Lecture 6: Multivariate Time Series and VARs

The theory on linear time series models developed for the univariate case extends in a natural way to the multivariate case.

Let \( \mathbf{x}_t = (x_{1,t}, \ldots, x_{k,t}) \) be a \( k \)-dimensional vector of univariate time series. Then \( x_t \) is weakly stationary if

\[
E x_t = \mu \quad \forall t
\]

and \( E (x_t - \mu)(x_{t+k} - \mu)' = \Gamma(k) \forall t \) exists and \( \|\Gamma(0)\| < \infty \) where \( \|A\| = (\text{tr} \ A A')^{1/2} \) is the euclidean matrix norm. It follows immediately for any \( n \) and any vectors \( a_1, \ldots, a_n \) that \( \sum_{i=1}^n \sum_{l=1}^n a_i \Gamma(i-l) a_l \geq 0 \). The spectral density matrix of \( y_t \) is defined as

\[
f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-i\lambda h}
\]

provided \( \sum_{h=-\infty}^{\infty} \|\Gamma(h)\| < \infty \). Note that the diagonal elements of \( f(\lambda) \) are the univariate spectral densities of \( x_{it} \). The off-diagonal elements of \( f(\lambda) \) are called the cross spectra between \( x_{it} \) and \( x_{mt} \). Using \( \gamma_{l,m}(h) = E (x_{lt} - \mu_l)(x_{mt+h} - \mu_m) \)

\[
[f(\lambda)]_{l,m} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{l,m}(h) e^{-i\lambda h}.
\]

Note that \( \gamma_{l,m}(h) \neq \gamma_{l,m}(-h) \) in general so that the off-diagonal elements of \( f(\lambda) \) are in general complex valued.

As for the univariate case there is an infinite moving average representation of \( x_t \). Assume that \( x_t \) is purely non-deterministic and weakly stationary, then

\[
y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} = \mu + \Psi(L) \varepsilon_t,
\]

where \( \Psi(L) = \sum_{j=0}^{\infty} \Psi_j L^j \) and \( \varepsilon_t \) is a multivariate sequence of white-noise processes such that

\[
E \varepsilon_t = 0
\]

\[
E \varepsilon_t \varepsilon'_t = \Sigma
\]

and \( E \varepsilon_t \varepsilon_j = 0 \) for \( t \neq s \).

The coefficients matrices \( \Psi_j \) of dimension \( k \times k \) satisfy \( \sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty \). If the polynomial \( \Psi(L) \) can be approximated by a rational matrix polynomial \( \Phi(L)^{-1} \Theta(L) \) then the model has an ARMA representation

\[
\Phi(L)(x_t - \mu) = \Theta(L) \varepsilon_t.
\]

The vector ARMA model is causal if \( \Phi(L)^{-1} \) is well defined, i.e., if it has a convergent power series expansion. This is the case if \( \Phi(z) \) is invertible for \( |z| \leq 1 \) or if \( \det \Phi(z) \neq 0 \) for \( |z| \leq 1 \).

In the same way the ARMA representation is invertible if \( \det \Theta(z) \neq 0 \) for \( |z| \leq 1 \). We can then write

\[
\Theta(L)^{-1} \Phi(L)(x_t - \mu) = \varepsilon_t.
\]
or

\[(x_t - \mu) = \sum_{i=1}^{\infty} \Pi_i (x_{t-i} - \mu) + \varepsilon_t\]

where \(I - \Pi(L) = I - \sum_{i=1}^{\infty} \Pi_i L^i = \Theta(L)^{-1}\Phi(L)\).

In practice, it is usually assumed that \(\Pi(L)\) can be approximated by a finite order polynomial. This leads to the VAR\((p)\) model

\[y_t = \Pi_1 y_{t-1} + \ldots + \Pi_p y_{t-p} + \varepsilon_t,\]

where \(y_t = x_t - \mu\). The VAR\((p)\) model can be represented in companion form by stacking the vectors \(y_t\) in the following way

\[
\begin{bmatrix}
y_t \\
\vdots \\
y_{t-p+1}
\end{bmatrix} =
\begin{bmatrix}
\Pi_1 & \cdots & \Pi_{p-1} & \Pi_p \\
I & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & I & 0
\end{bmatrix}
\begin{bmatrix}
y_{t-1} \\
\vdots \\
y_{t-p}
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_t \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

The autocovariance function of \(y_t\) can be found from considering the Yule Walker equations

\[
\Gamma(0) = E y_t y_t' = \sum_{i=1}^{p} \Pi_i E y_{t-i} y_t' + E \varepsilon_t y_t'
\]

\[
= \sum_{i=1}^{p} \Pi_i \Gamma(-i)' + \Sigma
\]

\[
= \sum_{i=1}^{p} \Gamma(-i) \Pi_i' + \Sigma
\]

and

\[
\Gamma(h) = \sum_{i=1}^{p} \Gamma(h-i) \Pi_i'
\]

by stacking \(\Pi' = [\Pi_1, \ldots, \Pi_p]\) and \(\Gamma_p' = [\Gamma(-1) \ldots \Gamma(-p)]\). These equations can be written as

\[
\Gamma(0) = \Gamma_p' \Pi + \Sigma
\]

and

\[
\Gamma_p = \Gamma_p' \Pi
\]

with

\[(\Gamma_p')_{ij} = \Gamma(i-j)\]

such that

\[
\Sigma = \Gamma(0) - \Pi' \Gamma_p' \Pi.
\]

For the AR\((1)\) case we have

\[
\Gamma(0) = \Gamma(-1) \Pi_1' + \Sigma
\]

\[
\Gamma(h) = \Gamma(h-1) \Pi_1'
\]

and

\[
\Gamma(1) = \Gamma(0) \Pi_1' \text{ or } \Gamma(-1) = \Pi_1 \Gamma(0)
\]

such that \(\Gamma(0) = \Pi_1 \Gamma(0) \Pi_1' + \Sigma\) and

\[
\text{vec} \Gamma(0) = (\Pi_1 \otimes \Pi_1) \text{vec} \Gamma(0) + \text{vec} \Sigma
\]
solving for $\text{vec}\Gamma(0)$ gives us

$$\text{vec}\Gamma(0) = (1 - \Pi_1 \otimes \Pi_1)^{-1} \text{vec}\Sigma.$$ 

The best linear predictor for the VAR($p$) model can be found in the same way as for the univariate case

$$\hat{y}_{t+1} = P_M y_{t+1} = \Pi_1 y_t + ... + \Pi_p y_{t-p+1}$$

$$= \Pi(L) y_{t+1}$$

$$= \Pi(L) (I - \Pi(L))^{-1} \varepsilon_{t+1}$$

$$= (\Psi(L) - I) \varepsilon_{t+1}$$

$$= \sum_{s=1}^{\infty} \Psi_s \varepsilon_{t-s+1},$$

where $\Psi(L) = \sum_{s=0}^{\infty} \Psi_s L^s = (I - \Pi(L))^{-1}$ and for the $h$-step ahead prediction error we have

$$\hat{y}_{t+h} = P_M y_{t+h} = \sum_{s=h}^{\infty} \Psi_s \varepsilon_{t-s+h}$$

with prediction error

$$y_{t+h} - \hat{y}_{t+h} = \sum_{s=0}^{h-1} \Psi_s \varepsilon_{t-s+h}.$$ 

The prediction error therefore has variance

$$\text{var} (y_{t+h} - \hat{y}_{t+h}) = \Sigma + \sum_{s=1}^{h-1} \Psi_s \Sigma \Psi'_s$$

6.1. Estimation of a VAR($p$)

We stack the vectors $y_t$ such that

$$x'_1 \in \mathbb{R}^{kp} = \begin{pmatrix} y'_{t-1}, y'_{t-2}, ..., y'_{t-p} \end{pmatrix} \quad \text{and} \quad \Pi' = \begin{bmatrix} \Pi_1, ..., \Pi_p \end{bmatrix}$$

then we can write

$$y_t = \Pi' x_t + \varepsilon_t$$

or $y'_t = x'_t \Pi + \varepsilon'_t$. We stack the variables into matrices

$$Y = \begin{bmatrix} y'_1 \\ \vdots \\ y'_T \end{bmatrix}, \quad X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_T \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_T \end{bmatrix}$$

such that the model can be written as

$$Y = X \Pi + \varepsilon.$$ 

In vectorized form this is

$$\text{vec} Y = (I \otimes X) \text{vec} \Pi + \text{vec} \varepsilon.$$ 

Note that $E(\text{vec} \varepsilon)(\text{vec} \varepsilon)' = \Sigma \otimes I_T$. The likelihood is then approximately proportional to

$$-\frac{Tk}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma| - \frac{1}{2} (\text{vec} \varepsilon)' (\Sigma^{-1} \otimes I_T) \text{vec} \varepsilon.$$ 

3
where
\[
(\text{vec } \varepsilon) \left( \Sigma^{-1} \otimes I_T \right) \text{vec } \varepsilon = \text{tr } \Sigma^{-1} \varepsilon \varepsilon' = \text{tr } \Sigma^{-1} (Y - \Pi T)(Y - \Pi T)'.
\]

The ML estimator is now immediately seen to be
\[
\text{vec } \hat{\Pi} = \left[ (I \otimes X') \left( \Sigma^{-1} \otimes I \right) (I \otimes X) \right]^{-1} (I \otimes X') \left( \Sigma^{-1} \otimes I \right) \text{vec } y
\]
\[
= (\Sigma^{-1} \otimes X'X)^{-1} \left( \Sigma^{-1} \otimes X' \right) \text{vec } y
\]
\[
= \left( I \otimes \left( X'X \right)^{-1} X' \right) \text{vec } y
\]
which shows that the ML estimator is equivalent to OLS carried out equation by equation.

Now
\[
\left( \text{vec } (\hat{\Pi} - \Pi) \right) = \left( I_k \otimes \left( X'X \right)^{-1} X' \right) \text{vec } \varepsilon
\]
\[
= \left( I \otimes \left( X'X \right)^{-1} \right) \left( I_k \otimes X' \right) \text{vec } \varepsilon,
\]
where \( \left( I \otimes \left( X'X \right)^{-1} \right) \xrightarrow{p} I \otimes \Gamma_p^{-1} \) and \( \frac{1}{\sqrt{T}} \left( I \otimes X' \right) \text{vec } \varepsilon \xrightarrow{d} N(0, \Sigma \otimes \Gamma_p) \) with \( \Gamma_p = Ex_t x'_t. \) This follows from noting that
\[
(I \otimes X) \text{vec } \varepsilon = \begin{bmatrix}
x_t \varepsilon_{1t} \\
\vdots \\
x_t \varepsilon_{kt}
\end{bmatrix},
\]
and
\[
\text{var } x_t \varepsilon_{lt} \varepsilon_{js} = \begin{cases}
0 & \text{if } t \neq s \\
\sigma_{tj} \Gamma_p & \text{otherwise.}
\end{cases}
\]
Therefore the distribution of the parameter estimates is asymptotically
\[
\sqrt{T} \text{vec } (\hat{\Pi} - \Pi) \xrightarrow{d} N \left(0, (I \otimes \Gamma_p^{-1}) \left( \Sigma \otimes \Gamma_p \right) (I \otimes \Gamma_p^{-1}) \right) = N \left(0, \Sigma \otimes \Gamma_p^{-1} \right).
\]

If we have blockwise restrictions as in the case of Granger-causality then we still can estimate the system equation by equation. If we have more general restrictions then we need to estimate the full system.

### 6.2. Prediction error variance decomposition

If we want to analyze the contributions of the error terms \( \varepsilon_t \) to the total forecast error variance then we need to orthogonalize the system. Let \( E \varepsilon_t \varepsilon'_t = \Sigma \) and \( R \Sigma R' = I \) where \( R \) is lower triangular. Then \( ER \varepsilon_t \varepsilon'_t R' = E \eta_t \eta'_t = I. \) We now look at the transformed model
\[
y_t = \sum_{j=0}^{\infty} \Psi_j R^{-1} \varepsilon_{t-j} = \sum C_j \eta_{t-j}.
\]
The forecast error of an \( h \)-step ahead forecast now is
\[
\text{var } (y_{t+h} - \hat{y}_{t+h}) = \Sigma + \sum_{j=1}^{h-1} C_j C_j' = \Sigma + \sum_{j=1}^{h-1} \Psi_j R^{-1} R^{-1} \Psi_j.'
\]
The coefficients $C_j$ can be obtained from

$$C_j = J' A^i J R^{-1}$$

where

$$J' = [I_k, 0, \ldots, 0], \quad A = \begin{bmatrix} \Pi_1 & \cdots & \Pi_{p-1} & \Pi_p \\ I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ I & 0 \\ \end{bmatrix}.$$ 

Then, according to Sims (1981), the proportion of the $h$-step ahead forecast error variance in variable $l$ accounted for by innovations in variable $\eta_{i,t}$ is given by

$$r^2_{l,i} + \sum_{j=1}^{h-1} c^2_{i,j},$$

where $r_{l,i} = [R^{-1}]_{li}$ and $c_{i,j} = [C_j]_{li}$. To see this, note that the forecast error is given by

$$y_{t+h} - \hat{y}_{t+h} = R^{-1} \eta_t + \sum_{j=1}^{h-1} C_j \eta_{t-j},$$

while the forecast error resulting from $\eta_{i,t-j}$ is given by

$$R^{-1}_{i} \eta_{i,t} + \sum_{j=1}^{h-1} C_{i,j} \eta_{i,t-j}$$

where $R^{-1}$ is the $i$th column of $R^{-1}$ and $C_{i,j}$ is the $i^{th}$ column of $C_j$. The relative forecast error variance is then given by

$$\frac{r^2_{l,i} + \sum_{j=1}^{h-1} c^2_{i,j}}{\text{var} \left( y_{t+h} - \hat{y}_{t+h} \right)}$$

where $\text{var} \left( y_{t+h} - \hat{y}_{t+h} \right)$ is the $l^{th}$ diagonal element of $\text{var} \left( y_{t+h} - \hat{y}_{t+h} \right)$.

If one is interested in innovations in the original variables rather than the orthogonalized innovations $\eta_t$ then an identification scheme as discussed in the next section is needed. In particular, since $\eta_t = R \varepsilon_t$ with $R$ lower triangular, we can identify the first element of $\varepsilon_t$ with the first element of $\eta_t$. Since the ordering of the vectors $y_t$ is arbitrary this identification scheme applies to all elements of $\varepsilon_t$.

6.3. Impulse response functions

Closely related to the concept of error variance decomposition is the concept of an impulse response function. We are interested in the effect of a shock $\varepsilon_{it}$ onto the variable $y_{t,t+h}$. Using the MA($\infty$) representation for $y_{t+h}$ we find

$$y_{t+h} = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t+h-j}$$

so the impact of $\varepsilon_t$ onto $y_{t+h}$ is $\Psi_h \varepsilon_t$. If we are interested in a unit variance shock of $\varepsilon_{it}$ then we need to take into account the fact that $\varepsilon_{it}$ is correlated with other shocks. This is again done by orthogonalizing the innovations

$$\eta_t = R \varepsilon_t.$$
where $R$ is lower triangular such that and $\eta_t$ is orthogonal. Then $\varepsilon_t = R^{-1}\eta_t$ and in particular, $\varepsilon_{1t} = r_{11}\eta_{1t}$.

Since the ordering of the variables is arbitrary we can restrict attention to the first innovation without loss of generality. Note that once the value of $\varepsilon_{1t}$ is fixed, the value of the other innovations follow from $\varepsilon_t = R^{-1}\eta_t$ where we set $\eta_{2t} = ... = \eta_{kt} = 0$. The impact of $\varepsilon_{1t}$ onto $y_{t+h}$ is therefore

$$[\Psi_h R^{-1}]_{1,1}$$

The impact onto variable $l$ is then

$$[\Psi_h R^{-1}]_{l,1}. $$

Typically the ordering of the variables is chosen to reflect a certain structure of shocks in the economy. For example, if we want to model monetary policy shocks as the original source of randomness then we would place a monetary variable in the first equation.

### 6.4. Granger causality

Assume that $y_t = (y_{1t}, y_{2t})$ is partitioned into two subvectors. Granger defined the concept of causality in terms of forecast performance. In this sense $y_{1t}$ does not $g$-cause $y_{2t}$ if it does not help in predicting $y_{2t}$. Formally,

**Definition 6.1 (Granger Causality).** Let $y_t$ be a stationary process. Define the linear subspaces

$$\mathcal{M}^t_1 = \mathcal{S}\{y_{1s}, s \leq t\}$$

$$\mathcal{M}^t_2 = \mathcal{S}\{y_{2s}, s \leq t\}$$

Then $y_{1t}$ causes $y_{2t}$ if

$$P_{\mathcal{M}^t_1 \cup \mathcal{M}^t_{2-1}}(y_{2t}) \neq P_{\mathcal{M}^t_{2-1}}(y_{2t})$$

and $y_{1t}$ causes $y_{2t}$ instantaneously if

$$P_{\mathcal{M}^t_1 \cup \mathcal{M}^t_{2-1}}(y_{2t}) \neq P_{\mathcal{M}^t_{2-1}}(y_{2t}).$$

It follows at once from the definition of Granger causality and the projection theorem that $y_{1t}$ does not $g$-cause $y_{2t}$ if

$$\text{var}(\varepsilon^1_t) = \text{var}(\varepsilon^2_t)$$

where $\varepsilon^1_t = y_{2t} - P_{\mathcal{M}^t_1 \cup \mathcal{M}^t_{2-1}}(y_{2t})$ and $\varepsilon^2_t = y_{2t} - P_{\mathcal{M}^t_{2-1}}(y_{2t})$. Another way to characterize Granger noncausality is by noting that

$$\text{cov}(y_{2t}, y_{1t-h} - P_{\mathcal{M}^t_{2-1}}(y_{1t-h})) = 0 \text{ for } h > 0.$$  

To see this note that by Granger causality

$$y_{2t} - P_{\mathcal{M}^t_{2-1}}(y_{2t}) \perp \mathcal{M}^t_{1-1} \cup \mathcal{M}^t_{2-1}$$

which implies that

$$\text{cov}(y_{2t} - P_{\mathcal{M}^t_{2-1}}(y_{2t}), y_{1t-h} - P_{\mathcal{M}^t_{2-1}}(y_{1t-h})) = 0 \text{ for } h > 0$$

since $y_{1t-h} - P_{\mathcal{M}^t_{2-1}}(y_{1t-h}) \in \mathcal{M}^t_{1-1} \cup \mathcal{M}^t_{2-1}$. But by the projection theorem it follows that $y_{1t-h} - P_{\mathcal{M}^t_{2-1}}(y_{1t-h}) \perp \mathcal{M}^t_{2-1}$ such that

$$\text{cov}(P_{\mathcal{M}^t_{2-1}}(y_{2t}), y_{1t-h} - P_{\mathcal{M}^t_{2-1}}(y_{1t-h})) = 0.$$  

It has to be emphasized that this notion of causality is strongly related to the notion of sequentivality in the sense that an event causing another event has to precede it in time. Moreover, the definition really is in terms of correlation rather than causation. Finding evidence of Granger causality can be an artifact of a spurious correlation. On the other hand, lack of Granger causality can be misleading too if the true causal link is of nonlinear form.

An alternative definition of causality is due to Sims.
Definition 6.2 (Sims Causality). For $y_{1t}$ and $y_{2t}$ stationary we say that $y_{1t}$ does not cause $y_{2t}$ if

$$\text{cov}(y_{2t+j}, y_{1t} - P_{M^2_1}(y_{1t})) = 0 \text{ for all } j \geq 1.$$ 

It can be seen immediately that this definition implies that all the coefficients $d_j$ for $j < 0$ in the projection

$$y_{1t} = \sum_{j=-\infty}^{\infty} d_j y_{2t-j} + w_t$$

are zero and the projection residual $w_t$ is uncorrelated with all future values $y_{2t+j}$.

Theorem 6.3. Granger Causality and Sims Causality are equivalent.

Proof. Assume $y^1_t$ does not Granger cause $y^2_t$. Then

$$\text{cov}(y_{2t}, y_{1t-h} - P_{M^2_{1-h}}(y_{1t-h})) = 0 \text{ for } h > 0.$$ 

Note that $P_{M^2_{1-h}}(y_{2t}) = P_{M^2_{1-h}} \big( P_{M^1_{1-h}} \big( y_{2t} \big) \big) = P_{M^1_{1-h} \cup M^2_{1-h}}(y_{2t})$ such that $y_{2t} - P_{M^2_{1-h}}(y_{2t}) \perp M^1_{t-1} \cup M^2_{t-1}$ for $h > 0$ such that

$$\text{cov}(y_{2t}, y_{1t-h} - P_{M^2_{1-h}}(y_{1t-h})) = 0 \text{ for } h > 0. \hspace{1cm} (6.1)$$

By stationarity this is equivalent to

$$\text{cov}(y_{2t+h}, y_{1t} - P_{M^2_1}(y_{1t})) = 0 \text{ for } h > 0.$$ 

The reverse implication follows from the fact that by (6.1) $y_{2t} - P_{M^2_{1-h}}(y_{2t}) \perp M^1_{t-1} \cup M^2_{t-1}$ which corresponds to Granger causality. ■

6.5. Granger causality in a VAR

Let

$$\begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where $\varepsilon_{1t}$ and $\varepsilon_{2s}$ are uncorrelated for all $t$ and $s$. Then $y_{1t}$ fails to Granger cause $y_{2t}$ if $\Phi_{21}(L) = 0$. This follows from the fact that

$$P_{M^1_{t-1} \cup M^2_{t-1}}(y_{2t}) = (\Phi_{22}(L) - I) y_{2t} + \Phi_{21}(L)y_{1t} = P_{M^2_{t-1}}(y_{2t})$$

if and only if $\Phi_{21}(L) = 0$. If $\Phi(L)^{-1}$ exists then the MA($\infty$) representation of the system is

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \Phi^{-1}(L) \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \Psi(L) \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Thus $\Phi(L)\Psi(L) = I$ so in particular

$$\Phi_{21}(L)\Psi_{11}(L) + \Phi_{22}(L)\Psi_{21}(L) = 0$$

which implies $\Psi_{21}(L) = 0$ if $\Phi_{22}(L) \neq 0$. We see that

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ 0 & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$
if \( y_{1t} \) fails to Granger cause \( y_{2t} \). Thus we have
\[
y_{1t} = \Psi_{11}(L)\varepsilon_{1t} + \Psi_{12}(L)\varepsilon_{2t}
\]
and \( \Phi_{22}(L)y_{2t} = \varepsilon_{2t} \) so \( y_{1t} = \Psi_{11}(L)\varepsilon_{1t} + \Psi_{12}(L)\Phi_{22}(L)y_{2t} \). Then
\[
P_{\psi_t}^2(y_{1t}) = \Psi_{12}(L)\Phi_{22}(L)y_{2t}
\]
since \( \Psi_{11}(L)\varepsilon_{1t} \) is orthogonal to \( M_t^2 \).

It now follows immediately that
\[
\text{cov} (y_{2t+j}, y_{1t} - \Psi_{12}(L)\Phi_{22}(L)y_{2t}) = \text{cov} (y_{2t+j}, \Psi_{11}(L)\varepsilon_{1t}) = 0 \quad \forall \ j > 0
\]

This establishes that Granger causality implies Sims causality. It can also be shown that Sims causality implies Granger causality. It thus follows that the two concepts are equivalent.

We can test the null of Granger non-causality by estimating the unrestricted VAR equation by equation by OLS and then test if the coefficients \( \Pi_{21} \ldots \Pi_{21p} \) in
\[
y_{2t} = \hat{c}_2 + \Pi_{21}y_{1t-1} + \ldots + \Pi_{21p}y_{1t-p} + \Pi_{221}y_{2t-1} + \ldots + \Pi_{22p}y_{2t-p}
\]
are jointly significantly different from zero. For a bivariate system this can be carried out by a standard \( F \)-test. Calculate the unrestricted residuals \( \text{RSS}_1 = \sum \hat{\varepsilon}_t^2 \) and residuals from the restricted regression
\[
y_{2t} = \hat{c}_2 + \Pi_{22}y_{2t-1} + \ldots + \Pi_{22p}y_{2t-p}.
\]
as \( \text{RSS}_0 = \sum \hat{\varepsilon}_t^2 \). Then under normality
\[
\frac{\text{RSS}_0 - \text{RSS}_1}{\text{RSS}_1}(T-2p-1)/p \sim F(p, T-2p-1)
\]

An asymptotically equivalent test is \( T(\text{RSS}_0 - \text{RSS}_1)/\text{RSS}_1 \overset{d}{\sim} \chi_p^2 \) under the null hypothesis of no Granger causality.

### 6.6. Structural VARs

Assume we have a structural economic model which is driven by behavioral sources of variation collected in a vector process \( \varepsilon(t) \). The structural model connects economic variables to current and past values of driving shocks
\[
\sum_{s=0}^{\infty} A_s y_{t-s} = \sum_{s=0}^{\infty} B_s \varepsilon_{t-s}
\]

(6.2)

We also assume that \( Y_t \) has an equivalent VAR(\( \infty \)) representation
\[
y_t = \sum \Pi_s y_{t-s} + u_t
\]

(6.3)

If the number of elements in \( \varepsilon_t \) is equal to the number of elements in \( y_t \) and if knowledge of \( A_s \) and \( B_s \) is enough to solve for \( \varepsilon_t \) in terms of lagged \( y_t \) then we can write
\[
B(L)^{-1} A(L) y_t = B_0 \varepsilon_t
\]

(6.4)

with
\[
B(L) = I + \sum_{i=1}^{\infty} B_i B_0^{-1} L^i
\]

where \( B(L)^{-1} \) has a polynomial expansion \( \sum_{i=0}^{\infty} C_i L^i \) and
\[
B(L)^{-1} A(L) = B(L)^{-1} [A_0 + (A(L) - A_0)].
\]
Since \( B(L)^{-1} \) satisfies \( I + \sum_{i=1}^{\infty} B_i B_0^{-1} L^i \) \( \sum_{i=0}^{\infty} C_i L^i \) = \( I \) it must hold that \( C_0 = I \). This establishes that (6.4) is

\[
A_0 y_t + \sum_{s=1}^{\infty} \tilde{C}_s y_{t-s} = B_0 \varepsilon_t
\]

with \( \tilde{C}_s \) such that

\[
\sum_{i=1}^{\infty} C_i A_0 L^i + B(L)^{-1} (A(L) - A_0) = \sum_{s=1}^{\infty} \tilde{C}_s L^s.
\]

Substitution from (6.3) then gives

\[
A_0 u_t + \sum_{s=1}^{\infty} (A_0 \Pi_s + \tilde{C}_s) y_{t-s} = B_0 \varepsilon_t
\]

If the structural and reduced forms are identical it has to hold that \( A_0 \Pi_s = -\tilde{C}_s \). Then the unrestricted innovations of the VAR are related to the behavioral innovations \( \varepsilon_t \) by

\[
u_t = A_0^{-1} B_0 \varepsilon_t.\]

Note that \( \Pi_s \) are unrestricted reduced form parameters that can always be estimated from the data. If the theoretical model (6.2) does not restrict the dynamics of the system then we can always set \( \Pi_s = -A_0^{-1} \tilde{C}_s \). Identification of the system then reduces to finding the matrices \( A_0 \) and \( B_0 \).

Since we can consistently estimate the reduced form residuals \( \hat{u}_t \) we can estimate \( \Sigma = \text{var}(u_t) \) by

\[
\hat{\Sigma} = \frac{1}{T} \sum \hat{u}_t \hat{u}_t'.
\]

If we now impose the restriction that the policy disturbances \( \varepsilon_t \) be uncorrelated and that \( \Omega = \text{var}(\varepsilon_t) \) is diagonal and that \( B_0 = I \) then

\[
\Sigma = A_0 \Omega A_0'.
\]

The matrix \( A_0 \) can then be identified by imposing that it is lower triangular. In other words, if the only restrictions on the system are that \( A_0 \) is the lower triangular and that \( \Omega \) is diagonal then the structural VAR is just identified.

It is clear that the just identified case with triangular matrix is only one of many possibilities to identify \( A_0 \).

Another interesting example is Blanchard and Quah’s decomposition. Their goal is to decompose GNP into permanent and transitory shocks. They postulate that demand side shocks have only temporary effects on GNP while supply side or technology shocks have permanent effects. Unemployment on the other hand is affected by both shocks. They postulate

\[
\begin{bmatrix}
\Delta Y_t \\
u_t
\end{bmatrix} = \begin{bmatrix}
c_{11}(L) & c_{12}(L) \\
c_{21}(L) & c_{22}(L)
\end{bmatrix} \begin{bmatrix}
\varepsilon_{dt} \\
\varepsilon_{st}
\end{bmatrix},
\]

with \( c_{11}(1) = 0 \) such that \( \varepsilon_{dt} \) has no long-run effect on \( \Delta Y_t \). Also assume E\( \varepsilon_{dt} \varepsilon_{st}' = I \).

The VAR(\( p \)) representation of the system is

\[
\begin{bmatrix}
\Delta Y_t \\
u_t
\end{bmatrix} = \begin{bmatrix}
a_{11}(L) & a_{12}(L) \\
a_{21}(L) & a_{22}(L)
\end{bmatrix} \begin{bmatrix}
\Delta Y_{t-1} \\
u_{t-1}
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

Since in this case \( A_0 = I \), it follows that

\[
\begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix} = \begin{bmatrix}
c_{11}(0) & c_{12}(0) \\
c_{21}(0) & c_{22}(0)
\end{bmatrix} \begin{bmatrix}
\varepsilon_{dt} \\
\varepsilon_{st}
\end{bmatrix}
\]
The goal is to estimate the structural residuals $\varepsilon_t$ which can be done if we know the coefficients of the matrix $C_0 = C(0)$. From $E\eta_t\eta_t^\prime = \Sigma$ we have

$$\Sigma = C_0 C_0^\prime$$

i.e.,

$$\begin{align*}
\text{var } \eta_1 &= c_{11}(0)^2 + c_{12}(0)^2 \\
\text{var } \eta_2 &= c_{21}(0)^2 + c_{22}(0)^2 \\
\text{cov } (\eta_1, \eta_2) &= c_{11}(0)c_{21}(0) + c_{12}(0)c_{22}(0)
\end{align*}$$

These are three restrictions for four variables. The fourth restriction can be obtained from the long-run restriction $c_{11}(1) = 0$. Note that $(I - A(L)L)^{-1} C_0 = C(L)$ by the MA(∞) representation of the VAR and $\eta_t = C_0 \varepsilon_t$. So in particular, $(I - A(1))^{-1} C_0 = C(1)$. Now

$$(I - A(1))^{-1} = \frac{1}{D} \begin{bmatrix} 1 - a_{22}(1) + a_{12}(1) \\ a_{21}(1) \\ a_{21}(1) - a_{11}(1) \end{bmatrix}$$

where $D = \det(I - A(1))$. The upper corner of $C(1)$ is zero by the long run restrictions such that we have an additional equation to determine the coefficients

$$(1 - a_{22}(1))c_{11}(0) + a_{12}(1)c_{21}(0) = 0.$$