Introduction

In the case of univariate time series processes, it was seen that the Wold decomposition provides a rich class of linear models to describe a scalar, stationary time series process. It turns out that extending linear ARMA processes to the linear multivariate ARMA case is relatively straightforward. The population characteristics of Vector Autoregressions (VARs) and Vector Autoregressive Moving Average (VARMA) processes appear to have been first derived by Quenouille (1957), although they were not routinely applied until the early 1980s.

An important issue concerns the relationship between a multivariate time series model and a dynamic econometric model estimated from time series data. In many ways the two are essentially the same. In particular, multivariate time series models may often have an excessive number of parameters, and economic theory is ideally necessary in order to impose restrictions on the parameter space. Also the pure VAR and VARMA multivariate time series models are essentially linear, and economic theory may provide ways of formulating non-linear relationships. However, in many cases, a dynamic econometric model may be approximated by a linear multivariate time series model. Hence, the two fields of multivariate time series analysis and dynamic econometric models are very closely linked and inter-related. The understanding and appreciation of dynamic econometric models inevitably requires a good working knowledge of time series analysis. A major problem area to be addressed later concerns how to best deal with the forms of non stationarity frequently encountered in much economic time series data. Initially, it is necessary to deal with the class of linear, stationary multivariate time series processes that forms the basis for all subsequent work.
**Vector Form of Wold’s Decomposition**

Consider a time series vector of g random variables denoted by

\[ y_t' = (y_{1t}, y_{2t}, \ldots, y_{gt}) \]

so that \( y_t' \) represents a g dimensional vector of random variables measured at discrete intervals of time. It is assumed that \( y_t' \) is covariance stationary, or integrated of order zero, i.e. \( y_t \sim I(0) \), so that each element of the g dimensional vector is stationary. The multivariate form of Wold’s decomposition then implies that apart from fixed period harmonics that \( y_t \) can be represented as an infinite order moving average process given by

\[
y_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j},
\]

where the \( B_j \) are g\times g dimensional coefficient matrices and \( \varepsilon_t \) is a g dimensional vector white noise process with the properties that

\[
E(\varepsilon_t) = 0
\]

and \( E(\varepsilon_t, \varepsilon_s') = \Omega, \ s = t \) and \( E(\varepsilon_t, \varepsilon_{s'}) = 0, \ s \neq t \). Hence \( \Omega \) is a symmetric, positive definite covariance matrix of the innovations.

**Autocovariance Function for Vector Processes**

The dependence between \( y_t \) and a value at a future or previous time period is represented by the autocovariance function, which for a stationary process \( y_t \) will be time invariant. The autocovariance function for a vector process becomes a matrix. In particular, the autocovariance function between \( y_t \) and \( y_{t-k} \) is given by \( \Gamma_k \) and is defined as
so that a typical element of $\Gamma_k$ is $\gamma_{ij}(k)$, which is the scalar autocovariance between $y_i$ and $y_{j-k}$. Note that in general $\gamma_{ij}(k) \neq \gamma_{ji}(-k)$. However, $\gamma_{ij}(k) = \gamma_{ji}(-k)$, so that

$$\Gamma_{-k} = E(y_i y'_{i+k})$$

and

$$\Gamma'_{-k} = E(y_i y'_{i+k})' = E(y_{i+k} y'_{i})' = \Gamma_k$$

The covariance matrix of the general vector process in Wold’s decomposition is
\[ \Gamma_0 = E(y_t y_t') = E\left( \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \right) \left( \sum_{i=0}^{\infty} B_i \varepsilon_{t-i} \right)' = E\left( \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \varepsilon_{t-j}' B_j' \right) \]

(4) \[ \Gamma_0 = \sum_{j=0}^{\infty} B_j \Omega B_j' \]

and the autocovariance matrix at lag \( k \) is

\[ \Gamma_k = E(y_t y_{t-k}') = E\left( \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \right) \left( \sum_{i=0}^{\infty} B_i \varepsilon_{t-i-k} \right)' \]

\[ \Gamma_k = E(\varepsilon_t + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + \ldots + B_k \varepsilon_{t-k} + B_{k+1} \varepsilon_{t-k-1} + \ldots) (\varepsilon_{t-k} + B_k \varepsilon_{t-k-1} + B_{2} \varepsilon_{t-k-2} + \ldots) \]

(5) \[ \Gamma_k = \left( \sum_{j=0}^{\infty} B_{j+k} \Omega B_j' \right), \]

which can be a useful formula for calculating autocovariance functions.

The VAR(1) Process

Before considering the general higher order VAR(p) model, it is very convenient to start with the VAR(1) model, which can also be modified to deal with many higher order multivariate time series models. The VAR(1) process is

(6) \[ y_t = A y_{t-1} + \varepsilon_t \]

By successive substitution,

\[ y_t = \varepsilon_t + A \varepsilon_{t-1} + A^2 \varepsilon_{t-2} + \ldots A^{t-1} \varepsilon_1 + A^t y_0 \]
and from the same arguments as with the scalar case, a necessary condition for stationarity is that \( \lim_{t \to \infty} (A') = 0 \), which is the null matrix. An alternative derivation is to write (6) as,

\[
A(L)y_t = (I - AL)y_t = \varepsilon_t,
\]

and to be invertible for a valid Wold decomposition, it is necessary for all the roots of the determinental polynomial \( |A(L)| \) to lie outside the unit circle. (See the appendix for further details). If the matrix \( A(L) \) is invertible, then the Wold decomposition of the VAR(1) is

\[
y_t = (I - AL)^{-1}\varepsilon_t = \sum_{j=0}^{\infty} A^j\varepsilon_{t-j}.
\]

The autocovariance function of the VAR(1) can also be found by direct analogy with the scalar AR(1) process. On transposing, \( y_t' = y_{t-1}'A' + \varepsilon_t' \) and on premultiplying by \( \varepsilon_t \)

\[
\varepsilon_t y_t' = \varepsilon_t y_{t-1}'A' + \varepsilon_t\varepsilon_t'.
\]

Since

\[
E(\varepsilon_t, y_{t-1}'A') = 0, \ k \geq 1
\]

it follows from taking expectations through equation (9) that \( E(\varepsilon_t, y_t') = \Omega \). On post-multiplying through (6) by \( y_t' \) and taking expectations:

\[
y_t y_t' = Ay_{t-1}y_t' + \varepsilon_t y_t'
\]

\[
\Gamma_0 = A\Gamma_1' + \Omega
\]
Also on post-multiplying through (6) by \( y'_{i-k} \) and taking expectations:

\[
\Gamma_k = A \Gamma_{k-1} + E(\varepsilon_t y'_{i-k})
\]

and hence

(12) \[ \Gamma_k = A \Gamma_{k-1}, \text{ for } k \geq 1 \]

Substituting for \( \Gamma'_i \) from (12) into (11) gives

(13) \[ \Gamma_0 = A \Gamma_0 A' + \Omega \]

On recalling that the column stacking operator, \( s(.) \), possesses the property that

\[
s(ABC) = (C' \otimes A)s(B),
\]

where \( s(\cdot) = vec(\cdot)' \), then from (13)

\[
s(\Gamma_0)(A \otimes A)s(\Gamma_0) + s(\Omega)
\]

(14) \[ s(\Gamma_0) = (I - A \otimes A)^{-1}s(\Omega), \]

which gives a direct expression for the covariance matrix of of \( y \) in terms of \( A \) and the covariance matrix of \( \varepsilon \), which is \( \Omega \). Also, by successive substitution from (12),

\[
\Gamma_k = A^T \Gamma_0,
\]

and hence
(15) \( s(\Gamma_k) = (I \otimes A^k) s(\Gamma_0) \)

An alternative expression for \( \Gamma_0 \) is available from the infinite moving average representation, or Wold decomposition in (1),

(16) \[
\Gamma_0 = E\left( \sum_{j=0}^{\infty} A^j \varepsilon_{t-j} \right) \left( \sum_{j=0}^{\infty} \varepsilon'_{t-j} A^j' \right) = \sum_{j=0}^{\infty} A^j \Omega A^j'
\]

**Example of a Stationary VAR(1) Process**

The following example is based on the bivariate \( g = 2 \) case with one lag, so that \( p = 1 \). Then,

(17) \[
\begin{bmatrix}
y'_{1t} \\
y'_{2t}
\end{bmatrix} = \begin{bmatrix}
(1/3) & -(1/6) \\
-(1/3) & (1/2)
\end{bmatrix} \begin{bmatrix}
y'_{1t-1} \\
y'_{2t-1}
\end{bmatrix} + \begin{bmatrix}
\varepsilon'_{1t} \\
\varepsilon'_{2t}
\end{bmatrix},
\]

which can be expressed in lag operator form as

\[
\begin{bmatrix}
1 - (1/3)L & (1/6)L \\
(1/3)L & 1 - (1/2)L
\end{bmatrix} \begin{bmatrix}
y'_{1t} \\
y'_{2t}
\end{bmatrix} = \begin{bmatrix}
\varepsilon'_{1t} \\
\varepsilon'_{2t}
\end{bmatrix}
\]

and

(18) \[
A(L) = [1 - (1/3)L][1 - (1/2)L] - (1/18)L^2 = 1 - (5/6)L + (1/9)L^2
\]

\[
= [1 - (1/6)L][1 - (2/3)L]
\]

Hence the determinental polynomial has roots of 6 and 3/2; both of which lie outside the unit circle. Hence the process is stationary. The infinite moving average representation, or
Wold decomposition is obtained from (7) as $y_t = \sum_{j=0}^{\infty} A^j \varepsilon_{t-j}$, and in this case the first four terms in the Wold decomposition are

$$
\begin{bmatrix}
y_{1t} \\
y_{2t}
\end{bmatrix} = 
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} + 
\begin{bmatrix}
.333 & -1.67 \\
-3.33 & .500
\end{bmatrix} 
\begin{bmatrix}
\varepsilon_{1t-1} \\
\varepsilon_{2t-1}
\end{bmatrix} + 
\begin{bmatrix}
.167 & -1.39 \\
-.278 & .306
\end{bmatrix} 
\begin{bmatrix}
\varepsilon_{1t-2} \\
\varepsilon_{2t-2}
\end{bmatrix}
+ 
\begin{bmatrix}
.102 & -0.97 \\
-.194 & 1.99
\end{bmatrix} 
\begin{bmatrix}
\varepsilon_{1t-3} \\
\varepsilon_{2t-3}
\end{bmatrix} + 
\begin{bmatrix}
.066 & -0.66 \\
-.131 & .132
\end{bmatrix} 
\begin{bmatrix}
\varepsilon_{1t-4} \\
\varepsilon_{2t-4}
\end{bmatrix} + \ldots.
$$

It can be seen that each element of $A^j$ is tending to zero for increasing powers of A, which is entirely consistent with the $y_t$ process being stationary. The autocovariance matrix of $y_t$ is $\Gamma_0$ and from (13),

$$
\Gamma_0 = A\Gamma_0 A' + \Omega
$$

$$
s(\Gamma_0) = (I - A \otimes A)^{-1} s(\Omega)
$$

Given $\Omega = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$, then,

$$
s(\Gamma_0) = 
\begin{bmatrix}
E(y_{1t}^2) \\
E(y_{1t}y_{2t}) \\
E(y_{1t}y_{2t}) \\
E(y_{2t}^2)
\end{bmatrix} = 
\begin{bmatrix}
1 \\
2 \\
2 \\
5
\end{bmatrix}
$$

$s(\Omega) = 
\begin{bmatrix}
1 \\
2 \\
2 \\
5
\end{bmatrix}$
\[ A \otimes A = \begin{bmatrix} (1/3) & -(1/6) \\ -(1/3) & (1/2) \end{bmatrix} \otimes \begin{bmatrix} (1/3) & -(1/6) \\ -(1/3) & (1/2) \end{bmatrix} = \begin{bmatrix} (1/9) & -(1/18) & -(1/18) & (1/36) \\ -(1/9) & (1/6) & (1/18) & -(1/12) \\ -(1/9) & (1/18) & (1/6) & -(1/12) \\ (1/9) & -(1/6) & -(1/6) & (1/4) \end{bmatrix} \]

Then

\[ s(\Gamma_0) = \begin{bmatrix} E(y_{1t}^2) \\ E(y_{1t}y_{2t}) \\ E(y_{1t}y_{2t}) \\ E(y_{2t}^2) \end{bmatrix} = \begin{bmatrix} 1.093 \\ 1.768 \\ 1.768 \\ 6.043 \end{bmatrix} \]

\[ \Gamma_1 = \begin{bmatrix} 0.070 & -0.418 \\ 0.520 & 2.432 \end{bmatrix}, \quad \Gamma_2 = \begin{bmatrix} -0.634 & -0.545 \\ 0.237 & 1.355 \end{bmatrix}, \text{ etc.} \]

**Stationary Vector Autoregression (VAR)**

The most widely used vector model used in time series econometric work is the Vector Autoregression of order \( p \), or VAR(\( p \)), which is

\[
(21) \quad y_t = \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t,
\]

where \( A_j \) are \( g \times g \) dimensional coefficient matrices and \( \varepsilon_t \) is the vector white noise process which was defined previously. The \( i \)'th equation of the VAR (\( p \)) will be

\[
(22) \quad y_{it} = \sum_{k=1}^{p} \sum_{j=1}^{n} a_{ij}(k) y_{j_{t-k}} + \varepsilon_{it},
\]
where $a_j(k)$ is the $(i,j)$ element of $A_k$. Hence all the last $p$ lagged values of each variable, including itself, explain the current value of $y_{it}$ plus the random innovation $\varepsilon_{it}$. In lag operator form the VAR ($p$) can be expressed as

$$A(L)y_i = \varepsilon_i$$

where

$$A(L) = I - \sum_{j=1}^{p} A_j L^j$$

**Analysis of the VAR(p) From Companion Form**

One great attraction with the VAR(1) model in (6) is that it can be easily used to derive the properties of higher order vector time series models. The increase in the number of equations in the VAR(1) poses little problems for the calculations required; while the increase in the order (i.e. value of $p$) is far more awkward. This fact was also apparent for the univariate AR(p) model compared with the univariate AR(1) process. For this reason, it is particularly convenient to express the VAR(p) in *companion form* as a VAR(1) model by stacking the system to allow for the original model to occupy the first $g$ rows of the VAR(1) and all the other rows to be identities. Then,

$$
\begin{bmatrix}
y_t \\
y_{t-1} \\
\vdots \\
y_{t-p+1}
\end{bmatrix}
= 
\begin{bmatrix}
A_1 & A_2 & A_p \\
I & 0 & \cdots & 0 \\
0 & I & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I
\end{bmatrix}
\begin{bmatrix}
y_{t-1} \\
y_{t-2} \\
\vdots \\
y_{t-p}
\end{bmatrix}
+ 
\begin{bmatrix}
\varepsilon_t \\
0 \\
\vdots \\
0
\end{bmatrix}
$$

which can be expressed as
(26) \[ Y_t = AY_{t-1} + \xi_t \]

where \( Y_t \) and \( \xi_t \) are \( gp \times 1 \) in dimension and \( A \) is \( gp \times gp \).

**Wold Decomposition of the VAR(p)**

On using the results for the pure VAR(1) model, the companion form VAR(p) can be written by successive substitution as

(27) \[ Y_t = \sum_{j=0}^{\infty} A^j \xi_{t-j} \]

On defining a \( g \times gp \) dimensional selection matrix, \( N \) as \( N = [I \ 0] \) the system of equations can then be pre-multiplied by \( N^t \) to obtain

\[ N^tY_t = N^tAY_{t-1} + N^t\xi_t \]

hence

(28) \[ y_t = N^tAY_{t-1} + \varepsilon_t \]

which is an alternative form of expressing the VAR(p) in (21) with \( N'A \) being the first \( g \) rows of \( A \) in (25). On pre-multiplying (27) by the selection matrix,

\[ y_t = \sum_{j=0}^{\infty} N'A^j \xi_{t-j} = \sum_{j=0}^{\infty} (N'A^jN)\varepsilon_{t-j} \]

so that the \( j \)th MAR coefficient matrix is the top left corner \( g \times g \) matrix of \( A^j \). Hence for the VAR (p) model the Wold decomposition matrices, or infinite moving average representation coefficient matrices \( B_j \) are defined as
\[(29) \quad B_j = N'A^jN\]

Since \(Y_j = \sum_{j=0}^{t-1} A^jN^t\), or equivalently for the g equation VAR(p) process,

\[
y_t = \sum_{j=0}^{t-1} (N'A^jN)e_{t-j} + (N'A)^{t-j}Y_0
\]

it again follows that for \(y_t\) to be stationary it must be time invariant, so that a necessary condition is for \(\lim_{t \to \infty} (A)^t = 0\). A formal test for stationarity of the VAR(p) in (21) is that all the roots of \(A(L)\) in (23) must lie outside the unit circle, which is an analogous result for the pure VAR(1). A more formal proof is in the appendix.

**Granger Causality:**

A variable \(x\) Granger causes another variable \(y\), if \(x\) has *predictable content* in the formation of predictions of future \(y\). That is if knowledge of past \(x\) significantly improves prediction of future \(y\), then \(x\) Granger Causes \(y\). More formally,

\[
MSE(E_y | y_t, y_{t-1}, y_{t-2}, \ldots) = \sigma^2_y
\]

\[
MSE(E_{y|x} | y_t, y_{t-1}, y_{t-2}, \ldots, x_t, x_{t-1}, x_{t-2}, \ldots) = \sigma^2_{y,x}
\]

If \(\sigma^2_{y,x} < \sigma^2_y\), then \(x\) Granger Causes \(y\).

Note that the concept of Granger Causality is designed for the situation in engineering terminology of a *black box*, that is where the input and output are observed, but the innermost workings of the black box are unknown.
Test for Linear Granger Causality:

\[ H_0 : \ y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} + \ldots + \phi_p y_{t-p+1} + \epsilon_{t+1} \]

\[ H_1 : \ y_{t+1} = \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} + \ldots + \phi_p y_{t-p+1} + \beta_1 x_t + \beta_2 x_{t-1} + \ldots + \beta_k x_{t-k+1} + \epsilon_{t+1} \]

Notes: (i) A test for \( \beta_1 = \beta_2 = \ldots = \beta_k = 0 \) can be done by either the Wald, Likelihood Ratio or LM tests.

(ii) The above equations can also be interpreted as one equation of a VAR, with and without restrictions. For example, consider the bivariate VAR\( p \) model,

\[ y_{1t} = \sum_{k=1}^{p} a_{11}(k) y_{1t-k} + \sum_{k=1}^{p} a_{12}(k) y_{2t-k} + \epsilon_{1t} \]

\[ y_{2t} = \sum_{k=1}^{p} a_{21}(k) y_{1t-k} + \sum_{k=1}^{p} a_{22}(k) y_{2t-k} + \epsilon_{2t} \]

In the general unrestricted case, there is bi-directional Granger causality between both variables. If \( a_{12}(1) = a_{12}(2) = \ldots = a_{12}(p) = 0 \), then \( y_{1t} \) only depends on its own past history and not on that of \( y_{2t} \). However, both the past of \( y_{1t} \) and \( y_{2t} \) are informative, (or have predictive content) for modeling the behavior of \( y_{2t} \). Hence \( y_{1t} \) is exogenous, but Granger causes \( y_{2t} \).

Conversely, if \( a_{21}(1) = a_{21}(2) = \ldots = a_{21}(p) = 0 \), then \( y_{2t} \) is exogenous, but Granger causes \( y_{1t} \).

(iii) Bi-directional causality is always a distinct possibility. So the test must be run in both directions.

(iv) Contemporaneous causality can occur through the off diagonal element of the error covariance matrix \( \Omega \). Genuine Granger causality must occur with past information reducing the prediction MSE rather than current information.
(v) Sims (1970) provided a different computational method for Granger Causality, which is harder and less intuitive. Sims applied the concept to the money and income debate at that time. However, conditioning on omitted, missing intermediate variables is a potential problem with all of these tests.

**Autoregressive Final Form (ARFF)**

Given a multivariate VARMA or VAR model, one interesting question concerns the univariate time series representation of a component of the \( y_i \) vector process. The VAR(p) model, \( A(L)y_i = \varepsilon_i \), can be inverted to give the Wold decomposition,

\[
y_i = A(L)^{-1} \varepsilon_i
\]

which can be expressed as

\[
(30) \quad y_i = \frac{A(L)^+}{|A(L)|} \varepsilon_i
\]

where \( A(L)^+ \) represents the adjoint matrix of \( A(L) \) and \( |A(L)| \) is the determinental polynomial associated with \( A(L) \) and will generally be a polynomial in the lag operator of order gp. On multiplying both sides of (30) by \( |A(L)| \), the ARFF equations are obtained.

\[
(31) \quad |A(L)|y_i = A(L)^+ \varepsilon_i
\]

From this representation, it can be seen that a typical element of \( y_i \), say \( y_{i_0} \) will follow a univariate ARMA(pg, p) process. This can be seen since the sum of g moving average processes, each of order p, will also be a moving average process of order at most p, which is one of the results found by Granger and Morris (1976) and was referenced in chapter 3.
Another aspect of the ARFF representation is that the autoregressive process will be the same for each univariate process in the \( y_t \) vector. To illustrate this, return to the previous example where

\[
\begin{bmatrix}
y_{1t} \\
y_{2t}
\end{bmatrix} = \begin{bmatrix}
(1/3) & -(1/6) \\
-(1/3) & (1/2)
\end{bmatrix} \begin{bmatrix}
y_{1t-1} \\
y_{2t-1}
\end{bmatrix} + \begin{bmatrix}
e_{1t} \\
e_{2t}
\end{bmatrix}
\]

Then from (31),

\[
\begin{bmatrix}
1 - (5/6)L + (1/9)L^2
\end{bmatrix} \begin{bmatrix}
y_{1t} \\
y_{2t}
\end{bmatrix} = \begin{bmatrix}
1 - (1/2)L & -(1/6)L \\
-(1/2)L & 1 - (1/3)L
\end{bmatrix} \begin{bmatrix}
e_{1t} \\
e_{2t}
\end{bmatrix}
\]

The univariate process for \( y_{1t} \) will be

\[
\begin{bmatrix}
1 - (5/6)L + (1/9)L^2
\end{bmatrix} y_{1t} = e_{1t} - (1/2)e_{1t-1} - (1/6)e_{2t-1}
\]

The right hand side is the sum of two component MA(1) processes, which will also be another MA(1) process, denoted by \( u_t \). Then

\[
u_{1t} = e_{1t} - (1/2)e_{1t-1} - (1/6)e_{2t-1} = \xi_t + \theta \xi_{t-1}
\]

\[
\gamma_u (0) = \left(1 + (1/4)\right) \sigma_1^2 + \left(1/36\right) \sigma_2^2 + \left(1/6\right) \sigma_{12}^2 = \left(1 + \theta^2\right) \sigma^2_\xi
\]

\[
\gamma_u (1) = -(1/2) \sigma_1^2 - (1/6) \sigma_{12}^2 = \theta \sigma^2_\xi
\]

To make progress with the link between the two component MA(1) processes and the aggregate one, \( \xi_t \), is difficult since it involves solving non linear equations. However, in order to provide a simple illustration, suppose
\[ E(\varepsilon \varepsilon') = \Omega \text{ and } \Omega = \begin{bmatrix} 4 & 0 \\ 0 & 36 \end{bmatrix} \]

Then \[ \gamma_u(0) = 6 = (1 + \theta^2)\sigma_{\xi}^2 \text{ and } \gamma_u(1) = -2 = \theta\sigma_{\xi}^2, \] which gives the first order autocorrelation coefficient on the aggregate MA(1) process to be

\[ \rho_u(1) = -\left(\frac{1}{3}\right) = \left(\frac{\theta}{1 + \theta^2}\right), \]

and on solving \((1/3)\theta^2 + \theta + (1/3) = 0\) gives \(\theta = -2.618\) and \(\theta = -0.382\). On taking the invertible root of \(-0.382\), leads to a solution for the variance of the white noise process of the component MA(1) to be \(\sigma_{\xi}^2 = (-2)/(-0.382) = 5.236\). Hence \(y_{tr} \sim ARMA(2,1)\) process

\[ y_{tr} = (5/6)y_{tr-1} - (1/9)y_{tr-2} + \xi_t - 0.382\xi_{t-1} \text{ and } \sigma_{\xi}^2 = 5.236 \]

Note that the variance of the innovation (i.e. one step ahead prediction error) is 5.236 which exceeds that of the corresponding innovation variance of 4 in the 'structural' model. This is an illustration of the fact that going to the “reduced form” univariate model for one variable will generally be inefficient, since it will have larger disturbance variance, which is the same as the one step ahead prediction MSE. A similar situation also arises for \(y_{2tr}\).

**Interpretation of Impulse Response Weights**

The standard VAR(p) model is set up so that only lagged values of other variables enter each equation and explain the left hand side variable. Any contemporaneous relationship between the elements of \(y_t\) will then be present in the off diagonal elements of the disturbance covariance matrix \(\Omega\). In the analysis of MARs and impulse response weights from an estimated VAR(p), the conventional MAR of
\[ y_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \]

has the normalization of \( B_0 = I \) and \( E(\varepsilon, \varepsilon') = \Omega \). Sims (1980) and others generally renormalize by premultiplying through (21) by a lower triangular matrix \( R \) to obtain

\[
R y_t = \sum_{j=1}^{p} (RA_j) y_{t-j} + \xi_t
\]

where \( \xi_t = Re_t \), and since \( R\Omega R' = I \), it follows that \( E(\xi_t) = 0 \), \( E(\xi_t \xi_{t'}) = 0 \), \( s \neq t \) and \( E(\xi_t \xi_{t'}) = I \). It also follows that \( R'R = \Omega^{-1} \). The advantage with this representation is that the effect of different innovations at different lags can be more clearly detected. Alternatively it is also possible to use another lower triangular matrix \( R \) which is defined so that

\[
\bar{R}D\bar{R} = \Omega
\]

and \( D \) is a diagonal matrix, although in most applications \( \bar{R} \) is usually chosen so that the covariance matrix of the disturbances will have an identity matrix. On returning to the previous example

\[
\begin{bmatrix}
    y_{1t} \\
    y_{2t}
\end{bmatrix} = \begin{bmatrix}
    (1/3) & -(1/6) \\
    -(1/3) & (1/2)
\end{bmatrix} \begin{bmatrix}
    y_{1t-1} \\
    y_{2t-1}
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

where, \( E(\varepsilon, \varepsilon') = \Omega = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \). We choose a matrix \( R \) such that \( R = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \), then

\[
R\Omega R' = I \quad \text{and} \quad R'R = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} = \Omega^{-1}
\]
The VAR(1) can then be expressed as

\[ y_{1t} = (1/3)y_{1t-1} - (1/6)y_{2t-1} + \xi_{1t} \]

\[ y_{2t} - 2y_{tt} = -(1/3)y_{tt-1} + (1/2)y_{2t-1} + \xi_{2t}, \]

where the new innovations \( \xi_{1t} \) and \( \xi_{2t} \) are not only uncorrelated at all lags, but contemporaneously also. The contemporaneous part of the relationship between \( y_{1t} \) and \( y_{2t} \) is now apparent in the second equation rather than in the matrix \( \Omega \).
Appendix: Conditions for Stationarity of the VAR(p)

The VAR(1) in companion form is \( Y_t = \sum_{j=0}^{t-1} A^j \zeta_{t-j} + A^t Y_0 \). However, only the first \( g \) rows containing the VAR(p) are of direct interest and can be seen to be

\[
y_t = \sum_{j=0}^{t-1} (N^t A^j N) \varepsilon_{t-j} + (N^t A)^t Y_0
\]

In order for \( y_t \) to be stationary it must be time invariant, so that a necessary condition is for

\[
\lim_{t \to \infty} (A)^t = 0.
\]

If \( A \) has \( g \) distinct eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_g \), which are stacked into the diagonal matrix,

\[
\Lambda = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \lambda_g
\end{bmatrix}
\]

and on defining \( H \) as the corresponding matrix of eigenvectors, so that \( H' = H^{-1} \), and \( A^{-1} = H \Lambda^{-1} H' \), then \( A' = \left( H \Lambda^{-1} H' \right) \left( H \Lambda^{-1} H \right) \ldots \left( H \Lambda^{-1} H \right) = \left( H \Lambda^{-1} H \right)^{t-1} \). In order to determine the eigenvalues of \( A \), it is therefore necessary to solve the determinental equation

\[
|A - \lambda I| = 0,
\]

and requiring all the eigenvalues to be inside the unit circle. This is equivalent to all the roots of \( |A(L)| \) to lie outside the unit circle. For example, consider the VAR(3), then
\[ |A - \lambda I| = \begin{vmatrix} A_1 & A_2 & A_3 \\ I & 0 & 0 \\ 0 & I & 0 \end{vmatrix} - \lambda \begin{vmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{vmatrix} = |A_1 - \lambda I - \lambda I 0 | \]

\[ = A_1 + \lambda A_2 + \lambda^2 A_3 - \lambda^3 I = \lambda^3 I - \sum_{j=1}^{\infty} \lambda^{3-j} A_j = 0 \]

In the general case for the VAR(p):

\[ |A - \lambda I| = \begin{vmatrix} A_1 - \lambda I & A_2 & A_3 & \cdots & A_p \\ I & -\lambda I & 0 & \cdots & 0 \\ 0 & I & -\lambda I & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & I -\lambda I \end{vmatrix} \]

\[ = |\lambda p I - \lambda^{p-1} A_1 - \lambda^{p-2} A_2 - \cdots - A_p| = |\lambda p I - \sum_{j=1}^{p} \lambda^{p-j} A_j| = 0 \]

The condition that all the roots of the above lie inside the unit circle is then equivalent to the condition for the VAR(p) model to be covariance stationary is for all the roots of

\[ |A(L)| = I - \sum_{j=1}^{p} A_j L^j \] to lie outside the unit circle.