $\mathbf{F}^{s} \rightarrow \mathbf{0}$ as $s \rightarrow \infty$ and $\mathbf{y}_{t}$ can be expressed as convergent sum of the history of $\varepsilon$ :

$$
\begin{equation*}
\mathbf{y}_{t}=\boldsymbol{\mu}+\boldsymbol{\varepsilon}_{t}+\boldsymbol{\Psi}_{1} \varepsilon_{t-1}+\boldsymbol{\Psi}_{2} \varepsilon_{t-2}+\boldsymbol{\Psi}_{3} \varepsilon_{t-3}+\ldots=\boldsymbol{\mu}+\boldsymbol{\Psi}(L) \varepsilon_{t} . \tag{18-9}
\end{equation*}
$$

The moving average matrices $\boldsymbol{\Psi}_{j}$ could equivalently be calculated as follows. The operator $\boldsymbol{\Phi}(L)\left(=\mathbf{I}_{k}-\boldsymbol{\Phi}_{1} L-\boldsymbol{\Phi}_{2} L^{2}-\ldots-\boldsymbol{\Phi}_{p} L^{p}\right)$ at (18-2) and $\boldsymbol{\Psi}(L)$ at (18-9) are related by

$$
\boldsymbol{\Psi}(L)=[\boldsymbol{\Phi}(L)]^{-1},
$$

requiring that

$$
\left[\mathbf{I}_{k}-\boldsymbol{\Phi}_{1} L-\boldsymbol{\Phi}_{2} L^{2}-\ldots-\boldsymbol{\Phi}_{p} L^{p}\right]\left[\mathbf{I}_{k}+\mathbf{\Psi}_{1} L+\boldsymbol{\Psi}_{2} L^{2}+\ldots\right]=\mathbf{I}_{k}
$$

Setting the coefficient on $L^{1}$ equal to the zero matrix produces

$$
\boldsymbol{\Psi}_{1}-\boldsymbol{\Phi}_{1}=\mathbf{0} .
$$

Similarly, setting the coefficient on $L^{2}$ equal to zero gives

$$
\boldsymbol{\Psi}_{2}=\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{1}+\boldsymbol{\Phi}_{2}
$$

and in general for $L^{s}$,

$$
\begin{equation*}
\boldsymbol{\Psi}_{s}=\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{s-1}+\boldsymbol{\Phi}_{2} \boldsymbol{\Psi}_{s-2}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\Psi}_{s-p} \quad \text { for } s=1,2, \ldots \tag{18-10}
\end{equation*}
$$

with $\boldsymbol{\Psi}_{0}=\mathbf{I}_{k}$ and $\boldsymbol{\Psi}_{s}=\mathbf{0}$ for $s<0$.
Note that the innovation in the $M A(\infty)$ representation in (18-9) is $\varepsilon_{t}$, the fundamental innovation for $\mathbf{y}_{t}$. There are alternative moving average representations based on vector white noise process other than $\boldsymbol{\varepsilon}_{t}$. Let $\mathbf{H}$ denote a nonsingular $(k \times k)$ matrix, and define

$$
\mathbf{u}_{t}=\mathbf{H} \varepsilon_{t} \quad(\text { Linear Combination of } \varepsilon) .
$$

Then certainly $\mathbf{u}_{t}$ is white noise:

$$
\begin{aligned}
E\left(\mathbf{u}_{t}\right) & =\mathbf{0} \text { and } \\
E\left(\mathbf{u}_{t} \mathbf{u}_{\tau}^{\prime}\right) & =\left\{\begin{array}{cl}
\mathbf{H} \Omega \mathbf{H}^{\prime} & \text { for } t=\tau \\
\mathbf{0} & \text { for } t \neq \tau
\end{array} .\right.
\end{aligned}
$$

Moreover, from (18-9) we could write

$$
\begin{aligned}
\mathbf{y}_{t} & =\boldsymbol{\mu}+\mathbf{H}^{-1} \mathbf{H} \varepsilon_{t}+\mathbf{\Psi}_{1} \mathbf{H}^{-1} \mathbf{H} \varepsilon_{t-1}+\mathbf{\Psi}_{2} \mathbf{H}^{-1} \mathbf{H} \varepsilon_{t-2}+\mathbf{\Psi}_{3} \mathbf{H}^{-1} \mathbf{H} \varepsilon_{t-3}+\ldots \\
& =\boldsymbol{\mu}+\mathbf{J}_{0} \mathbf{u}_{t}+\mathbf{J}_{1} \mathbf{u}_{t-1}+\mathbf{J}_{2} \mathbf{u}_{t-2}+\mathbf{J}_{3} \mathbf{u}_{t-3}+\ldots
\end{aligned}
$$

where

$$
\begin{equation*}
\mathbf{J}_{s}=\mathbf{\Psi}_{s} \mathbf{H}^{-1} \tag{18-11}
\end{equation*}
$$

One possible choice of $\mathbf{H}$ could be any matrix that diagonalize $\Omega$,

$$
\mathbf{H} \Omega \mathbf{H}^{\prime}=\mathbf{D},
$$

with $\mathbf{D}$ a diagonal matrix. For such a choice of $\mathbf{H}$, the element of $\mathbf{u}_{t}$ are uncorrelated with one another:

$$
E\left(\mathbf{u}_{t} \mathbf{u}_{t}^{\prime}\right)=\mathbf{H} \boldsymbol{\Omega} \mathbf{H}^{\prime}=\mathbf{D} .
$$

Thus, it is always possible to write a stationary $V A R(p)$ process as a infinite moving average of a white noise vector $\mathbf{u}_{t}$ whose elements are mutually uncorrelated.

### 2.1.3 Computation of Autocovariances of an Stationary $\operatorname{VAR}(p)$ Process

We now consider to express the second moments for $\mathbf{y}_{t}$ following a $\operatorname{VAR}(p)$. Recall that as in the univariate $A R(p)$ process, the Yule-Walker equation are obtained by postmultiplying (18-3) with $\left(\mathbf{y}_{t-j}-\boldsymbol{\mu}\right)^{\prime}$ and taking expectations. For $j=0$, using $\boldsymbol{\Gamma}_{j}=\boldsymbol{\Gamma}_{-j}^{\prime}$,

$$
\begin{aligned}
\boldsymbol{\Gamma}_{0}= & E\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime} \\
= & \mathbf{\Phi}_{1} E\left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime}+\mathbf{\Phi}_{2} E\left(\mathbf{y}_{t-2}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime} \\
& +\ldots+\boldsymbol{\Phi}_{p} E\left(\mathbf{y}_{t-p}-\boldsymbol{\mu}\right)\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime}+E \boldsymbol{\varepsilon}_{t}\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime} \\
= & \mathbf{\Phi}_{1} \boldsymbol{\Gamma}_{-1}+\boldsymbol{\Phi}_{2} \boldsymbol{\Gamma}_{-2}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\Gamma}_{-p}+\boldsymbol{\Omega} \\
= & \mathbf{\Phi}_{1} \boldsymbol{\Gamma}_{1}^{\prime}+\mathbf{\Phi}_{2} \boldsymbol{\Gamma}_{2}^{\prime}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\Gamma}_{p}^{\prime}+\boldsymbol{\Omega}
\end{aligned}
$$

and for $j>0$,

$$
\begin{equation*}
\boldsymbol{\Gamma}_{j}=\boldsymbol{\Phi}_{1} \boldsymbol{\Gamma}_{j-1}+\boldsymbol{\Phi}_{2} \boldsymbol{\Gamma}_{j-2}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\Gamma}_{j-p} \tag{18-12}
\end{equation*}
$$

These equations may be used to compute the $\boldsymbol{\Gamma}_{j}$ recursively for $j \geq p$ if $\boldsymbol{\Phi}_{1}, \ldots, \boldsymbol{\Phi}_{p}$ and $\boldsymbol{\Gamma}_{p-1}, \ldots, \boldsymbol{\Gamma}_{0}$ are known.

To obtain the initial $\boldsymbol{\Gamma}_{p-1}, \ldots, \boldsymbol{\Gamma}_{0}$, we proceed as follows. Let $\boldsymbol{\xi}_{t}$ be as defined in (18-4) and let $\boldsymbol{\Sigma}$ denote the variance of $\boldsymbol{\xi}_{t}$,

$$
\boldsymbol{\Sigma}=E\left(\boldsymbol{\xi}_{t} \boldsymbol{\xi}_{t}^{\prime}\right)
$$

$$
=E\left\{\left[\begin{array}{c}
\mathbf{y}_{t}-\boldsymbol{\mu} \\
\mathbf{y}_{t-1}-\boldsymbol{\mu} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{y}_{t-p+1}-\boldsymbol{\mu}
\end{array}\right] \times\left[\begin{array}{lll}
\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)^{\prime} & \left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)^{\prime} & \cdot \cdot\left(\mathbf{y}_{t-p+1}-\boldsymbol{\mu}\right)^{\prime}
\end{array}\right]^{\prime}\right\}
$$

$$
=\left[\begin{array}{cccccc}
\boldsymbol{\Gamma}_{0} & \boldsymbol{\Gamma}_{1} & \cdot & \cdot & \boldsymbol{\Gamma}_{p-1} \\
\boldsymbol{\Gamma}_{1}^{\prime} & \boldsymbol{\Gamma}_{0} & \cdot & \cdot & \cdot & \boldsymbol{\Gamma}_{p-2} \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
\boldsymbol{\Gamma}_{p-1}^{\prime} & \boldsymbol{\Gamma}_{p-2}^{\prime} & \cdot & \cdot & \boldsymbol{\Gamma}_{0}
\end{array}\right]
$$

Post-multiplying (18-4) by its own transpose and taking expectation gives

$$
E\left[\boldsymbol{\xi}_{t} \boldsymbol{\xi}_{t}^{\prime}\right]=E\left[\left(\mathbf{F} \boldsymbol{\xi}_{t-1}+\mathbf{v}_{t}\right)\left(\mathbf{F} \boldsymbol{\xi}_{t-1}+\mathbf{v}_{t}\right)^{\prime}\right]=\mathbf{F} E\left(\boldsymbol{\xi}_{t-1} \boldsymbol{\xi}_{t-1}^{\prime}\right) \mathbf{F}^{\prime}+E\left(\mathbf{v}_{t} \mathbf{v}_{t}^{\prime}\right)
$$

or

$$
\begin{equation*}
\boldsymbol{\Sigma}=\mathbf{F} \boldsymbol{\Sigma} \mathbf{F}^{\prime}+\mathbf{Q} \tag{18-13}
\end{equation*}
$$

From the result of vec operator on p. 13 of Ch. 1, we have

$$
\begin{equation*}
\operatorname{vec}(\boldsymbol{\Sigma})=(\mathbf{F} \otimes \mathbf{F}) \cdot \operatorname{vec}(\boldsymbol{\Sigma})+\operatorname{vec}(\mathbf{Q})=\mathcal{A} \cdot \operatorname{vec}(\boldsymbol{\Sigma})+\operatorname{vec}(\mathbf{Q}) \tag{18-14}
\end{equation*}
$$

where

$$
\mathcal{A} \equiv(\mathbf{F} \otimes \mathbf{F}) .
$$

Let $r=k p$, so that $\mathbf{F}$ is an $(r \times r)$ matrix and $\mathcal{A}$ is an $\left(r^{2} \times r^{2}\right)$ matrix. Equation (18-14) has the solution

$$
\begin{equation*}
\operatorname{vec}(\boldsymbol{\Sigma})=\left[\mathbf{I}_{r^{2}}-\mathcal{A}\right]^{-1} \operatorname{vec}(\mathbf{Q}) \tag{18-15}
\end{equation*}
$$

provided that the matrix $\left[\mathbf{I}_{r^{2}}-\mathcal{A}\right]$ is nonsingular. Thus, the $\boldsymbol{\Gamma}_{j}, j=-p+1, \ldots, p-1$ are obtained from (18-15).

## Example.

Consider the three-dimensional $V A R(1)$ process

$$
\mathbf{y}_{t}=\mathbf{c}+\left[\begin{array}{ccc}
0.5 & 0 & 0 \\
0.1 & 0.1 & 0.3 \\
0 & 0.2 & 0.3
\end{array}\right] \mathbf{y}_{t-1}+\varepsilon_{t}
$$

with $E\left(\varepsilon_{t} \varepsilon_{t}^{\prime}\right)=\boldsymbol{\Omega}=\left[\begin{array}{ccc}2.25 & 0 & 0 \\ 0 & 1.0 & 0.5 \\ 0 & 0.5 & 0.74\end{array}\right]$.
For this process the reverse characteristic polynomial is

$$
\begin{aligned}
\operatorname{det}\left[\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right]-\left[\begin{array}{ccc}
0.5 & 0 & 0 \\
0.1 & 0.1 & 0.3 \\
0 & 0.2 & 0.3
\end{array}\right]\right] & =\operatorname{det}\left[\begin{array}{ccc}
1-0.5 z & 0 & 0 \\
-0.1 z & 1-0.1 z & -0.3 z \\
0 & -0.2 z & 1-0.3 z
\end{array}\right] \\
& =(1-0.5 z)\left(1-0.4 z-0.03 z^{2}\right) .
\end{aligned}
$$

The roots of this polynomial are easily seen to be

$$
z_{1}=2, \quad z_{2}=2.1525, \quad z_{3}=-15.4858
$$

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

We next obtain the $M A(\infty)$ coefficients from this $V A R(1)$ process using the relation as described in equation (18-10).

$$
\begin{aligned}
\boldsymbol{\Psi}_{0}=\mathbf{I}_{3}, \boldsymbol{\Psi}_{1} & =\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{0}=\left[\begin{array}{ccc}
0.5 & 0 & 0 \\
0.1 & 0.1 & 0.3 \\
0 & 0.2 & 0.3
\end{array}\right], \boldsymbol{\Psi}_{2}=\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{1}=\left[\begin{array}{ccc}
0.25 & 0 & 0 \\
0.01 & 0.07 & 0.12 \\
0.02 & 0.08 & 0.15
\end{array}\right], \\
\boldsymbol{\Psi}_{3} & =\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{2}=\left[\begin{array}{ccc}
0.125 & 0 & 0 \\
0.037 & 0.031 & 0.057 \\
0.018 & 0.038 & 0.069
\end{array}\right], \boldsymbol{\Psi}_{4}=\ldots
\end{aligned}
$$

We finally calculate the autocovariance of this process. From (18-15) we have

$$
\begin{aligned}
\operatorname{vec}(\boldsymbol{\Sigma})= & \operatorname{vec}\left(\boldsymbol{\Gamma}_{0}\right)=\left[\mathbf{I}_{9}-\boldsymbol{\Phi}_{1} \otimes \mathbf{\Phi}_{1}\right]^{-1} \operatorname{vec}(\boldsymbol{\Omega}) \\
& {\left[\begin{array}{cccccccc}
0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.05 & 0.95 & -0.15 & 0 & 0 & 0 & 0 & 0 \\
0 & -0.1 & 0.85 & 0 & 0 & 0 & 0 & 0 \\
0 \\
-0.05 & 0 & 0 & 0.95 & 0 & 0 & -0.15 & 0 \\
-0.01 & -0.01 & -0.03 & -0.01 & 0.99 & -0.03 & -0.03 & -0.03 \\
0 & -0.02 & -0.03 & 0 & -0.02 & 0.97 & 0 & -0.06 \\
-0.09 \\
0 & 0 & 0 & -0.01 & 0 & 0 & 0.85 & 0 \\
0 & 0 & 0 & -0.02 & -0.02 & -0.06 & -0.03 & 0.97 \\
0 & 0 & 0 & 0 & -0.04 & -0.06 & 0 & -0.060 \\
0.09
\end{array}\right]^{-1} } \\
& {\left[\begin{array}{c}
2.25 \\
0 \\
0 \\
0 \\
1.0 \\
0.5 \\
0 \\
0.5 \\
0.74
\end{array}\right]=\left[\begin{array}{c}
3.000 \\
0.161 \\
0.019 \\
0.161 \\
1.172 \\
0.674 \\
0.019 \\
0.674 \\
0.954
\end{array}\right] . }
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \boldsymbol{\Gamma}_{0}=\left[\begin{array}{lll}
3.000 & 0.161 & 0.019 \\
0.161 & 1.172 & 0.674 \\
0.019 & 0.674 & 0.954
\end{array}\right] \quad \boldsymbol{\Gamma}_{1}=\boldsymbol{\Phi}_{1} \boldsymbol{\Gamma}_{0}=\left[\begin{array}{ccc}
1.5000 & 0.080 & 0.009 \\
0.322 & 0.3365 & 0.355 \\
0.038 & 0.437 & 0.421
\end{array}\right] \\
& \boldsymbol{\Gamma}_{2}=\boldsymbol{\Phi}_{1} \boldsymbol{\Gamma}_{1}=\left[\begin{array}{ccc}
0.75 & 0.040 & 0.005 \\
0.194 & 0.173 & 0.163 \\
0.076 & 0.198 & 0.197
\end{array}\right], \boldsymbol{\Gamma}_{3}=\ldots . .
\end{aligned}
$$

## Exercise 1.

Consider the two-dimensional $\operatorname{VAR}(2)$ process

$$
\mathbf{y}_{t}=\mathbf{c}+\left[\begin{array}{ll}
0.5 & 0.1 \\
0.4 & 0.5
\end{array}\right] \mathbf{y}_{t-1}+\left[\begin{array}{cc}
0 & 0 \\
0.25 & 0
\end{array}\right] \mathbf{y}_{t-2}+\varepsilon_{t}
$$

with $E\left(\varepsilon_{t} \varepsilon_{t}^{\prime}\right)=\boldsymbol{\Omega}=\left[\begin{array}{cc}0.09 & 0 \\ 0 & 0.04\end{array}\right]$. Please
(a). Check whether this process is stationary or not.
(b). Find the coefficient matrices in its $M A(\infty)$ representation, $\boldsymbol{\Psi}_{j}, j=0,1,2,3$.
(c). Find its autocovariance matrices $\boldsymbol{\Gamma}_{j}, j=0,1,2,3$.

### 2.1.4 Linear Forecast

From (18-9) we see that $\mathbf{y}_{t-j}$ is a linear function of $\varepsilon_{t-j}, \varepsilon_{t-j-1}, \ldots$ each is uncorrelated with $\varepsilon_{t+1}$ for $j=0,1, \ldots$ It follows that $\varepsilon_{t+1}$ is uncorrelated with $\mathbf{y}_{t-j}$ for any $j \geq 0$. Thus, the linear forecast of $\mathbf{y}_{t+1}$ on the basis of $\mathbf{y}_{t}, \mathbf{y}_{t-1}, \ldots$ is given by

$$
\hat{\mathbf{y}}_{t+1 \mid t}=\boldsymbol{\mu}+\boldsymbol{\Phi}_{1}\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)+\boldsymbol{\Phi}_{2}\left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)+\ldots+\boldsymbol{\Phi}_{p}\left(\mathbf{y}_{t-p+1}-\boldsymbol{\mu}\right),
$$

and $\varepsilon_{t+1}$ can be interpreted as the fundamental innovation for $\mathbf{y}_{t+1}$, that is, the error in forecasting $\mathbf{y}_{t+1}$ on the basis of a linear function of a constant and $\mathbf{y}_{t}, \mathbf{y}_{t-1}, \ldots$ Moregenally, it follows from (18-7) that a forecast of $\mathbf{y}_{t+s}$ on the basis of $\mathbf{y}_{t}, \mathbf{y}_{t-1}, \ldots$ will take the form

$$
\hat{\mathbf{y}}_{t+s \mid t}=\boldsymbol{\mu}+\mathbf{F}_{11}^{(s)}\left(\mathbf{y}_{t}-\boldsymbol{\mu}\right)+\mathbf{F}_{12}^{(s)}\left(\mathbf{y}_{t-1}-\boldsymbol{\mu}\right)+\ldots+\mathbf{F}_{1 p}^{(s)}\left(\mathbf{y}_{t-p+1}-\boldsymbol{\mu}\right) .
$$

### 2.2 Estimation: MLE for an Unrestricted $V A R$

Consider the estimation of the following $k$-dimensional Gaussian $V A R(p)$ process, i.e.

$$
\begin{equation*}
\mathbf{y}_{t}=\mathbf{c}+\boldsymbol{\Phi}_{1} \mathbf{y}_{t-1}+\boldsymbol{\Phi}_{2} \mathbf{y}_{t-2}+\ldots+\boldsymbol{\Phi}_{p} \mathbf{y}_{t-p}+\boldsymbol{\varepsilon}_{t} \tag{18-16}
\end{equation*}
$$

where $\boldsymbol{\varepsilon}_{t} \sim$ i.i.d. $N(\mathbf{0}, \boldsymbol{\Omega})$ and all the roots of $\operatorname{det}\left(\mathbf{I}_{k}-\boldsymbol{\Phi}_{1} z-\mathbf{\Phi}_{2} z^{2}-\ldots-\boldsymbol{\Phi}_{p} z^{p}\right)=0$ lie outside the unit circle. In this case, the parameters to be estimated are $\boldsymbol{\theta}=$ $\left[\mathbf{c}, \boldsymbol{\Phi}_{1}, \ldots, \boldsymbol{\Phi}_{p}, \boldsymbol{\Omega}\right]$.

Suppose that we have a sample of size $(T+p)$, as in the scalar $A R$ process, the simplest approach is to condition on the first $p$ observations (denoted $\mathbf{y}_{-p+1}, \mathbf{y}_{-p+2}, \ldots, \mathbf{y}_{0}$ ) and to base estimation on the last $T$ observations (denoted $\left.\mathbf{y}_{1}, \mathbf{y}_{2}, \ldots, \mathbf{y}_{T}\right)$. The objective is then to form the conditional likelihood

$$
\begin{equation*}
f_{\mathbf{Y}_{T}, \mathbf{Y}_{T-1}, \ldots, \mathbf{Y}_{1} \mid \mathbf{Y}_{0}, \mathbf{Y}_{-1}, \ldots, \mathbf{Y}_{-p+1}}\left(\mathbf{y}_{T}, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_{1} \mid \mathbf{y}_{0}, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1} ; \boldsymbol{\theta}\right) \tag{18-17}
\end{equation*}
$$

and maximize with respect to $\boldsymbol{\theta}, V A R$ are invariably estimated on the basis of the conditional likelihood function (18-17) rather than the full-sample unconditional likelihood. For brevity, we will hereafter refer to (18-17) simply as the "likelihood function" and the value of $\boldsymbol{\theta}$ that maximize (18-17) as the "maximum likelihood estimator".

### 2.2.1 The Conditional Likelihood Function for a Vector Autoregression

The likelihood function is calculated in the same way as for a scalar autoregression. Conditional on the value of $\mathbf{y}$ observed through date $t-1$, the value of $\mathbf{y}$ for date $t$ is equal to a constant

$$
\begin{equation*}
\mathbf{c}+\boldsymbol{\Phi}_{1} \mathbf{y}_{t-1}+\boldsymbol{\Phi}_{2} \mathbf{y}_{t-2}+\ldots+\boldsymbol{\Phi}_{p} \mathbf{y}_{t-p} \tag{18-18}
\end{equation*}
$$

plus a $N(\mathbf{0}, \boldsymbol{\Omega})$ variables. Thus, for $t \geq 1$,

$$
\begin{equation*}
\mathbf{y}_{t} \mid \mathbf{y}_{t-1}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{0}, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1} \sim N\left(\mathbf{c}+\boldsymbol{\Phi}_{1} \mathbf{y}_{t-1}+\boldsymbol{\Phi}_{2} \mathbf{y}_{t-2}+\ldots+\boldsymbol{\Phi}_{p} \mathbf{y}_{t-p}, \boldsymbol{\Omega}\right) \tag{18-19}
\end{equation*}
$$

It will be convenient to use a more compact expression for the conditional mean (18-18). Let $\mathbf{x}_{t}((k p+1) \times 1)$ denote a vector containing a constant terms and $p$ lags of each of the elements of $\mathbf{y}_{t}$ :

$$
\mathbf{x}_{t} \equiv\left[\begin{array}{c}
1 \\
\mathbf{y}_{t-1} \\
\mathbf{y}_{t-2} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{y}_{t-p}
\end{array}\right]
$$

Let $\ddot{\mathbf{X}}_{t}$ denote the $(k \times(k p+1) k)$ matrix

$$
\ddot{\mathbf{X}}_{\mathbf{t}}=\left[\begin{array}{cccccc}
\mathrm{x}_{\mathbf{t}}^{\prime} & \mathbf{0} & . & . & . & \mathbf{0} \\
\mathbf{0} & \mathrm{x}_{\mathbf{t}}^{\prime} & \mathbf{0} & . & . & \mathbf{0} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}_{\mathbf{t}}^{\prime}
\end{array}\right]
$$

and let the $((k p+1) k \times 1)$ vector $\boldsymbol{\beta}=\operatorname{vec}\left[\left(\mathbf{c}, \boldsymbol{\Phi}_{1}, \ldots, \boldsymbol{\Phi}_{p}\right)^{\prime}\right] \equiv \operatorname{vec}\left(\boldsymbol{\theta}^{* \prime}\right)$, it is easy to see that

$$
\begin{equation*}
\ddot{\mathbf{X}}_{\mathbf{t}} \cdot \boldsymbol{\beta}=\mathbf{c}+\boldsymbol{\Phi}_{1} \mathbf{y}_{t-1}+\boldsymbol{\Phi}_{2} \mathbf{y}_{t-2}+\ldots+\boldsymbol{\Phi}_{p} \mathbf{y}_{t-p} \tag{18-20}
\end{equation*}
$$

## Example

To see the above result, for example $k=2$ and $p=1$, then we have

$$
\mathbf{c}=\left[\begin{array}{l}
c_{1} \\
c_{2}
\end{array}\right], \quad \boldsymbol{\Phi}_{1}=\left[\begin{array}{ll}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{array}\right], \quad \mathbf{y}_{t-1}=\left[\begin{array}{l}
y_{1, t-1} \\
y_{2, t-1}
\end{array}\right] .
$$

In this case, $\mathbf{x}_{t}^{\prime}=\left[\begin{array}{lll}1 & y_{1, t-1} & y_{2, t-1}\end{array}\right]$ and $\boldsymbol{\theta}^{*}=\left[\begin{array}{lll}c_{1} & \phi_{11} & \phi_{12} \\ c_{2} & \phi_{21} & \phi_{22}\end{array}\right]$.
Therefore

$$
\ddot{\mathbf{X}}_{\mathbf{t}}=\left[\begin{array}{cccccc}
1 & y_{1, t-1} & y_{2, t-1} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & y_{1, t-1} & y_{2, t-1}
\end{array}\right]
$$

and

$$
\boldsymbol{\beta}=\operatorname{vec}\left(\boldsymbol{\theta}^{* \prime}\right)=\operatorname{vec}\left[\begin{array}{cc}
c_{1} & c_{2} \\
\phi_{11} & \phi_{21} \\
\phi_{12} & \phi_{22}
\end{array}\right]=\left[\begin{array}{c}
c_{1} \\
\phi_{11} \\
\phi_{12} \\
c_{2} \\
\phi_{21} \\
\phi_{22}
\end{array}\right] .
$$

It is easy to see that

$$
\begin{aligned}
\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta} & =\left[\begin{array}{cccccc}
1 & y_{1, t-1} & y_{2, t-1} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & y_{1, t-1} & y_{2, t-1}
\end{array}\right]\left[\begin{array}{c}
c_{1} \\
\phi_{11} \\
\phi_{12} \\
c_{2} \\
\phi_{21} \\
\phi_{22}
\end{array}\right] \\
& =\left[\begin{array}{c}
c_{1}+\phi_{11} y_{1, t-1}+\phi_{12} y_{2, t-1} \\
c_{2}+\phi_{21} y_{1, t-1}+\phi_{22} y_{2, t-1}
\end{array}\right] \\
& =\left[\begin{array}{c}
c_{1} \\
c_{2}
\end{array}\right]+\left[\begin{array}{ll}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{array}\right]\left[\begin{array}{l}
y_{1, t-1} \\
y_{2, t-1}
\end{array}\right] \\
& =\mathbf{c}+\mathbf{\Phi}_{1} \mathbf{y}_{t-1} .
\end{aligned}
$$

Using this notation, (18-19) can be written more compactly as

$$
\begin{equation*}
\mathbf{y}_{t} \mid \mathbf{y}_{t-1}, \mathbf{y}_{t-1}, . ., \mathbf{y}_{0}, \mathbf{y}_{-1}, . ., \mathbf{y}_{-p+1} \sim N\left(\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}, \boldsymbol{\Omega}\right) \tag{18-21}
\end{equation*}
$$

Thus, the conditional density of the $t$ th observation is

$$
\begin{aligned}
& f_{\mathbf{Y}_{t} \mid \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots, \mathbf{Y}_{-p+1}}\left(\mathbf{y}_{t} \mid \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{-p+1} ; \boldsymbol{\theta}\right) \\
& =(2 \pi)^{-k / 2}\left|\boldsymbol{\Omega}^{-1}\right|^{1 / 2} \exp \left[(-1 / 2)\left(\mathbf{y}_{t}-\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}\left(\mathbf{y}_{t}-\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}\right)\right]
\end{aligned}
$$

The log likelihood function of the fully sample $\mathbf{y}_{T}, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_{1}$ conditioned on $\mathbf{y}_{0}, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}$ is therefore

$$
\begin{equation*}
\mathcal{L}^{*}(\boldsymbol{\theta})=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\boldsymbol{\Omega}^{-1}\right|-(1 / 2) \sum_{t=1}^{T}\left[\left(\mathbf{y}_{t}-\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}\left(\mathbf{y}_{t}-\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}\right)\right] . \tag{18-22}
\end{equation*}
$$

### 2.2.2 MLE of $\boldsymbol{\beta}$

The MLE of $\boldsymbol{\beta}$ is the value $\hat{\boldsymbol{\beta}}$ maximize (18-22). At first glance it is not a trivial work to find $\hat{\boldsymbol{\beta}}$. However, at a close look, $\ddot{\mathbf{X}}_{\mathrm{t}}$ is a special matrix as the matrix $\ddot{\mathbf{X}}_{\mathrm{t}}$ in Section 5.1 of Ch. 10, i.e. $\mathbf{x}_{1 t}=\mathbf{x}_{2 t}=\ldots=\mathbf{x}_{M t}$, or the same regressors. Therefore $\hat{\boldsymbol{\beta}}$ is simply obtained from OLS regression of $y_{i t}$ on $\mathbf{x}_{t}$ from the results of SURE model.

### 2.2.3 MLE of $\Omega$

When evaluated at the MLE $\hat{\boldsymbol{\beta}}$, the log likelihood (18-22) is

$$
\begin{equation*}
\mathcal{L}^{*}(\boldsymbol{\Omega}, \hat{\boldsymbol{\beta}})=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\boldsymbol{\Omega}^{-1}\right|-(1 / 2) \sum_{t=1}^{T} \hat{\varepsilon}_{t}^{\prime} \boldsymbol{\Omega}^{-1} \hat{\varepsilon}_{t}, \tag{18-23}
\end{equation*}
$$

where $\hat{\varepsilon}_{t}=\mathbf{y}_{t}-\ddot{\mathbf{X}}_{\mathbf{t}} \hat{\boldsymbol{\beta}}$.
Differentiate (18-23) with respect to $\boldsymbol{\Omega}^{-1}$ (see p. 23 of Ch 1 ) we obtain

$$
\begin{aligned}
& \frac{\partial \mathcal{L}^{*}(\boldsymbol{\Omega}, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\Omega}^{-1}}=(T / 2) \frac{\partial \ln \left|\boldsymbol{\Omega}^{-1}\right|}{\partial \boldsymbol{\Omega}^{-1}}-(1 / 2) \sum_{t=1}^{T} \frac{\partial \hat{\varepsilon}_{t}^{\prime} \boldsymbol{\Omega}^{-1} \hat{\varepsilon}_{t}}{\partial \boldsymbol{\Omega}^{-1}} \\
&=(T / 2) \boldsymbol{\Omega}^{\prime}-(1 / 2) \sum_{t=1}^{T} \hat{\varepsilon}_{t} \hat{\varepsilon}^{\prime} \\
& t
\end{aligned}
$$

The likelihood is maximized when this derivative is set to zero, or when

$$
\hat{\boldsymbol{\Omega}}^{\prime}=(1 / T) \sum_{t=1}^{T} \hat{\varepsilon}_{t} \hat{\varepsilon}^{\prime}{ }_{t}
$$

Since $\boldsymbol{\Omega}$ is symmetric, we have

$$
\hat{\boldsymbol{\Omega}}=(1 / T) \sum_{t=1}^{T} \hat{\varepsilon}_{t} \hat{\varepsilon}^{\prime}{ }_{t}
$$

also. The row $i$, column $j$ elements of $\hat{\boldsymbol{\Omega}}$ is

$$
\hat{\sigma}_{i j}=(1 / T) \sum_{t=1}^{T} \hat{\varepsilon}_{i t} \hat{\varepsilon}_{j t},
$$

which is the average product of the OLS residual for variable $i$ and the OLS residual for variable $j$.

### 2.2.4 Likelihood Ratio Tests about the Lag Order of VAR

To perform a likelihood ratio test, we need to calculate the maximum value achieved for (18-22). Thus consider

$$
\begin{equation*}
\mathcal{L}^{*}(\hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\beta}})=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\hat{\boldsymbol{\Omega}}^{-1}\right|-(1 / 2) \sum_{t=1}^{T} \hat{\varepsilon}_{t}^{\prime} \hat{\boldsymbol{\Omega}}^{-1} \hat{\varepsilon}_{t} . \tag{18-24}
\end{equation*}
$$

The last term in (18-24) is

$$
\begin{aligned}
(1 / 2) \sum_{t=1}^{T} \hat{\varepsilon}_{t}^{\prime} \hat{\Omega}^{-1} \hat{\varepsilon}_{t} . & =(1 / 2) \operatorname{trace}\left[\sum_{t=1}^{T} \hat{\varepsilon}_{t}^{\prime} \hat{\Omega}^{-1} \hat{\varepsilon}_{t}\right] \\
& =(1 / 2) \operatorname{trace}\left[\sum_{t=1}^{T} \hat{\Omega}^{-1} \hat{\varepsilon}_{t} \hat{\varepsilon}_{t}^{\prime}\right] \\
& =(1 / 2) \operatorname{trace}\left[\hat{\Omega}^{-1}(T \hat{\Omega})\right] \\
& =(1 / 2) \operatorname{trace}\left(T \cdot \mathbf{I}_{k}\right) \\
& =T k / 2 .
\end{aligned}
$$

Substituting this into (18-24) produces

$$
\mathcal{L}^{*}(\hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\beta}})=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\hat{\boldsymbol{\Omega}}^{-1}\right|-(T k / 2)
$$

Suppose we want to test the null hypothesis that was generated from a Gaussian $V A R$ with $p_{0}$ lags against the alternative specification of $p_{1}>p_{0}$ lags. Then we may estimate the model with $M L E$ under $H_{0}$ of $p_{0}$ lags and under $H_{1}$ of $p_{1}$ lags and obtains the maximum value for the log likelihood value

$$
\mathcal{L}_{0}^{*}=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\hat{\Omega}_{0}^{-1}\right|-(T k / 2) .
$$

and

$$
\mathcal{L}_{1}^{*}=(-T k / 2) \ln (2 \pi)+(T / 2) \ln \left|\hat{\Omega}_{1}^{-1}\right|-(T k / 2),
$$

respectively.
Twice the log likelihood ratio is then (see p. 4 of Ch.1)

$$
\begin{aligned}
2\left(\mathcal{L}_{1}^{*}-\mathcal{L}_{0}^{*}\right) & =T\left(\ln \left|\hat{\mathbf{\Omega}}_{1}^{-1}\right|\right)-T\left(\ln \left|\hat{\mathbf{\Omega}}_{0}^{-1}\right|\right) \\
& =T \ln \left(\frac{1}{\left|\hat{\boldsymbol{\Omega}}_{1}\right|}\right)-T \ln \left(\frac{1}{\left|\hat{\mathbf{\Omega}}_{0}\right|}\right) \\
& =-T \ln \left|\hat{\mathbf{\Omega}}_{1}\right|+T \ln \left|\hat{\mathbf{\Omega}}_{0}\right| \\
& =T\left\{\ln \left|\hat{\mathbf{\Omega}}_{0}\right|-\ln \left|\hat{\mathbf{\Omega}}_{1}\right|\right\} .
\end{aligned}
$$

Under the null hypothesis, this asymptotically has a $\chi^{2}$ distribution with degrees of freedom equal to the number of restrictions imposed under $H_{0}$.

## $\mathfrak{E x a m p l e}$.

See the example on page 297 of Hamilton.

### 2.2.5 AIC and BIC in VAR

We postulate that the true process VAR is a $k$-dimensional autoregression of order $p_{0}$. Abstracting from deterministic regressors (such as seasonal dummies or intercepts), the most common used lag-order selection criteria are:

$$
\begin{aligned}
A I C & =\ln |\hat{\boldsymbol{\Omega}}(p)|+\frac{2}{T}\left(k^{2} p\right) \\
B I C & =\ln |\hat{\boldsymbol{\Omega}}(p)|+\frac{\ln T}{T}\left(k^{2} p\right)
\end{aligned}
$$

where $T$ is the effective sample size and $\hat{\boldsymbol{\Omega}}(p)$ is the maximum likelihood estimate of the innovation covariance matrix $\boldsymbol{\Omega}$ (see Sin and White (1996) for further discussion of the theoretical rationale for these criteria). The lag order estimate $\hat{p}$ is chosen to minimize the value of the criterion function for $\{p: 1 \leq p \leq \bar{p}\}$.

### 2.3 Bivariate Granger Causality Tests

### 2.3.1 Definitions of Causality

Granger (1969) has defined a concept of causality which, under suitable conditions, is fairly easy to deal with in the context of $V A R$ models. Therefore it has become quite popular in recent years. The idea is that a cause can not come after the effect. Thus, if a variable $Y$ affect a variable $X$, the former should help improving the predictions of the latter variable.

To formalize this idea, we said that $Y$ fail to Granger cause $X$ if for all $s>0$ the mean squares error of a forecast of $X_{t+s}$ based on ( $\left.X_{t}, X_{t-1}, \ldots\right)$ is the same as the MSE of a forecast of $X_{t+s}$ that use both $\left(X_{t}, X_{t-1}, \ldots\right)$ and ( $\left.Y_{t}, X_{t-1}, \ldots\right)$. If we restrict ourselves to linear functions, $Y$ fails to Granger-cause $X$ if

$$
\operatorname{MSE}\left[\hat{E}\left(X_{t+s} \mid X_{t}, X_{t-1}, \ldots\right)\right]=\operatorname{MSE}\left[\hat{E}\left(X_{t+s} \mid X_{t}, X_{t-1}, \ldots, Y_{t}, Y_{t-1}, \ldots\right)\right]
$$

### 2.3.2 Alternative Implications of Granger Causality, VAR

In a bivariate $V A R$ describing $X$ and $Y, Y$ does not Granger-cause $X$ if the coefficients $\boldsymbol{\Phi}_{j}$ are lower triangular for all $j$ :

$$
\begin{aligned}
{\left[\begin{array}{c}
X_{t} \\
Y_{t}
\end{array}\right]=} & {\left[\begin{array}{l}
c_{1} \\
c_{2}
\end{array}\right]+\left[\begin{array}{cc}
\phi_{11}^{(1)} & 0 \\
\phi_{21}^{(1)} & \phi_{22}^{(1)}
\end{array}\right]\left[\begin{array}{c}
X_{t-1} \\
Y_{t-1}
\end{array}\right]+\left[\begin{array}{cc}
\phi_{11}^{(2)} & 0 \\
\phi_{21}^{(2)} & \phi_{22}^{(2)}
\end{array}\right]\left[\begin{array}{c}
X_{t-2} \\
Y_{t-2}
\end{array}\right]+\ldots } \\
& +\left[\begin{array}{cc}
\phi_{11}^{(p)} & 0 \\
\phi_{21}^{(p)} & \phi_{22}^{(p)}
\end{array}\right]\left[\begin{array}{c}
X_{t-p} \\
Y_{t-p}
\end{array}\right]+\left[\begin{array}{c}
\varepsilon_{X t} \\
\varepsilon_{Y t}
\end{array}\right] .
\end{aligned}
$$

From the first row of this system, the optimal one-period-ahead forecast of $x$ depends only on its own lagged values and not on lagged $y$ :

$$
\operatorname{MSE}\left[\hat{E}\left(X_{t+1} \mid X_{t}, X_{t-1}, \ldots, Y_{t}, Y_{t-1}, \ldots\right)\right]=c_{1}+\phi_{11}^{(1)} X_{t}+\phi_{11}^{(2)} X_{t-1}+\ldots+\phi_{11}^{(p)} X_{t-p+1}
$$

By induction, the same is true of an $s$-period-ahead forecast. Thus for the bivariate $V A R, y$ does not Granger-cause $X$ if $\boldsymbol{\Phi}_{j}$ is lower triangular for all $j$.

### 2.3.3 Alternative Implications of Granger Causality, VMA

Recall from (18-10) that

$$
\boldsymbol{\Psi}_{s}=\boldsymbol{\Phi}_{1} \boldsymbol{\Psi}_{s-1}+\boldsymbol{\Phi}_{2} \boldsymbol{\Psi}_{s-2}+\ldots+\boldsymbol{\Phi}_{p} \boldsymbol{\Psi}_{s-p} \quad \text { for } s=1,2, \ldots
$$

with $\boldsymbol{\Psi}_{0}=\mathbf{I}_{k}$ and $\boldsymbol{\Psi}_{s}=\mathbf{0}$ for $s<0$. This expression implies that if $\boldsymbol{\Phi}_{j}$ is lower triangular for all $j$, then the moving average matrices $\Psi j$ for the fundamental representation will be lower triangular for all $s$. Thus if $y$ fails to Granger-cause $x$, then the $\operatorname{VMA(\infty )}$ representation can be written as

$$
\left[\begin{array}{c}
X_{t} \\
Y_{t}
\end{array}\right]=\left[\begin{array}{l}
\mu_{1} \\
\mu_{2}
\end{array}\right]+\left[\begin{array}{cc}
\psi_{11}(L) & 0 \\
\psi_{21}(L) & \psi_{22}(L)
\end{array}\right]\left[\begin{array}{l}
\varepsilon_{X t} \\
\varepsilon_{Y t}
\end{array}\right]
$$

where

$$
\psi_{i j}(L)=\psi_{i j}^{(0)}+\psi_{i j}^{(1)} L+\psi_{i j}^{(2)} L^{2}+\psi_{i j}^{(3)} L^{3}+\ldots
$$

with $\psi_{11}^{(0)}=\psi_{22}^{(0)}=1$ and $\psi_{21}^{(0)}=0$.

### 2.3.4 Econometric Tests for Granger Causality

A simple approach to test whether a particular series $Y$ "Granger Cause" $X$ can be based on the $V A R$. To implement this test, we assume a particular autoregressive lag length $p$ and estimate

$$
\begin{equation*}
X_{t}=c_{1}+\alpha_{1} X_{t-1}+\alpha_{2} X_{t-2}+\ldots+\alpha_{p} X_{t-p}+\beta_{1} Y_{t-1}+\beta_{2} Y_{t-2}+\ldots .+\beta_{p} Y_{t-p}+u_{t} \tag{18-25}
\end{equation*}
$$

by OLS. We then conduct a $F$ test of the null hypothesis

$$
H_{0}: \beta_{1}=\beta_{2}=\ldots=\beta_{p}=0
$$

Recalling section 4.2.1 of Chapter 6 , one way to implement this test is to calculate the sum of the squared residuals from (18-25),

$$
R S S_{u}=\sum_{t=1}^{T} \hat{u}_{t}^{2}
$$

and compare this with the sum of squared residuals of an univariate autoregression for $x_{t}$,

$$
R S S_{r}=\sum_{t=1}^{T} \hat{e}_{t}^{2}
$$

where

$$
\begin{equation*}
X_{t}=\hat{c}_{0}+\hat{\gamma}_{1} X_{t-1}+\hat{\gamma}_{2} X_{t-2}+\ldots+\hat{\gamma}_{p} X_{t-p}+e_{t} \tag{18-26}
\end{equation*}
$$

is also estimated by OLS. If

$$
S \equiv \frac{\left(R S S_{r}-R S S_{u}\right) / p}{R S S_{u} /(T-2 p-1)}
$$

is greater than $5 \%$ critical value of an $F(p, T-2 p-1)$ distribution, then we reject the null hypothesis that $Y$ does not Granger-cause $X$; that is, if $S$ is sufficiently large, we conclude that $Y$ does Granger-cause $X$.

## Exercise 2.

Please specify a bivariate $V A R$ model for Taiwan's GDP and Stock Index from $L R$ test and from this model to test the Granger-causality between these two variables.

### 2.4 The Impulse-Response Function

In equation (18-9) a $V A R$ can be written in vector $M A(\infty)$ form as

$$
\begin{equation*}
\mathbf{y}_{t}=\boldsymbol{\mu}+\varepsilon_{t}+\mathbf{\Psi}_{1} \varepsilon_{t-1}+\mathbf{\Psi}_{2} \varepsilon_{t-2}+\mathbf{\Psi}_{3} \varepsilon_{t-3}+\ldots \tag{18-27}
\end{equation*}
$$

Thus, the matrix $\boldsymbol{\Psi}_{s}$ has the interpretation

$$
\frac{\partial \mathbf{y}_{t+s}}{\partial \varepsilon_{t}^{\prime}}=\Psi_{s}
$$

that is, the row $i$, column $j$ element of $\Psi_{s}$ identifies the consequence of a one-unit increase in the $j$ th variables's innovation at date $t\left(\varepsilon_{j, t}\right)$ for the values of the $i$ th variable at time $t+s\left(Y_{i, t+s}\right)$, holding all other innovations at all date constant.

A plot of the row $i$, column $j$ element of $\boldsymbol{\Psi}_{s}$,

$$
\begin{equation*}
\frac{\partial Y_{i, t+s}}{\partial \varepsilon_{j, t}} \tag{18-28}
\end{equation*}
$$

as a function of $s$ is called the impulse response function. It describe the response of $y_{i, t+s}$ to one-time impulse in $Y_{j, t}$ with other variables dated $t$ or earlier held constant.

Suppose that we are told that date $t$ value of the first observation in the autoregression, $Y_{1, t}$, was higher than expected, so that $\varepsilon_{1 t}$ is positive. How does this cause to revise our forecast of $Y_{i, t+s}$ ? In other word, what is the response of

$$
\frac{\partial Y_{i, t+s}}{\partial \varepsilon_{1 t}}, \quad i=1,2, . ., k
$$

when we consider that the elements of $\varepsilon_{t}$ are contemporaneously correlated with one another, the fact that $\varepsilon_{1 t}$ is positive gives us some useful new information about the value of $\varepsilon_{2 t}, \ldots, \varepsilon_{k t}$. This implication has further implications for the value of $Y_{i, t+s}$. Thus, we would think the impulse response function so defined in (18-28) is a special case when $E\left(\varepsilon_{t} \varepsilon_{t}^{\prime}\right)=\Omega$ is a diagonal matrix.

Of course in general $\boldsymbol{\Omega}$ is not diagonal. However we may proceed as in section 2.1.2 of this chapter to find matrices $\mathbf{A}$ and $\mathbf{D}$ such that

$$
\begin{equation*}
\Omega=\mathbf{A D A}^{\prime} \tag{18-29}
\end{equation*}
$$

where $\mathbf{A}$ is a lower triangular matrix with 1 s along the principal diagonal and $\mathbf{D}$ is a diagonal matrix with positive entries along the principal diagonal.

Using this matrix A we can construct an $(k \times 1)$ vector $\mathbf{u}_{t}$ from

$$
\begin{equation*}
\mathbf{u}_{t}=\mathbf{A}^{-1} \varepsilon_{t} \tag{18-30}
\end{equation*}
$$

[^1]then we see that the elements of $\mathbf{u}_{t}$ are uncorrelated with each other:
\[

$$
\begin{aligned}
E\left(\mathbf{u}_{t} \mathbf{u}_{t}^{\prime}\right) & =\left[\mathbf{A}^{-1}\right] E\left(\varepsilon_{t} \varepsilon_{t}^{\prime}\right)\left[\mathbf{A}^{-1}\right]^{\prime} \\
& =\left[\mathbf{A}^{-1}\right] \boldsymbol{\Omega}\left[\mathbf{A}^{\prime}\right]^{-1} \\
& =\left[\mathbf{A}^{-1}\right] \mathbf{A D A} \mathbf{A}^{\prime}\left[\mathbf{A}^{\prime}\right]^{-1} \\
& =\mathbf{D}
\end{aligned}
$$
\]

From (18-11) we have

$$
\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_{t}^{\prime}}=\mathbf{\Psi}_{s} \mathbf{A}
$$

or

$$
\left[\begin{array}{cccccc}
\frac{\partial y_{1, t+s}}{\partial u_{1}} & \frac{\partial y_{1, t+s} \partial u_{2}}{\partial u_{2}} & \cdot & \cdot & \frac{\partial y_{1, t+s}}{\partial u_{k}} \\
\frac{\partial y_{2, t+s}}{\partial u_{1}} & \frac{\partial y_{2, t+s}}{\partial u_{2}} & \cdot & \cdot & \cdot & \frac{\partial y_{2, t+s}}{\partial u_{k}} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\frac{\partial y_{k, t+s}}{\partial u_{1}} & \frac{\partial y_{k, t+s}}{\partial u_{2}} & \cdot & \cdot & \cdot & \frac{\partial y_{k, t+s}}{\partial u_{k}}
\end{array}\right]=\left[\begin{array}{lllll}
\boldsymbol{\Psi}_{s} \mathbf{a}_{1} & \Psi_{s} \mathbf{a}_{2} & \cdot & \cdot & . \Psi_{s} \mathbf{a}_{k}
\end{array}\right]
$$

where $\mathbf{a}_{j}$ are the $j$ th column of $\mathbf{A}$.
A plot (how many figures ?) of

$$
\begin{equation*}
\frac{\partial \mathbf{y}_{t+s}}{\partial u_{j t}}=\boldsymbol{\Psi}_{s} \mathbf{a}_{j} \tag{18-31}
\end{equation*}
$$

as a functions of $s$ is known as an orthogonalized impulse - response function.
For a given observed sample of size $T$, we would estimate the autoregressive coefficients $\hat{\boldsymbol{\Phi}}_{1}, \hat{\boldsymbol{\Phi}}_{2}, . ., \hat{\boldsymbol{\Phi}}_{p}$ by $C S S$ (or conditional MLE; that is, OLS from each single equation) and construct $\hat{\mathbf{\Psi}}_{s}$ from (18-10). OLS estimation would also provided the estimate $\hat{\boldsymbol{\Omega}}=(1 / T) \sum_{t=1}^{T} \hat{\varepsilon}_{t} \hat{\varepsilon}_{t}^{\prime}$. Matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{D}}$ satisfying $\hat{\boldsymbol{\Omega}}=\hat{\mathbf{A}} \hat{\mathbf{D}} \hat{\mathbf{A}}^{\prime}$ could then be constructed from $\hat{\boldsymbol{\Omega}}$. The sample estimate of (18-31) is then

$$
\hat{\mathbf{\Psi}}_{s} \hat{\mathbf{a}}_{j}
$$

Another popular form is also implemented and reported. Recall that $\mathbf{D}$ is a diagonal matrix whose $(j, j)$ element is the variance of $u_{j t}$. Let $\mathbf{D}^{1 / 2}$ denote the diagonal matrix whose $(j, j)$ element is the standard deviation of $u_{j t}$. Note that (18-29) could be written as

$$
\begin{equation*}
\boldsymbol{\Omega}=\mathbf{A D}^{1 / 2} \mathbf{D}^{1 / 2} \mathbf{A}^{\prime}=\mathbf{P P}^{\prime} \tag{18-32}
\end{equation*}
$$


where

$$
\mathbf{P}=\mathbf{A} \mathbf{D}^{1 / 2}
$$

Expression (18-32) is the Cholesky decomposition of the matrix $\boldsymbol{\Omega}$. Note that, like A, the $(k \times k)$ matrix $\mathbf{P}$ is lower triangular and has standard deviation of $\mathbf{u}_{t}$ along its principal diagonal.

In place of $\mathbf{u}_{t}$ defined in (18-30), some researcher use

$$
\mathbf{v}_{t}=\mathbf{P}^{-1} \varepsilon_{t}=\mathbf{D}^{-1 / 2} \mathbf{A}^{-1} \varepsilon_{t}=\mathbf{D}^{-1 / 2} \mathbf{u}_{t}
$$

Thus, $v_{j t}$ is just $u_{j t}$ divided by its standard deviation $\sqrt{d_{j j}}$. A one-unit increase in $v_{j t}$ is the same as one-standard-deviation increase in $u_{j t}$.

In place of the dynamic multiplier $\partial y_{i, t+s} / \partial u_{j t}$, these researchers then report $\partial y_{i, t+s} / \partial v_{j t}$. Denote the $j$ th column of $\mathbf{P}$ by $\mathbf{p}_{j}$, we have

$$
\begin{equation*}
\frac{\partial \mathbf{y}_{t+s}}{\partial v_{j t}}=\boldsymbol{\Psi}_{s} \mathbf{p}_{j} \tag{18-33}
\end{equation*}
$$

from the results of (18-31). We also note that

$$
\mathbf{p}_{j}=\mathbf{A d} d_{j}^{1 / 2}=\mathbf{a}_{j} \sqrt{d_{j j}},
$$

where $\mathbf{d}_{j}^{1 / 2}$ is the $j$ th column of $\mathbf{D}^{1 / 2}$.

### 2.5 Forecast Error Variance Decomposition

The forecast error of a $V A R s$ periods into the future would be

$$
\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s \mid t}=\boldsymbol{\varepsilon}_{t+s}+\mathbf{\Psi}_{1} \varepsilon_{t+s-1}+\mathbf{\Psi}_{2} \varepsilon_{t+s-2}+\ldots+\mathbf{\Psi}_{s-1} \varepsilon_{t+1} .
$$

The mean-squared error of this $s$-period-ahead forecast is thus

$$
\begin{align*}
\operatorname{MSE}\left(\hat{\mathbf{y}}_{t+s \mid t}\right) & =E\left[\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s \mid t}\right)\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s \mid t}\right)^{\prime}\right] \\
& =\boldsymbol{\Omega}+\mathbf{\Psi}_{1} \Omega \mathbf{\Psi}_{1}^{\prime}+\mathbf{\Psi}_{2} \boldsymbol{\Omega} \mathbf{\Psi}_{2}^{\prime}+\ldots+\mathbf{\Psi}_{s-1} \boldsymbol{\Omega} \mathbf{\Psi}_{s-1}^{\prime} . \tag{18-34}
\end{align*}
$$

Let us now consider how each of the orthogonalized disturbance ( $u_{1 t}, \ldots, u_{k t}$ ) contributes to this MSE. Write (18-30) as

$$
\varepsilon_{t}=\mathbf{A} \mathbf{u}_{t}=\mathbf{a}_{1} u_{1 t}+\mathbf{a}_{2} u_{2 t}+\ldots+\mathbf{a}_{k} u_{k t} .
$$

Then

$$
\begin{aligned}
\boldsymbol{\Omega} & =E\left(\varepsilon_{t} \varepsilon_{t}^{\prime}\right) \\
& =\mathbf{a}_{1} \mathbf{a}_{1}^{\prime} \operatorname{Var}\left(u_{1 t}\right)+\mathbf{a}_{2} \mathbf{a}_{2}^{\prime} \operatorname{Var}\left(u_{2 t}\right)+\ldots+\mathbf{a}_{k} \mathbf{a}_{k}^{\prime} \operatorname{Var}\left(u_{k t}\right) .
\end{aligned}
$$

Substituting this result into (18-34), the MSE of the $s$-period-ahead forecast can be written as the sums of $k$ terms, one arising from each of the disturbance $u_{j t}$;

$$
\begin{equation*}
M S E\left(\hat{\mathbf{y}}_{t+s \mid t}\right)=\sum_{j=1}^{k}\left\{\operatorname{Var}\left(u_{j t}\right) \cdot\left[\mathbf{a}_{j} \mathbf{a}_{j}^{\prime}+\boldsymbol{\Psi}_{1} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{1}^{\prime}+\boldsymbol{\Psi}_{2} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{2}^{\prime}+\ldots+\boldsymbol{\Psi}_{s-1} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{s-1}^{\prime}\right]\right\} . \tag{18-35}
\end{equation*}
$$

With this expression, we can calculate the contribution of the $j$ th orthogonalized innovation to the MSE of the s-period-ahead forecast:

$$
\begin{equation*}
\operatorname{Var}\left(u_{j t}\right) \cdot\left[\mathbf{a}_{j} \mathbf{a}_{j}^{\prime}+\boldsymbol{\Psi}_{1} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{1}^{\prime}+\boldsymbol{\Psi}_{2} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{2}^{\prime}+\ldots+\boldsymbol{\Psi}_{s-1} \mathbf{a}_{j} \mathbf{a}_{j}^{\prime} \boldsymbol{\Psi}_{s-1}^{\prime}\right] \tag{18-36}
\end{equation*}
$$

The ration of (18-36) to $\operatorname{MSE}(18-35)$ is called the forecast error variance decomposition.

Alternatively, recalling that $\mathbf{p}_{j}=\mathbf{A} \mathbf{d}_{j}^{1 / 2}=\mathbf{a}_{j} \sqrt{\operatorname{Var}\left(u_{j t}\right)}$, we may express the $M S E$ as

$$
\operatorname{MSE}\left(\hat{\mathbf{y}}_{t+s \mid t}\right)=\sum_{j=1}^{k}\left[\mathbf{p}_{j} \mathbf{p}_{j}^{\prime}+\boldsymbol{\Psi}_{1} \mathbf{p}_{j} \mathbf{p}_{j}^{\prime} \boldsymbol{\Psi}_{1}^{\prime}+\boldsymbol{\Psi}_{2} \mathbf{p}_{j} \mathbf{p}_{j}^{\prime} \boldsymbol{\Psi}_{2}^{\prime}+\ldots+\boldsymbol{\Psi}_{s-1} \mathbf{p}_{j} \mathbf{p}_{j}^{\prime} \boldsymbol{\Psi}_{s-1}^{\prime}\right]
$$

also.

## Exercise 3.

Please plot the impulse response function and forecast error variance decomposition from a bivariate VAR(4) model with Taiwan's GDP and Stock Index data set (first difference of the data may be necessary).

## 3 Appendix: Collection of Results Between Least Square and MLE Estimators

In this section, we collect the results of the relation between the least square and the MLE estimators under linear regression model. In this setting, we allow the sample to be time series data-sub $t$, or cross section data-sub $i$, or both.

### 3.1 OLS and MLE under Ideal Conditions, sub $t$ or sub $i$

Let the linear regression model

$$
Y_{t}=\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}+\varepsilon_{t}, \quad t=1,2, \ldots, T
$$

Collect the sample in matrix form, this relationship is written as

$$
\mathbf{y}=\left[\begin{array}{c}
Y_{1}  \tag{18-37}\\
Y_{2} \\
\cdot \\
\cdot \\
\cdot \\
Y_{T}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{x}_{1}^{\prime} \\
\mathbf{x}_{2}^{\prime} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{x}_{T}^{\prime}
\end{array}\right] \boldsymbol{\beta}+\left[\begin{array}{c}
\varepsilon_{1} \\
\varepsilon_{2} \\
\cdot \\
\cdot \\
\cdot \\
\varepsilon_{T}
\end{array}\right]=\mathbf{X} \boldsymbol{\beta}+\boldsymbol{\varepsilon}
$$

where $\mathbf{y}$ is $T \times 1$ vector, $\mathbf{X}$ is an $T \times k$ matrix with rows $\mathbf{x}_{t}^{\prime}, \boldsymbol{\beta}=\left[\beta_{1}, \beta_{2}, \cdots, \beta_{k}\right]^{\prime}$ and $\boldsymbol{\varepsilon}$ is an $T \times 1$ vector with element $\varepsilon_{t}$. We assume that $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^{2} \mathbf{I}_{T}\right)$.

### 3.1.1 OLS Estimation of $\boldsymbol{\beta}$

Let us first consider the Ordinary Least Square estimator $(O L S)$ which is the value for $\boldsymbol{\beta}$ that minimizes the sum of squared errors denoted as $S S E$ (or residuals, remember the principal of estimation at Ch. 3)

$$
\begin{align*}
S S E(\boldsymbol{\beta}) & =(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta}) \\
& =\sum_{t=1}^{T}\left(Y_{t}-\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}\right)^{2} \tag{18-38}
\end{align*}
$$

### 3.1.2 MLE Estimation of $\boldsymbol{\beta}$

From (18-37) it is apparent that the sample joint density is

$$
\begin{aligned}
\mathbf{y} & \sim N\left(\mathbf{X} \boldsymbol{\beta}, \sigma^{2} \mathbf{I}\right), \quad i, e \\
f\left(\mathbf{y}, \mathbf{X} ; \boldsymbol{\beta}, \sigma^{2}\right) & =(2 \pi)^{-T / 2}\left|\sigma^{2} \mathbf{I}\right|^{-1 / 2} \exp \left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}\left(\sigma^{2} \mathbf{I}\right)^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\} \\
& =(2 \pi)^{-T / 2}\left(\sigma^{2}\right)^{-T / 2} \exp \left\{-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\}
\end{aligned}
$$

the log-likelihood function therefore is

$$
\begin{align*}
\ln L\left(\boldsymbol{\beta}, \sigma^{2} ; \mathbf{y}\right) & =-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})  \tag{18-39}\\
& =-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2 \sigma^{2}} \sum_{t=1}^{T}\left(Y_{t}-\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}\right)^{2}  \tag{18-40}\\
& =-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2 \sigma^{2}} \sum_{t=1}^{T}\left(\varepsilon_{t}^{2}\right)
\end{align*}
$$

From (18-38),(18-39) and (18-40), the results that OLS estimators of $\boldsymbol{\beta}$ is identical to that of MLE is clear.

## Example

The conditional log likelihood function of an $A R(1)$ process. See eq.(5.2.27) on p. 122 of Hamilton (1994)..

### 3.2 GLS and MLE under Known Nonspherical Disturbance, sub $t$ or sub $i$

Let the linear regression model

$$
Y_{t}=\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}+\varepsilon_{t}, \quad t=1,2, \ldots, T
$$

Collect the sample in matrix form, this relationship is written as

$$
\mathbf{y}=\left[\begin{array}{c}
Y_{1}  \tag{18-41}\\
Y_{2} \\
\cdot \\
\cdot \\
\cdot \\
Y_{T}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{x}_{1}^{\prime} \\
\mathbf{x}_{2}^{\prime} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{x}_{T}^{\prime}
\end{array}\right] \boldsymbol{\beta}+\left[\begin{array}{c}
\varepsilon_{1} \\
\varepsilon_{2} \\
\cdot \\
\cdot \\
\cdot \\
\varepsilon_{T}
\end{array}\right]=\mathbf{X} \boldsymbol{\beta}+\boldsymbol{\varepsilon}
$$

where $\mathbf{y}$ is $T \times 1$ vector, $\mathbf{X}$ is an $T \times k$ matrix with rows $\mathbf{x}_{t}^{\prime}, \boldsymbol{\beta}=\left[\beta_{1}, \beta_{2}, \cdots, \beta_{k}\right]^{\prime}$ and $\boldsymbol{\varepsilon}$ is an $T \times 1$ vector with element $\varepsilon_{t}$. We assume that $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^{2} \boldsymbol{\Omega}\right)$ with a known $\boldsymbol{\Omega}$.

### 3.2.1 GLS Estimation of $\boldsymbol{\beta}$

Since by assumption $\boldsymbol{\Omega}$ is known, it can find a matrix $\mathbf{P}$ such that

$$
\mathbf{P}^{\prime} \mathbf{P}=\Omega^{-1}
$$

Let us consider the Generalized Least Square estimator ( $G L S$ ) which is the value for $\boldsymbol{\beta}$ that minimizes the Sum of Squared Errors denoted as SSE of the transformed equation

$$
\begin{align*}
S S E(\boldsymbol{\beta}) & =(\mathbf{P} \mathbf{y}-\mathbf{P} \mathbf{X} \boldsymbol{\beta})^{\prime}(\mathbf{P} \mathbf{y}-\mathbf{P X} \boldsymbol{\beta}) \\
& =[\mathbf{P}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})]^{\prime}[(\mathbf{P}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})] \\
& =(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \mathbf{P}^{\prime} \mathbf{P}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})  \tag{18-42}\\
& =(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \mathbf{\Omega}^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})  \tag{18-43}\\
& \neq \sum_{t=1}^{T}\left(Y_{t}-\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}\right)^{2},
\end{align*}
$$

where

$$
\mathrm{Py}=\mathbf{P X} \boldsymbol{\beta}+\mathbf{P} \varepsilon
$$

which satisfy

$$
E(\mathbf{P} \boldsymbol{\varepsilon})(\mathbf{P} \boldsymbol{\varepsilon})^{\prime}=\sigma^{2} \mathbf{I}_{T}
$$

### 3.2.2 MLE Estimation of $\boldsymbol{\beta}$

From (18-41) it is apparent that the sample joint density is

$$
\begin{aligned}
\mathbf{y} & \sim N\left(\mathbf{X} \boldsymbol{\beta}, \sigma^{2} \boldsymbol{\Omega}\right), \quad i, e \\
f\left(\mathbf{y}, \mathbf{X} ; \boldsymbol{\beta}, \sigma^{2}\right) & =(2 \pi)^{-T / 2}\left|\sigma^{2} \boldsymbol{\Omega}\right|^{-1 / 2} \exp \left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}\left(\sigma^{2} \boldsymbol{\Omega}\right)^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\} \\
& =(2 \pi)^{-T / 2}\left(\sigma^{2}\right)^{-T / 2}|\boldsymbol{\Omega}|^{-1 / 2} \exp \left\{-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\}
\end{aligned}
$$

the log-likelihood function therefore is

$$
\begin{align*}
\ln L\left(\boldsymbol{\beta}, \sigma^{2} ; \mathbf{X}, \mathbf{y}\right) & =-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2} \ln |\boldsymbol{\Omega}|-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta}) \\
& \neq-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2} \ln |\boldsymbol{\Omega}|-\frac{1}{2 \sigma^{2}} \sum_{t=1}^{T}\left(Y_{t}-\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}\right)^{2} \tag{18-44}
\end{align*}
$$

From (18-43) and (18-44), the results that GLS estimators of $\boldsymbol{\beta}$ is identical to that of MLE is clear.

### 3.3 Feasible GLS and MLE under Unknown Nonspherical Disturbance, sub $t$ or sub $i$

Let the linear regression model

$$
Y_{t}=\mathbf{x}_{t}^{\prime} \boldsymbol{\beta}+\varepsilon_{t}, \quad t=1,2, \ldots, T
$$

Collect the sample in matrix form, this relationship is written as

$$
\mathbf{y}=\left[\begin{array}{c}
Y_{1}  \tag{18-45}\\
Y_{2} \\
\cdot \\
\cdot \\
\cdot \\
Y_{T}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{x}_{1}^{\prime} \\
\mathbf{x}_{2}^{\prime} \\
\cdot \\
\cdot \\
\cdot \\
\mathbf{x}_{T}^{\prime}
\end{array}\right] \boldsymbol{\beta}+\left[\begin{array}{c}
\varepsilon_{1} \\
\varepsilon_{2} \\
\cdot \\
\cdot \\
\cdot \\
\varepsilon_{T}
\end{array}\right]=\mathbf{X} \boldsymbol{\beta}+\boldsymbol{\varepsilon}
$$

where $\mathbf{y}$ is $T \times 1$ vector, $\mathbf{X}$ is an $T \times k$ matrix with rows $\mathbf{x}_{t}^{\prime}, \boldsymbol{\beta}=\left[\beta_{1}, \beta_{2}, \cdots, \beta_{k}\right]^{\prime}$ and $\boldsymbol{\varepsilon}$ is an $T \times 1$ vector with element $\varepsilon_{t}$. We assume that $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^{2} \boldsymbol{\Omega}\right)$. Here $\boldsymbol{\Omega}=\boldsymbol{\Omega}(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ is a vector of few unknown parameters.

### 3.3.1 FGLS Estimation of $\boldsymbol{\beta}$

Since by assumption $\boldsymbol{\Omega}$ is unknown, it cannot find a matrix $\mathbf{P}$ such that

$$
\mathbf{P}^{\prime} \mathbf{P}=\Omega^{-1}
$$

The Feasible Generalized Least Square estimator (FGLS) is

$$
\begin{equation*}
\check{\boldsymbol{\beta}}=\left(\mathbf{X}^{\prime} \hat{\mathbf{\Omega}}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime} \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}, \tag{18-46}
\end{equation*}
$$

where $\hat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$, i.e. $\hat{\boldsymbol{\Omega}}=\boldsymbol{\Omega}(\hat{\boldsymbol{\theta}})$ in which $\hat{\boldsymbol{\theta}} \xrightarrow{p} \hat{\boldsymbol{\theta}}$.

### 3.3.2 MLE Estimation of $\boldsymbol{\beta}$

From (52) it is apparent that the sample joint density is

$$
\begin{aligned}
\mathbf{y} & \sim N\left(\mathbf{X} \boldsymbol{\beta}, \sigma^{2} \boldsymbol{\Omega}\right), \quad i, e \\
f\left(\mathbf{y}, \mathbf{X} ; \boldsymbol{\beta}, \sigma^{2} \boldsymbol{\Omega}\right) & =(2 \pi)^{-T / 2}\left|\sigma^{2} \boldsymbol{\Omega}\right|^{-1 / 2} \exp \left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime}\left(\sigma^{2} \boldsymbol{\Omega}\right)^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\} \\
& =(2 \pi)^{-T / 2}\left(\sigma^{2}\right)^{-T / 2}|\boldsymbol{\Omega}|^{-1 / 2} \exp \left\{-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})\right\}
\end{aligned}
$$

the log-likelihood function therefore is

$$
\begin{equation*}
\ln L\left(\boldsymbol{\beta}, \sigma^{2} \boldsymbol{\Omega} ; \mathbf{y}, \mathbf{X}\right)=-\frac{T}{2} \ln \left(2 \pi \sigma^{2}\right)-\frac{1}{2} \ln |\boldsymbol{\Omega}|-\frac{1}{2 \sigma^{2}}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta})^{\prime} \boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X} \boldsymbol{\beta}) \tag{18-47}
\end{equation*}
$$

Since the MLE of $\boldsymbol{\beta}$ is the value to satisfy the FOC in (18-47), which is possible highly nonlinear. From (18-46) and (18-47), the results that FGLS estimators of $\boldsymbol{\beta}$ is not identical to that of MLE is clear.

## Example.

The exact log-likelihood function of an $A R$ process. See eq.(5.2.9) on p. 119 of Hamilton (1994).

### 3.4 System of Equations: GLS and MLE Under Known Nonspherical Disturbance, sub $t$ and sub $i$, SURE Model

An alternative way of developing the SURE estimator-which does not involve Kronecker products - is to write the $M$ equations together as

$$
\ddot{\mathbf{y}}_{t}=\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}+\ddot{\varepsilon}_{t}, \quad t=1,2, \ldots, T
$$

where

$$
\ddot{\mathbf{y}}_{t}=\left[\begin{array}{c}
y_{1 t} \\
y_{2 t} \\
. \\
. \\
. \\
y_{M t}
\end{array}\right], \quad \ddot{\mathbf{X}}_{t}=\left[\begin{array}{cccccc}
\mathbf{x}_{1 t}^{\prime} & \mathbf{0} & . & . & . & \mathbf{0} \\
\mathbf{0} & \mathbf{x}_{2 t}^{\prime} & \mathbf{0} & . & . & \mathbf{0} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}_{M t}^{\prime}
\end{array}\right], \boldsymbol{\beta}=\left[\begin{array}{c}
\boldsymbol{\beta}_{1} \\
\boldsymbol{\beta}_{2} \\
. \\
. \\
. \\
\boldsymbol{\beta}_{M}
\end{array}\right]
$$

and

If the $T$ equation are stacked in the usual way, we have

$$
\begin{equation*}
\ddot{\mathrm{y}}=\ddot{\mathrm{X}} \boldsymbol{\beta}+\ddot{\boldsymbol{\varepsilon}}, \tag{18-48}
\end{equation*}
$$

where

$$
\ddot{\mathbf{y}}=\left[\begin{array}{c}
\ddot{\mathbf{y}}_{1} \\
\ddot{\mathbf{y}}_{2} \\
\cdot \\
\cdot \\
. \\
\ddot{\mathbf{y}}_{T}
\end{array}\right], \ddot{\mathbf{x}}=\left[\begin{array}{c}
\ddot{\mathbf{x}}_{1} \\
\cdot \\
\cdot \\
\ddot{\mathbf{x}}_{T}
\end{array}\right] \text {, and } \ddot{\varepsilon}=\left[\begin{array}{c}
\ddot{\varepsilon}_{1} \\
\ddot{\varepsilon}_{2} \\
\cdot \\
\cdot \\
\cdot \\
\ddot{\varepsilon}_{T}
\end{array}\right] \text {. }
$$

The covariance matrix of the disturbance in the stacked equation is

$$
\begin{aligned}
& E\left(\ddot{\varepsilon} \ddot{\varepsilon}^{\prime}\right)=E\left[\begin{array}{c}
\ddot{\varepsilon}_{1} \\
\ddot{\varepsilon}_{2} \\
\cdot \\
\cdot \\
\ddot{\varepsilon}_{T}
\end{array}\right]\left[\begin{array}{lllll}
\ddot{\varepsilon}_{1}^{\prime} & \ddot{\varepsilon}_{2}^{\prime} & . & . & \ddot{\varepsilon}_{T}^{\prime}
\end{array}\right]
\end{aligned}
$$

and we assume that $\boldsymbol{\varepsilon}$ is normally distributed and $\boldsymbol{\Omega}$ (and therefore $\boldsymbol{\Lambda}$ ) is known.

### 3.4.1 GLS Estimation of $\boldsymbol{\beta}$

Since by assumption $\boldsymbol{\Omega}$ is known, it can find a matrix $\mathbf{P}$ such that

$$
\mathbf{P}^{\prime} \mathbf{P}=\Lambda^{-1}=\mathbf{I}_{T} \otimes \mathbf{\Omega}^{-1}
$$

Let us consider the Generalized Least Square estimator ( $G L S$ ) which is the value for $\boldsymbol{\beta}$ that minimizes the sum of squared errors denoted as SSE of the transformed equation

$$
\begin{align*}
S S E(\boldsymbol{\beta}) & =(\mathbf{P} \ddot{\mathbf{y}}-\mathbf{P} \ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime}(\mathbf{P} \ddot{\mathbf{y}}-\mathbf{P} \ddot{\mathbf{X}} \boldsymbol{\beta}) \\
& =[\mathbf{P}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})]^{\prime}[\mathbf{P}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})] \\
& =(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime} \mathbf{P}^{\prime} \mathbf{P}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta}) \\
& =(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime} \boldsymbol{\Lambda}^{-1}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta}) \\
& =\sum_{t=1}^{T}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right), \tag{18-49}
\end{align*}
$$

where

$$
P \ddot{y}=P \ddot{X} \boldsymbol{\beta}+P \ddot{\varepsilon}
$$

satisfy

$$
E(\mathbf{P} \ddot{\varepsilon})(\mathbf{P} \ddot{\varepsilon})^{\prime}=\mathbf{I}_{M T} .
$$

### 3.4.2 MLE Estimation of $\boldsymbol{\beta}$

From (18-48) it is apparent that the sample joint density is

$$
\begin{aligned}
\ddot{\mathbf{y}} & \sim N(\ddot{\mathbf{X}} \boldsymbol{\beta}, \boldsymbol{\Lambda}), \quad i, e, \\
f(\ddot{\mathbf{y}}, \ddot{\mathbf{X}} ; \boldsymbol{\beta}) & =(2 \pi)^{-M T / 2}|\boldsymbol{\Lambda}|^{-1 / 2} \exp \left\{-\frac{1}{2}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime}(\boldsymbol{\Lambda})^{-1}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})\right\} \\
& =(2 \pi)^{-M T / 2}|\boldsymbol{\Lambda}|^{-1 / 2} \exp \left\{-\frac{1}{2}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime} \boldsymbol{\Lambda}^{-1}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})\right\}
\end{aligned}
$$

the log-likelihood function therefore is

$$
\begin{align*}
\ln L(\boldsymbol{\beta} ; \ddot{\mathbf{X}}, \ddot{\mathbf{y}}) & =-\frac{M T}{2} \ln (2 \pi)-\frac{1}{2} \ln |\boldsymbol{\Lambda}|-\frac{1}{2}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta})^{\prime} \boldsymbol{\Lambda}^{-1}(\ddot{\mathbf{y}}-\ddot{\mathbf{X}} \boldsymbol{\beta}) \\
& =-\frac{M T}{2} \ln (2 \pi)-\frac{T}{2} \ln |\boldsymbol{\Omega}|-\frac{1}{2} \sum_{t=1}^{T}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right) \\
& =-\frac{M T}{2} \ln (2 \pi)+\frac{T}{2} \ln \left|\boldsymbol{\Omega}^{-1}\right|-\frac{1}{2} \sum_{t=1}^{T}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right) \tag{18-50}
\end{align*}
$$

From (18-49) and (18-50), the results that GLS estimators of $\boldsymbol{\beta}$ is identical to that of MLE is clear.

### 3.5 System of Equations: FGLS and MLE under Unknown Nonspherical Disturbance, sub $t$ and sub $i$, SURE Model, VAR

An alternative way of developing the SURE estimator-which does not involve Kronecker products - is to write the $M$ equations together as

$$
\ddot{\mathbf{y}}_{\mathbf{t}}=\ddot{\mathbf{X}}_{\mathbf{t}} \boldsymbol{\beta}+\ddot{\varepsilon}_{\boldsymbol{t}}, \quad t=1,2, \ldots, T
$$

where

$$
\ddot{\mathbf{y}}_{\mathbf{t}}=\left[\begin{array}{c}
y_{1 t} \\
y_{2 t} \\
\cdot \\
. \\
. \\
y_{M t}
\end{array}\right], \quad \ddot{\mathbf{X}}_{\mathbf{t}}=\left[\begin{array}{cccccc}
\mathbf{x}_{\mathbf{1 t}}^{\prime} & \mathbf{0} & . & . & . & \mathbf{0} \\
\mathbf{0} & \mathbf{x}_{\mathbf{2 t}}^{\prime} & \mathbf{0} & . & . & \mathbf{0} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
\mathbf{0} & . & . & . & \mathbf{0} & \mathbf{x}_{\mathbf{M t}}^{\prime}
\end{array}\right], \boldsymbol{\beta}=\left[\begin{array}{c}
\boldsymbol{\beta}_{\mathbf{1}} \\
\boldsymbol{\beta}_{\mathbf{2}} \\
. \\
. \\
. \\
\boldsymbol{\beta}_{M}
\end{array}\right]
$$

and

If the $T$ equation are stacked in the usual way, we have

$$
\begin{equation*}
\ddot{\mathbf{y}}=\ddot{\mathbf{X}} \boldsymbol{\beta}+\ddot{\boldsymbol{\varepsilon}}, \tag{18-51}
\end{equation*}
$$

where

$$
\ddot{\mathbf{y}}=\left[\begin{array}{c}
\ddot{\mathbf{y}}_{1} \\
\ddot{\mathbf{y}}_{2} \\
\cdot \\
\cdot \\
. \\
\ddot{\mathbf{y}}_{\mathbf{T}}
\end{array}\right], \ddot{\mathrm{X}}=\left[\begin{array}{c}
\ddot{\mathbf{X}}_{1} \\
\cdot \\
\cdot \\
\ddot{\mathbf{X}}_{\mathrm{T}}
\end{array}\right] \text {, and } \ddot{\varepsilon}=\left[\begin{array}{c}
\ddot{\varepsilon}_{1} \\
\ddot{\varepsilon}_{2} \\
\cdot \\
\cdot \\
. \\
\ddot{\varepsilon}_{T}
\end{array}\right] \text {. }
$$

The covariance matrix of the disturbance in the stacked equation is

$$
\begin{aligned}
& E\left(\ddot{\varepsilon} \ddot{\varepsilon}^{\prime}\right)=E\left[\begin{array}{c}
\ddot{\varepsilon}_{1} \\
\ddot{\varepsilon}_{2} \\
\cdot \\
\cdot \\
\cdot \\
\ddot{\varepsilon}_{T}
\end{array}\right]\left[\begin{array}{lllll}
\ddot{\varepsilon}_{1}^{\prime} & \ddot{\varepsilon}_{2}^{\prime} & \cdot & & \ddot{\varepsilon}_{T}^{\prime}
\end{array}\right]
\end{aligned}
$$

and we assume that $\boldsymbol{\varepsilon}$ is normally distributed and $\boldsymbol{\Omega}=\boldsymbol{\Omega}(\boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ is a vector of few unknown parameters.

### 3.5.1 FGLS Estimation of $\boldsymbol{\beta}$

Since by assumption $\boldsymbol{\Omega}$ is unknown, it cannot find a matrix $\mathbf{P}$ such that

$$
\mathbf{P}^{\prime} \mathbf{P}=\Lambda^{-1}
$$

The Feasible Generalized Least Square estimator (FGLS) is

$$
\begin{equation*}
\check{\boldsymbol{\beta}}=\left(\ddot{\mathbf{X}}^{\prime} \hat{\boldsymbol{\Lambda}}^{-1} \ddot{\mathbf{X}}\right)^{-1} \ddot{\mathbf{X}}^{\prime} \hat{\boldsymbol{\Lambda}}^{-1} \ddot{\mathbf{y}} \tag{18-52}
\end{equation*}
$$

where $\hat{\boldsymbol{\Lambda}}$ is a consistent estimator of $\boldsymbol{\Lambda}$, i.e., $\hat{\boldsymbol{\Lambda}}=\boldsymbol{\Lambda}(\hat{\boldsymbol{\theta}})$ in which $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$.

### 3.5.2 MLE Estimation of $\boldsymbol{\beta}$

From (18-51), the log-likelihood function therefore is

$$
\begin{equation*}
\ln L(\boldsymbol{\beta}, \boldsymbol{\Omega}(\boldsymbol{\theta}) ; \ddot{\mathbf{y}})=-\frac{M T}{2} \ln (2 \pi)+\frac{T}{2} \ln \left|\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\right|-\frac{1}{2} \sum_{t=1}^{T}\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right)^{\prime} \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\left(\ddot{\mathbf{y}}_{t}-\ddot{\mathbf{X}}_{t} \boldsymbol{\beta}\right) . \tag{18-53}
\end{equation*}
$$

Since the MLE of $\boldsymbol{\beta}$ (and $\boldsymbol{\theta}$ ) is the value to satisfy the FOC in (18-53), which is possible highly nonlinear. From (18-52) and (18-53), the results that GLS estimators of $\boldsymbol{\beta}$ is not identical to that of MLE is clear.

## Example.

The first look of condition log likelihood function of an $\operatorname{VAR}(p)$ process. See eq.(11.1.10) on p. 293 of Hamilton (1994).


The Long Gallery of Social Science Building, NSYSU.

## End of this Chapter


[^0]:    ${ }^{1}$ Here, $L$ is $1 \times 1$.

[^1]:    2017 by $\mathfrak{P r o f}$. $\mathfrak{C h i n g n u n} \mathfrak{L e v}$

