5: MULTIVARATE STATIONARY PROCESSES
1 Some Preliminary Definitions and Concepts

- **Random Vector**: A vector $X = (X_1, ..., X_n)$ whose components are scalar-valued random variables on the same probability space.

- **Vector Random Process**: A family of random vectors $\{X_t, t \in T\}$ defined on a probability space, where $T$ is a set of time points. Typically $T = \mathbb{R}$, $T = \mathbb{Z}$ or $T = \mathbb{N}$, the sets or real, integer and natural numbers, respectively.

- **Time Series Vector**: A particular realization of a vector random process.
1.1 The Lag operator

- The lag operator $L$ maps a sequence $\{X_t\}$ into a sequence $\{Y_t\}$ such that $Y_t = LX_t = X_{t-1}$, for all $t$.

- If we apply $L$ repeatedly on a process, for instance $L(L(LX_t))$, we will use the convention $L(L(LX_t)) = L^3X_t = X_{t-3}$.

- If we apply $L$ to a constant $c$, $Lc = c$.

- Inversion: $L^{-1}$ is the inverse of $L$, such that $L^{-1}(L)X_t = X_t$, $L^{-1} = X_{t+1}$.

- The difference operator $\Delta$ is the filter $1 - L$. When applied to $X_t$, $(1 - L)X_t = \Delta X_t = X_t - X_{t-1}$
1.2 Polynomials in the lag operator

- We can form (univariate) polynomials: \( \phi(L) = \phi_0 L^0 + \phi_1 L + \phi_2 L^2 + ... + \phi_p L^p \) is a polynomial in the lag operator of order \( p \) and is such that \( \phi(L)X_t = \phi_0 X_t + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} \). Recall the AR(1) process: \( (1 - \phi L)Y_t = \varepsilon_t \), with \( \varepsilon_t \sim WN \).

- Lag polynomials can also be inverted. For a polynomial \( \phi(L) \), we are looking for the values of the coefficients \( \alpha_i \) of \( \phi(L)^{-1} = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + ... \) such that \( \phi(L)^{-1}\phi(L) = 1 \).

Case 1: \( p = 1 \). Let \( \phi(L) = (1 - \phi L) \) with \( |\phi| < 1 \). To find the inverse write

\[
(1 - \phi L)(\alpha_0 + \alpha_1 L + \alpha_2 L^2 + ...)^{-1} = 1
\]

note that all the coefficients of the non-zero powers of \( L \) must be equal to zero. This gives

\[
\begin{align*}
\alpha_0 &= 1 \\
-\phi + \alpha_1 &= 0 \quad \Rightarrow \quad \alpha_1 = \phi
\end{align*}
\]
\[-\phi \alpha_1 + \alpha_2 = 0 \quad \Rightarrow \quad \alpha_2 = \phi^2\]
\[-\phi \alpha_2 + \alpha_3 = 0 \quad \Rightarrow \quad \alpha_3 = \phi^3\]

and so on. In general \(\alpha_k = \phi^k\), so \((1 - \phi L)^{-1} = \sum_{j=0}^\infty \phi^j L^j\) provided that \(|\phi| < 1\).

It is easy to check this because

\[
(1 - \phi L)(1 + \phi L + \phi^2 L^2 + \ldots + \phi^k L^k) = 1 - \phi^{k+1} L^{k+1}
\]

so

\[
(1 + \phi L + \phi^2 L^2 + \ldots + \phi^k L^k) = \frac{1 - \phi^{k+1} L^{k+1}}{(1 - \phi L)}
\]

and \(k \to \infty \sum_{j=0}^k \phi^j L^j \to \frac{1}{(1-\phi L)}\).
Case 2: $p = 2$. Let $\phi(L) = (1 - \phi_1 L - \phi_2 L^2)$. To find the inverse it is useful to factor the polynomial in the following way

$$(1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

where the $\lambda_1, \lambda_2$ are the reciprocal of the roots of the above left-hand side polynomial or equivalently the eigenvalues of the matrix

$$\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}$$

Suppose $\lambda_1, \lambda_2 < 0$ and $\lambda_1 \neq \lambda_2$. We have that $(1 - \phi_1 L - \phi_2 L^2)^{-1} = (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}$. Therefore we can use what we have seen above for the case $p = 1$. We can write

$$(1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1} = (\lambda_1 - \lambda_2)^{-1} \left[ \frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L} \right]$$

$$= \frac{\lambda_1}{\lambda_1 - \lambda_2} \left[ 1 + \lambda_1 L + \lambda_1 L^2 + ... \right] - \frac{\lambda_2}{\lambda_1 - \lambda_2} \left[ 1 + \lambda_2 L + \lambda_2 L^2 + ... \right]$$

$$= (c_1 + c_2) + (c_1 \lambda_1 + c_2 \lambda_2) L + (c_1 \lambda_1^2 + c_2 \lambda_2^2) L^2 + ...$$
where \( c_1 = \lambda_1/(\lambda_1 - \lambda_2), \ c_2 = -\lambda_2/(\lambda_1 - \lambda_2) \)

- **Matrix of polynomial in the lag operator**: \( \Phi(L) \) if its elements are polynomial in the lag operator, i.e.

\[
\Phi(L) = \begin{pmatrix}
1 & -0.5L \\
L & 1 + L
\end{pmatrix} = \Phi_0 + \Phi_1 L
\]

where

\[
\Phi_0 = \begin{pmatrix}
1 & 1 \\
0 & 2
\end{pmatrix}, \quad \Phi_1 = \begin{pmatrix}
0 & -0.5 \\
1 & 1
\end{pmatrix}, \quad \Phi_{j>1} = \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix}.
\]

When applied to a vector \( X_t \) we obtain

\[
\Phi(L)X_t = \begin{pmatrix}
1 & -0.5L \\
L & 1 + L
\end{pmatrix} \begin{pmatrix}
X_{1t} \\
X_{2t}
\end{pmatrix} = \begin{pmatrix}
X_{1t} - 0.5X_{2t-1} \\
X_{1t-1} + X_{2t} + X_{2t-1}
\end{pmatrix}
\]
The inverse is matrix such that $\Phi(L)^{-1}\Phi(L) = I$. Suppose $\Phi(L) = (I - \Phi_1L)$. Its inverse $\Phi(L)^{-1} = A(L)$ is a matrix such that $(A_0 + A_1L + A_2L^2 + ...)\Phi = I$. That is

$$A_0 = I$$
$$A_1 - \Phi_1 = 0 \implies A_1 = \Phi_1$$
$$A_2 - A_1\Phi_1 = 0 \implies A_2 = \Phi^2$$
$$\vdots$$
$$A_k - A_{k-1}\Phi_1 = 0 \implies A_k = \Phi^k$$

(1)
1.3 Covariance Stationarity

Let $Y_t$ be a $n$-dimensional vector of time series, $Y'_t = [Y_{1t}, \ldots, Y_{nt}]$. Then $Y_t$ is covariance (weakly) stationary if $E(Y_t) = \mu$, and the autocovariance matrix $\Gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu)'$ for all $t,j$, that is are independent of $t$ and both finite.

- Stationarity of each of the components of $Y_t$ does not imply stationarity of the vector $Y_t$. Stationarity in the vector case requires that the components of the vector are stationary and costationary.

- Although $\gamma_j = \gamma_{-j}$ for a scalar process, the same is not true for a vector process. The correct relation is

$$\Gamma_j = \Gamma'_{-j}$$

**Example:** $n = 2$ and $\mu = 0$

$$\begin{align*}
\Gamma_1 &= \begin{pmatrix}
E(Y_{1t}Y_{1t-1}) & E(Y_{1t}Y_{2t-1}) \\
E(Y_{2t}Y_{1t-1}) & E(Y_{2t}Y_{2t-1}) \\
E(Y_{1t+1}Y_{1t}) & E(Y_{1t+1}Y_{2t}) \\
E(Y_{2t+1}Y_{1t}) & E(Y_{2t+1}Y_{2t})
\end{pmatrix} \\
&= \begin{pmatrix}
E(Y_{1t}Y_{1t+1}) & E(Y_{1t}Y_{2t+1}) \\
E(Y_{2t}Y_{1t+1}) & E(Y_{2t}Y_{2t+1})
\end{pmatrix}' = \Gamma'_{-1}
\end{align*}$$

1.4 Convergence of random variables

We refresh three concepts of stochastic convergence in the univariate case and then we extend the three cases to the multivariate framework.

Let \( \{x_T, T = 1, 2, \ldots\} \) be a sequence of random variables.

- **Convergence in probability** \( \{x_T\}_{T=1}^{\infty} \) converges in probability to \( c \), (written \( x_T \xrightarrow{p} c \)), if

  \[
  \lim_{T \to \infty} p(|x_T - c| > \epsilon) = 0
  \]

When the above condition is satisfied, the number \( c \) is called *probability limit* or *plim* of the sequence \( \{x_T\} \), indicated as

\[
\text{plim} x_T = c
\]

- **Convergence in mean square**. The sequence of random variables \( \{x_T\} \) converges in mean square to \( c \), denoted by \( x_T \xrightarrow{m.s.} c \), if

  \[
  \lim_{T \to \infty} E(x_T - c)^2 = 0
  \]
• *Convergence in distribution* Let $\{x_T\}_{T=1}^\infty$ be a sequence of random variables and let $F_T$ denote the cumulative distribution function of $x_T$ and $F$ the cumulative distribution of the scalar $x$. The sequence is said to converge in distribution written as $x_T \xrightarrow{d} x$ (or $x_T \xrightarrow{L} x$), if for all real numbers $c$ for which $F$ is continuous

$$\lim_{T \to \infty} F_T(c) = F(c).$$
The concepts of stochastic convergence can be generalized to the multivariate setting. Suppose \( \{X_T\} \) is a sequence of \( n \)-dimensional random vectors and \( X \) is a \( n \)-dimensional random vector. Then

1. \( X_T \overset{p}{\to} X \) if \( X_{jT} \overset{p}{\to} X_j \) for \( j = 1, \ldots, n \)
2. \( X_T \overset{m.s.}{\to} X \) if \( \lim_{T \to \infty} E(X_T - X)'(X_t - X) = 0 \)
3. \( X_T \overset{d}{\to} X \) if \( \lim_{T \to \infty} F_T(c) = F(c) \) where \( F \) and \( F_T \) are the joint distribution of \( X \) and \( X_T \).

**Proposition C.1 L.** Suppose \( \{X_T\}_{T=1}^{\infty} \) is a sequence of \( n \)-dimensional random vectors. Then the following relations hold:

(a) \( X_T \overset{m.s.}{\to} X \Rightarrow X_T \overset{p}{\to} X \Rightarrow X_T \overset{d}{\to} X \)

(c) (Slutsky’s Theorem) If \( g : \mathbb{R}^n \to \mathbb{R}^m \) is a continuous function, then

\[
X_T \overset{p}{\to} X \Rightarrow g(X_T) \overset{p}{\to} g(X) \quad \text{(i.e. } \text{plim } g(X_T) = g(\text{plim } X_T)\text{)}
\]

\[
X_T \overset{d}{\to} X \Rightarrow g(X_T) \overset{d}{\to} g(X)
\]
Proposition C.2 L. Suppose \( \{X_T\} \) and \( \{Y_T\} \) are sequences of \( n \times 1 \) random vectors and \( A_T \) is a sequence of \( n \times n \) random matrices, \( x \) is a \( n \times 1 \) random vector, \( c \) is a fixed \( n \times 1 \) vector, and \( A \) is a fixed \( n \times n \) matrix.

1. If \( \text{plim} X_T, \text{plim} Y_T \) and \( \text{plim} A_T \) exist then
   
   (a) \( \text{plim} (X_T \pm Y_T) = \text{plim} X_T \pm \text{plim} Y_T, \)
   
   (b) \( \text{plim} c'X_T = c'\text{plim} X_T \)
   
   (c) \( \text{plim} X'_T Y_T = (\text{plim} X_T)'(\text{plim} Y_T) \)
   
   (d) \( \text{plim} A_T X_T = (\text{plim} A_T)(\text{plim} X_T) \)

2. If \( X_T \xrightarrow{d} X \) and \( \text{plim} (X_T - Y_T) = 0 \) then \( Y_T \xrightarrow{d} X. \)

3. If \( X_T \xrightarrow{d} X \) and \( \text{plim} Y_T = c \), then
   
   (a) \( X_T + Y_T \xrightarrow{d} X + c \)
   
   (b) \( Y'_T X_t \xrightarrow{d} c'X \)

4. If \( X_T \xrightarrow{d} X \) and \( \text{plim} A_T = A \) then \( A_T X_T \xrightarrow{d} AX \)

5. If \( X_T \xrightarrow{d} X \) and \( \text{plim} A_T = 0 \) then \( \text{plim} A_T X_T = 0 \)
Example Let \( \{X_T\} \) be a sequence of \( n \times 1 \) random vectors with \( X_T \xrightarrow{d} N(\mu, \Omega) \), and Let \( \{Y_T\} \) be a sequence of \( n \times 1 \) random vectors with \( Y_T \xrightarrow{p} C \). Then \( Y'_T X_T \xrightarrow{d} N(C'\mu, C\Omega C') \).
1.5 Infinite sums of random variables

- An infinite sequence of real numbers \( \{a_i\}, i = 0, \pm 1, \pm 2, \ldots \) is absolutely summable if \( \lim_{n \to \infty} \sum_{i=-n}^{n} |a_i| < 0 \)

- A sequence of \( k \times k \) matrices \( \{A_i\} \) is absolutely summable if each sequence (element) is absolutely summable, \( \lim_{n \to \infty} \sum_{i=-n}^{n} |a_{mn,i}| < 0 \), where \( a_{mn,i} \) is the element \((m,n)\) of \( A_i \). Equivalently it is absolutely summable if its Euclidean norm, \( ||A_i|| = (\sum_{m} \sum_{n} a_{mn,i})^{1/2} \), is absolutely summable.

Proposition C.9 L Suppose \( \{A_i\} \) is an absolutely summable sequence of \( k \times k \) matrices and \( \{z_t\} \) is a sequence of \( k \) dimensional random vectors satisfying \( E(z_t'z_t) \leq c \), \( t = 0, \pm 1, \pm 2, \ldots \), for some finite constant \( c \). Then there exists a sequence of \( k \)-dimensional random variables \( \{y_t\} \) such that \( \sum_{i=-n}^{n} A_i z_t \overset{m.s.}{\to} y_t \)
1.6 Limit Theorems

The Law of Large Numbers and the Central Limit Theorem are the most important results for computing the limits of sequences of random variables.

There are many versions of LLN and CLT that differ on the assumptions about the dependence of the variables.

**Proposition C.12 L** (Weak law of large numbers)

1. (iid sequences) Let \( \{y_t\} \) be an i.i.d sequence of random variables with \( E(y_t) = \mu < \infty \). Then

\[
\bar{y}_T = T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{P} \mu
\]

2. (independent sequences) Let \( \{y_t\} \) be a sequence of independent random variables with \( E(y_t) = \mu < \infty \) and \( E|y_t|^{1+\epsilon} \leq c < \infty \) for some \( \epsilon > 0 \) and a finite constant \( c \). Then \( T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{P} \mu \).

3. (stationary processes) Let \( y_t \) be a convariance stationary process with finite \( E(y_t) = \mu \) and \( E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \) with absolutely summable auto-
covariances \( \sum_{j=0}^{\infty} |\gamma_j| < \infty \). Then \( \bar{y}_T \xrightarrow{m.s.} \mu \) hence \( \bar{y}_T \xrightarrow{p} \mu \).

- Weak stationarity and absolutely summable covariances are sufficient conditions for a law of large numbers to hold.

**Proposition C.13L** (Central limit theorem)

1. (i.i.d. sequences) Let \( \{Y_T\} \) be a sequence of \( k \)-dimensional iid random vectors with mean \( \mu \) and covariance matrix \( \Sigma \). Then

   \[
   \sqrt{T} (\bar{Y}_T - \mu) \xrightarrow{d} N(0, \Sigma)
   \]

2. (stationary processes) Let \( Y_t = \mu + \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j} \) be a \( n \)-dimensional stationary random process with, \( E(Y_t) = \mu < \infty \ \sum_{j=0}^{\infty} \|\Phi_j\| < \infty \) and \( \varepsilon_t \sim i.i.d(0, \Sigma) \). Then

   \[
   \sqrt{T} (\bar{Y}_T - \mu) \xrightarrow{d} N \left( 0, \sum_{j=-\infty}^{\infty} \Gamma_j \right)
   \]

   where \( \Gamma_j \) is the autocovariance matrix at lag \( j \).
2 Some stationary processes

2.1 White Noise (WN)

A $n$-dimensional vector white noise $\epsilon'_t = [\epsilon_{1t}, ..., \epsilon_{nt}] \sim WN(0, \Omega)$ is such if $E(\epsilon_t) = 0$ and $\Gamma_k = \Omega$ ($\Omega$ a symmetric positive definite matrix) if $k = 0$ and 0 if $k \neq 0$. If $\epsilon_t$, $\epsilon_\tau$ are independent the process is an independent vector White Noise (i.i.d). If also $\epsilon_t \sim N$ the process is a Gaussian WN.

Important: A vector whose components are white noise is not necessarily a white noise. Example: let $u_t$ be a scalar white noise and define $\epsilon_t = (u_t, u_{t-1})'$. Then $E(\epsilon_t \epsilon'_t) = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}$ and $E(\epsilon_t \epsilon'_{t-1}) = \begin{pmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{pmatrix}$. 

2.2 Vector Moving Average (VMA)

Given the $n$-dimensional vector White Noise $\epsilon_t$ a vector moving average of order $q$ is defined as

$$Y_t = \mu + \epsilon_t + C_1 \epsilon_{t-1} + \ldots + C_q \epsilon_{t-q}$$

where $C_j$ are $n \times n$ matrices of coefficients.

- **The VMA(1)**

Let us consider the VMA(1)

$$Y_t = \mu + \epsilon_t + C_1 \epsilon_{t-1}$$

with $\epsilon_t \sim WN(0, \Omega)$, $\mu$ is the mean of $Y_t$. The variance of the process is given by

$$\Gamma_0 = E[(Y_t - \mu)(Y_t - \mu)'] = \Omega + C_1 \Omega C_1'$$

with autocovariances

$$\Gamma_1 = C_1 \Omega, \quad \Gamma_{-1} = \Omega C_1', \quad \Gamma_j = 0 \text{ for } |j| > 1$$
• The VMA($q$)

Let us consider the VMA($q$)

$$Y_t = \mu + \epsilon_t + C_1\epsilon_{t-1} + \ldots + C_q\epsilon_{t-q}$$

with $\epsilon_t \sim WN(0, \Omega)$, $\mu$ is the mean of $Y_t$. The variance of the process is given by

$$\Gamma_0 = E[(Y_t - \mu)(Y_t - \mu)'] = \Omega + C_1\Omega C_1' + C_2\Omega C_2' + \ldots + C_q\Omega C_q'$$

with autocovariances

$$\Gamma_j = C_j\Omega + C_{j+1}\Omega C_1' + C_{j+2}\Omega C_2' + \ldots + C_q\Omega C_{q-j}' \quad \text{for} \quad j = 1, 2, \ldots, q$$

$$\Gamma_j = \Omega C'_j + C_1\Omega C'_{j+1} + C_2\Omega C'_{j+2} + \ldots + C_{q+j}\Omega C_q' \quad \text{for} \quad j = -1, -2, \ldots, -q$$

$$\Gamma_j = 0 \quad \text{for} \quad |j| > q$$
• The VMA(∞)

A useful process, as we will see, is the VMA(∞)

\[ Y_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} \]  

the process can be thought as the limiting case of a VMA(q) (for \( q \to \infty \)). Recall the previous result the process converges in mean square if \( \{C_j\} \) is absolutely summable.

**Proposition (10.2H)**. Let \( Y_t \) be an \( n \times 1 \) vector satisfying

\[ Y_t = \mu + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} \]

where \( \varepsilon_t \) is a vector WN with \( E(\varepsilon_{t-j}) = 0 \) and \( E(\varepsilon_t \varepsilon'_{t-j}) = \Omega \) for \( j = 0 \) and zero otherwise and \( \{C_j\}_{j=0}^{\infty} \) is absolutely summable. Let \( Y_{it} \) denote the \( i \)th element of \( Y_t \) and \( \mu_i \) the \( i \)th element of \( \mu \). Then

(a) The autocovariance between the \( i \)th variable at time \( t \) and the \( j \)th variable at time \( s \) periods earlier, \( E(Y_{it} - \mu_i)(Y_{jt-s} - \mu_j) \) exists and is given by
the row i column j element of

\[ \Gamma_s = \sum_{v=0}^{\infty} C_{s+v} \Omega C'_s \]

for \( s = 0, 1, 2, \ldots \).

(b) The sequence of matrices \( \{\Gamma_s\}_{s=0}^{\infty} \) is absolutely summable.

If furthermore \( \{\varepsilon_t\}_{t=-\infty}^{\infty} \) is an i.i.d. sequence with \( E|\varepsilon_{i_1 t} \varepsilon_{i_2 t} \varepsilon_{i_3 t} \varepsilon_{i_4 t}| \leq \infty \) for \( i_1, i_2, i_3, i_4 = 1, 2, \ldots, n \), then also

(c) \( E|Y_{i_1 t_1} Y_{i_2 t_2} Y_{i_3 t_3} Y_{i_4 t_4}| \leq \infty \) for all \( t_1, t_2, t_3, t_4 \)

(d) \( (1/T) \sum_{t=1}^{T} y_{i t} y_{j t-s} \overset{p}{\to} E(y_{i t} y_{j t-s}) \), for \( i, j = 1, 2, \ldots, n \) and for all \( s \)

Implications:

1. Result (a) implies that the second moments of a \( MA(\infty) \) with absolutely summable coefficients can be found by taking the limit of the autocovariance of an \( MA(q) \).

2. Result (b) ensures ergodicity for the mean

3. Result (c) says that \( Y_t \) has bounded fourth moments

22
4. Result (d) says that $Y_t$ is ergodic for second moments

Note: the vector $MA(\infty)$ representation of a stationary VAR satisfies the absolute summability condition so that assumption of the previous proposition hold.
2.3 Invertible and fundamental VMA

- The VMA is invertible, i.e. it possesses a VAR representation, if and only if the determinant of $C(L)$ vanishes only outside the unit circle, i.e. if $\det(C(z)) \neq 0$ for all $|z| \leq 1$.

**Example** Consider the process

$$
\begin{pmatrix}
Y_{1t} \\
Y_{2t}
\end{pmatrix} =
\begin{pmatrix}
1 & L \\
0 & \theta - L
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{pmatrix}
$$

$\det(C(z)) = \theta - z$ which is zero for $z = \theta$. The process is invertible if and only if $|\theta| > 1$.

- The VMA is fundamental if and only if the $\det(C(z)) \neq 0$ for all $|z| < 1$. In the previous example the process is fundamental if and only if $|\theta| \geq 1$. In the case $|\theta| = 1$ the process is fundamental but noninvertible.

- Provided that $|\theta| > 1$ the MA process can be inverted and the shock can be obtained as a combination of present and past values of $Y_t$. That is the VAR
(Vector Autoregressive) representation can be recovered. The representation will entail infinitely many lags of $Y_t$ with absolutely summable coefficients, so that the process converges in mean square.

Considering the above example

$$
\begin{pmatrix}
1 & -\frac{L}{\theta - L} \\
0 & \frac{1}{\theta - L}
\end{pmatrix}
\begin{pmatrix}
Y_{1t} \\
Y_{2t}
\end{pmatrix}
= 
\begin{pmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{pmatrix}
$$

or

$$
Y_{1t} + \frac{L}{\theta} \frac{1}{1 - \frac{1}{\theta} L} Y_{2t} = \varepsilon_{1t}
$$

$$
\frac{1}{\theta} \frac{1}{1 - \frac{1}{\theta} L} Y_{2t} = \varepsilon_{2t}
$$

(3)
2.4 Wold Decomposition

Any zero-mean stationary vector process $Y_t$ admits the following representation

$$Y_t = C(L) \varepsilon_t + \mu_t$$

(4)

where $C(L) \varepsilon_t$ is the stochastic component with $C(L) = \sum_{i=0}^{\infty} C_i L^i$ and $\mu_t$ the purely deterministic component, the one perfectly predictable using linear combinations of past $Y_t$.

If $\mu_t = 0$ the process is said regular. Here we only consider regular processes.

(4) represents the Wold representation of $Y_t$ which is unique and for which the following properties hold:

(b) $\varepsilon_t$ is the innovation for $Y_t$, i.e. $\varepsilon_t = Y_t - \text{Proj}(Y_t|Y_{t-1}, Y_{t-1}, ...)$, i.e. the shock is fundamental.

(b) $\varepsilon_t$ is White noise, $E\varepsilon_t = 0$, $E\varepsilon_t \varepsilon'_\tau = 0$, for $t \neq \tau$, $E\varepsilon_t \varepsilon'_t = \Omega$

(c) The coefficients are square summable $\sum_{j=0}^{\infty} \|C_j\|^2 < \infty$.

(d) $C_0 = I$
• The result is very powerful since holds for any covariance stationary process.

• However the theorem does not implies that (4) is the true representation of the process. For instance the process could be stationary but non-linear or non-invertible.
2.5 Other fundamental MA(∞) Representations

- It is easy to extend the Wold representation to the general class of invertible MA(∞) representations. For any non singular matrix $R$ of constant we define $u_t = R^{-1}\epsilon_t$ and we have

$$Y_t = C(L)Ru_t = D(L)u_t$$

where $u_t \sim WN(0, R^{-1}\Omega R^{-1'})$.

- Notice that all these representations obtained as linear combinations of the Wold representations are fundamental. In fact, $det(C(L)R) = det(C(L))det(R)$. Therefore if $det(C(L)R) \neq 0 \forall |z| < 1$ so will $det(C(L)R)$. 

3 VAR: representations

- Every stationary vector process $Y_t$ admits a Wold representation. If the MA matrix of lag polynomials is invertible, then a unique VAR exists.

- We define $C(L)^{-1}$ as an $(n \times n)$ lag polynomial such that $C(L)^{-1}C(L) = I$; i.e. when these lag polynomial matrices are matrix-multiplied, all the lag terms cancel out. This operation in effect converts lags of the errors into lags of the vector of dependent variables.

- Thus we move from MA coefficient to VAR coefficients. Define $A(L) = C(L)^{-1}$. Then given the (invertible) MA coefficients, it is easy to map these into the VAR coefficients:

$$
Y_t = C(L)\epsilon_t
$$

$$
A(L)Y_t = \epsilon_t
$$

(5)

where $A(L) = A_0 - A_1L^1 - A_2L^2 - \ldots$ and $A_j$ for all $j$ are $(n \times n)$ matrices of coefficients.
• To show that this matrix lag polynomial exists and how it maps into the coefficients in $C(L)$, note that by assumption we have the identity

$$(A_0 - A_1 L^1 - A_2 L^2 - ...) (I + C_1 L^1 + C_2 L^2 + ...) = I$$

After distributing, the identity implies that coefficients on the lag operators must be zero, which implies the following recursive solution for the VAR coefficients:

$$
A_0 = I \\
A_1 = A_0 C_1 \\
A_k = A_0 C_k + A_1 C_k + ... + A_{k-1} C_1
$$

• As noted, the VAR is of infinite order (i.e. infinite number of lags required to fully represent joint density).

• In practice, the VAR is usually restricted for estimation by truncating the lag-length. Recall that the AR coefficients are absolutely summable and vanish at long lags.
The \textit{pth-order vector autoregression}, denoted VAR(p) is given by
\begin{equation}
Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t
\end{equation}

\textit{Note}: Here we are considering zero mean processes. In case the mean of \(Y_t\) is not zero we should add a constant in the VAR equations.

* \textit{VAR(1) representation} Any VAR(p) can be rewritten as a VAR(1). To form a VAR(1) from the general model we define: \(e'_t = [\epsilon, 0, ..., 0]\), \(Y'_t = [Y'_t, Y'_{t-1}, ..., Y'_{t-p+1}]\)

\[
A = \begin{pmatrix}
A_1 & A_2 & ... & A_{p-1} & A_p \\
I_n & 0 & ... & 0 & 0 \\
0 & I_n & ... & 0 & 0 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
0 & ... & ... & I_n & 0
\end{pmatrix}
\]

Therefore we can rewrite the VAR(p) as a VAR(1)
\[
Y_t = AY_{t-1} + e_t
\]

This is also known as the companion form of the VAR(p)
• **SUR representation** The VAR($p$) can be stacked as

$$Y = X\Gamma + u$$

where $X = [X_1, \ldots, X_T]'$, $X_t = [Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-p}]'$, $Y = [Y_1, \ldots, Y_T]'$, $u = [\epsilon_1, \ldots, \epsilon_T]'$ and $\Gamma = [A_1, \ldots, A_p]'$

• **Vec representation** Let $vec$ denote the stacking columns operator, i.e $X =

$$
\begin{pmatrix}
X_{11} & X_{12} \\
X_{21} & X_{22} \\
X_{31} & X_{32}
\end{pmatrix}
$$

then $vec(X) =

$$
\begin{pmatrix}
X_{11} \\
X_{21} \\
X_{31} \\
X_{12} \\
X_{22} \\
X_{32}
\end{pmatrix}
$$

Let $\gamma = vec(\Gamma)$, then the VAR can be rewritten as

$$Y_t = (I_n \otimes X'_t)\gamma + \epsilon_t$$
4 VAR: Stationarity

4.1 Stability and stationarity

• Consider the VAR(1)

\[ Y_t = \mu + AY_{t-1} + \varepsilon_t \]

Substituting backward we obtain

\[ Y_t = \mu + AY_{t-1} + \varepsilon_t \]
\[ = \mu + A(\mu + AY_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \]
\[ = (I + A)\mu + A^2Y_{t-2} + A\varepsilon_{t-1} + \varepsilon_t \]
\[ \vdots \]

\[ Y_t = (I + A + \ldots + A^{j-1})\mu + A^jY_{t-j} + \sum_{i=0}^{j-1} A^i\varepsilon_{t-i} \]

• The eigenvalues of \( A, \lambda \), solve \( \det(A - I\lambda) = 0 \). If all the eigenvalues of \( A \) are smaller than one in modulus the sequence \( A^i, i = 0, 1, \ldots \) is absolutely summable. Therefore

1. the infinite sum \( \sum_{i=0}^{j-1} A^i\varepsilon_{t-i} \) exists in mean square;
2. \((I + A + ... + A^{j-1})\mu \rightarrow (I - A)^{-1}\) and \(A^j \rightarrow 0\) as \(j\) goes to infinity.

Therefore if the eigenvalues are smaller than one in modulus then \(Y_t\) has the following representation

\[Y_t = (I - A)^{-1}\mu + \sum_{i=0}^{\infty} A^i \varepsilon_{t-i}\]

- Note that the the eigenvalues correspond to the reciprocal of the roots of the determinant of \(A(z) = I - Az\). A VAR(1) is called \textit{stable} if
  \[\det(I - Az) \neq 0 \text{ for } |z| \leq 1.\]

- For a VAR(p) the stability condition also requires that all the eigenvalues of \(A\) (the AR matrix of the companion form of \(Y_t\)) are smaller than one in modulus. Therefore we have that a VAR(p) is called \textit{stable} if
  \[\det(I - A_1z - A_2z^2, ..., A_pz^p) \neq 0 \text{ for } |z| \leq 1.\]

- \textit{A condition for stationarity:} A stable VAR process is stationary.

- Notice that the converse is not true. An unstable process can be stationary.
4.2 Back the Wold representation

• If the VAR is stationary $Y_t$ has the following Wold representation

$$Y_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} C_j \epsilon_t$$

where the sequence $\{C_j\}$ is absolutely summable, $\sum_{j=0}^{\infty} |C_j| < \infty$

• How can we find it? Let us rewrite the VAR(p) as a VAR(1)

• We know how to find the MA($\infty$) representation of a stationary AR(1). We can proceed similarly for the VAR(1). Substituting backward in the companion form we have

$$Y_t = A^j Y_{t-j} + A^{j-1} \epsilon_{t-j+1} + \ldots + A^1 \epsilon_{t-1} + \ldots + \epsilon_t$$

If conditions for stationarity are satisfied, the series $\sum_{i=1}^{\infty} A^j$ converges and $Y_t$ has an VMA($\infty$) representation in terms of the Wold shock $\epsilon_t$ given by

$$Y_t = (I - AL)^{-1} \epsilon_t$$
\[ \sum_{i=1}^{\infty} A^i e_{t-j} = C(L)e_t \]

where \( C_0 = A_0 = I, C_1 = A_1, C_2 = A^2, \ldots, C_k = A^k \). \( C_j \) will be the \( n \times n \) upper left matrix of \( C_j \).
Example A stationary VAR(1)

\[
\begin{pmatrix}
Y_{1t} \\
Y_{2t}
\end{pmatrix} = \begin{pmatrix}
0.5 & 0.3 \\
0.02 & 0.8
\end{pmatrix} \begin{pmatrix}
Y_{1t-1} \\
Y_{2t-1}
\end{pmatrix} + \begin{pmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{pmatrix}
\]

\[
E(\epsilon_t \epsilon_t') = \Omega = \begin{pmatrix}
1 & 0.3 \\
0.3 & 0.1
\end{pmatrix} \quad \lambda = \begin{pmatrix}
0.81 \\
0.48
\end{pmatrix}
\]

Figure 1: Blu: $Y_1$, green $Y_2$. 
5 VAR: second moments

- Second Moments of a VAR(p).
Let us consider the companion form of a stationary (zero mean for simplicity) VAR(p) defined earlier

\[ Y_t = AY_{t-1} + e_t \]  \hspace{1cm} (7)

The variance of \( Y_t \) is given by

\[ \tilde{\Sigma} = \mathbf{E}[(Y_t)(Y_t)'] = A\tilde{\Sigma}A' + \tilde{\Omega} \]  \hspace{1cm} (8)

A closed form solution to (7) can be obtained in terms of the vec operator. Let \( A, B, C \) be matrices such that the product \( ABC \) exists. A property of the vec operator is that

\[ \text{vec}(ABC) = (C' \otimes A)\text{vec}(B) \]

Applying the vec operator to both sides of (7) we have

\[ \text{vec}(\tilde{\Sigma}) = (A \otimes A)\text{vec}(\tilde{\Sigma}) + \text{vec}(\tilde{\Omega}) \]

If we define \( \mathcal{A} = (A \otimes A) \) then we have

\[ \text{vec}(\tilde{\Sigma}) = (I - \mathcal{A})^{-1}\text{vec}(\tilde{\Omega}) \]
where

\[ \tilde{\Gamma}_0 = \tilde{\Sigma} = \begin{pmatrix}
\Gamma_0 & \Gamma_1 & \ldots & \Gamma_{p-1} \\
\Gamma_{-1} & \Gamma_0 & \ldots & \Gamma_{p-2} \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{-p+1} & \Gamma_{-p+2} & \ldots & \Gamma_0
\end{pmatrix} \]

The variance \( \Sigma = \Gamma_0 \) of the original series \( Y_t \) is given by the first \( n \) rows and columns of \( \tilde{\Sigma} \).

The \( j \)th autocovariance of \( Y_t \) (denoted \( \tilde{\Gamma}_j \)) can be found by post multiplying (6) by \( Y_{t-j} \) and taking expectations:

\[
E(Y_t Y_{t-j}) = A E(Y_t Y_{t-j}) + E(e_t Y_{t-j})
\]

Thus

\[ \tilde{\Gamma}_j = A \tilde{\Gamma}_{j-1} \]

or

\[ \tilde{\Gamma}_j = A^j \tilde{\Gamma}_0 = A^j \tilde{\Sigma} \]
The autocovariances $\Gamma_j$ of the original series $Y_t$ are given by the first $n$ rows and columns of $\tilde{\Gamma}_j$ and are given by

$$\Gamma_h = A_1 \Gamma_{h-1} + A_2 \Gamma_{h-2} + \ldots + A_p \Gamma_{h-p}$$

known as Yule-Walker equations.