Massachusetts Institute of Technology
Department of Economics
Time Series 14.384

Guido Kuersteiner

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 $x_t^{'} = (x_{1,t}, ..., x_{k,t})$ be a k-dimensional vector of univariate time series. Then x_t is weakly stationary if

$$Ex_t = \mu \qquad \forall\, t$$

and $E(x_t - \mu)(x_{t+k} - \mu)^{'} = \Gamma(k)$ $\forall t$ exists and $\|\Gamma(0)\| < \infty$ where $\|A\| = (\text{tr } AA')^{1/2}$ is the euclidean matrix norm. It follows immediately for any n and any vectors $a_1, ..., a_n$ 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 $x_{l,t}$ and x_{mt}. Using $\gamma_{l,m}(h) = E(x_{lt} - \mu_l)(x_{m,t+h} - \mu_m)$

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

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 ε_t is a multivariate sequence of white-noise processes such that

$$\begin{aligned} E\varepsilon_t &= 0 \\ E\varepsilon_t\varepsilon_t^{'} &= \Sigma \\ \text{and } E\varepsilon_t\varepsilon_j^{'} &= 0 \qquad \text{for } t \neq s \end{aligned}$$

The coefficients matrices Ψ_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

$$\Theta(L)^{-1}\Phi(L)\left(x_t - \mu\right) = \varepsilon_t$$

or

$$\left(x_t - \mu\right) = \sum_{i=1}^{\infty} \Pi_i \left(x_{t-i} - \mu\right) + \varepsilon_t$$

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

$$y_t = \Pi_1 y_{t-1} + ... + \Pi_p y_{t-p} + \varepsilon_t,$$

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

$$\begin{array}{ccccc} \left[\begin{array}{c} y_t \\ \vdots \\ y_{t-p+1} \end{array}\right] & = & \left[\begin{array}{cccc} \Pi_1 & \cdots & \Pi_{p-1} & \Pi_p \\ I & & & 0 \\ & \ddots & & \vdots \\ & & I & 0 \end{array}\right] & \left[\begin{array}{c} y_{t-1} \\ \vdots \\ y_{t-p} \end{array}\right] & + & \left[\begin{array}{c} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{array}\right] \\ Y_t & = & A & Y_{t-1} & + & e_t \end{array}$$

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

$$\begin{aligned} \Gamma(0) &= E y_t y_t^{'} = \sum_{i=1}^{p} \Pi_i E y_{t-i} y_t^{'} + E \varepsilon_t y_t^{'} \\ &= \sum_{i=1}^{p} \Pi_i \Gamma(-i)^{'} + \Sigma \\ &= \sum_{i=1}^{p} \Gamma(-i) \Pi_i^{'} + \Sigma \end{aligned}$$

and

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

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

$$\Gamma(0) = \Gamma_p^{\prime} \Pi + \Sigma$$

and

$$\Gamma_p = \Gamma_{(p)} \Pi$$

with

$$\left[\Gamma_{(p)}\right]_{ij} = \Gamma(i-j)$$

such that

$$\Sigma = \Gamma(0) - \Pi^{'} \Gamma_{(p)}^{'} \Pi.$$

For the AR(1) case we have

$$\begin{aligned} \Gamma(0) &= \Gamma(-1)\Pi_1^{\prime} + \Sigma \\ \Gamma(h) &= \Gamma(h-1)\Pi_1^{\prime} \end{aligned}$$

and

$$\Gamma(1) = \Gamma(0)\Pi_1^{'} \text{ or } \Gamma(-1) = \Pi_1\Gamma(0)$$

such that $\Gamma(0) = \Pi_1\Gamma(0)\Pi_1^{'} + \sum$ and

$$\operatorname{vec}\Gamma(0) = (\Pi_1 \otimes \Pi_1)\operatorname{vec}\Gamma(0) + \operatorname{vec}\Sigma$$

solving for $\operatorname{vec}\Gamma(0)$ gives us

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

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

$$\begin{aligned}
\hat{y}_{t+1} &= P_{\mathcal{M}_t} y_{t+1} = \Pi_1 y_t + ... + \Pi_p y_{t-p+1} \\
&= \Pi(L) y_{t+1} \\
&= \Pi(L)\left(I - \Pi(L)\right)^{-1} \varepsilon_{t+1} \\
&= \left(\Psi(L) - I\right)\varepsilon_{t+1} \\
&= \sum_{s=1}^{\infty} \Psi_s \varepsilon_{t-s+1},
\end{aligned}$$

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

$$\hat{y}_{t+h} = P_{\mathcal{M}_t} y_{t+h} = \sum_{s=h}^{\infty} \Psi_s \varepsilon_{t-s+h}$$

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

$$\operatorname{var}\left(y_{t+h} - \hat{y}_{t+h}\right) = \Sigma + \sum_{s=1}^{h-1} \Psi_s \Sigma \Psi_s^{'}$$

6.1. Estimation of a VAR(p)

We stack the vectors y_t such that

$$\underset{1\times kp}{x_t^{'}} = \left(y_{t-1}^{'}, y_{t-2}^{'}, ..., y_{t-p}^{'}\right) \qquad \text{and} \qquad \underset{k\times kp}{\Pi^{'}} = \left[\Pi_1, ..., \Pi_p\right]$$

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

$$Y = \begin{bmatrix} y_1^{'} \\ \vdots \\ y_T^{'} \end{bmatrix}, \; X = \begin{bmatrix} x_1^{'} \\ \vdots \\ x_T^{'} \end{bmatrix}, \; \varepsilon = \begin{bmatrix} \varepsilon_1^{'} \\ \vdots \\ \varepsilon_T^{'} \end{bmatrix}$$

such that the model can be written as

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

In vectorized form this is

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

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

$$-\frac{Tk}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}(\operatorname{vec}\ \varepsilon)^{'}\left(\Sigma^{-1} \otimes I_T\right)\operatorname{vec}\varepsilon,$$

where

$$(\operatorname{vec}\ \varepsilon)^{'}\left(\Sigma^{-1} \otimes I_T\right)\operatorname{vec}\ \varepsilon = \operatorname{tr}\ \Sigma^{-1}\varepsilon\varepsilon^{'} = \operatorname{tr}\ \Sigma^{-1}(Y - X\Pi)(Y - X\Pi)^{'}.$$

The ML estimator is now immediately seen to be

$$\begin{aligned}
\operatorname{vec}\ \hat{\Pi} &= \left[\left(I \otimes X^{'}\right)\left(\Sigma^{-1} \otimes I\right)(I \otimes X)\right]^{-1}\left(I \otimes X^{'}\right)\left(\Sigma^{-1} \otimes I\right)\operatorname{vec} y \\
&= \left(\Sigma^{-1} \otimes X'X\right)^{-1}\left(\Sigma^{-1} \otimes X^{'}\right)\operatorname{vec} y \\
&= \left(I \otimes \left(X^{'}X\right)^{-1} X^{'}\right)\operatorname{vec} y
\end{aligned}$$

which shows that the ML estimator is equivalent to OLS carried out equation by equation.

Now

$$\begin{aligned}
\left(\operatorname{vec}\left(\hat{\Pi} - \Pi\right)\right) &= \left(I_k \otimes \left(X^{'}X\right)^{-1} X^{'}\right)\operatorname{vec}\varepsilon \\
&= \left(I \otimes \left(X^{'}X\right)^{-1}\right)\left(I_k \otimes X^{'}\right)\operatorname{vec}\varepsilon,
\end{aligned}$$

where $\left(I \otimes \left(\frac{1}{T}X^{'}X\right)^{-1}\right) \xrightarrow{p} I \otimes \Gamma_p^{-1}$ and $\frac{1}{\sqrt{T}}\left(I \otimes X^{'}\right)\operatorname{vec}\ \varepsilon \xrightarrow{d} N(0, \Sigma \otimes \Gamma_p)$ with $\Gamma_p = E x_t x_t^{'}$. This follows from noting that

$$(I \otimes X)\operatorname{vec}\ \varepsilon = \begin{bmatrix} \sum x_t \varepsilon_{1t} \\ \vdots \\ \sum x_t \varepsilon_{kt} \end{bmatrix},$$

and

$$\operatorname{var}\ x_t \varepsilon_{\ell t} \varepsilon_{js} x_s^{'} = \begin{cases} 0 & \text{if} \quad t \neq s \\ \sigma_{\ell j}\Gamma_p & \text{otherwise.} \end{cases}$$

Therefore the distribution of the parameter estimates is asymptotically

$$\sqrt{T}\operatorname{vec}\left(\hat{\Pi} - \Pi\right) \xrightarrow{d} N\left(0, \left(I \otimes \Gamma_p^{-1}\right)\left(\Sigma \otimes \Gamma_p\right)\left(I \otimes \Gamma_p^{-1}\right)\right) = N\left(0, \Sigma \otimes \Gamma_p^{-1}\right).$$

If we have blockwise restrictions as in the case of Granger-non 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 ε_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

$$y_t = \sum_{j=0}^{\infty} \Psi_j R^{-1} R\varepsilon_{t-j} = \sum C_j \eta_{t-j}.$$

The forecast error of an h-step ahead forecast now is

$$\operatorname{var}(y_{t+h} - \hat{y}_{t+h}) = \Sigma + \sum_{j=1}^{h-1} C_j C_j^{'} = \Sigma + \sum_{j=1}^{h-1} \Psi_j R^{-1} R^{-1\prime} \Psi_j^{'}.$$

The coefficients C_j can be obtained from

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

where

$$J' = \underbrace{[I_k, 0...0]}_{k\times kp}, \qquad A = \begin{bmatrix} \Pi_1 & \cdots & \Pi_{p-1} & \Pi_p \\ I & & & 0 \\ & \ddots & & \vdots \\ & & I & 0 \end{bmatrix}.$$

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

$$r_{l,i}^2 + \sum_{j=1}^{h-1} c_{li,j}^2,$$

where $r_{li} = \left[R^{-1}\right]_{li}$ and $c_{li,j} = [C_j]_{li}$. 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

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

where R_i^{-1} is the ith 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

$$\frac{r_{li}^2 + \sum_{j=1}^{h-1} c_{li,j}^2}{\operatorname{var}(y_{t+h} - \hat{y}_{t+h})_l}$$

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

If one is interested in innovations in the original variables rather than the orthogonalized innovations η_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 ε_t with the first element of η_t. Since the ordering of the vectors y_t is arbitrary this identification scheme applies to all elements of ε_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 ε_{it} onto the variable $y_{l,t+h}$. Using the MA(∞) representation for y_{t+h} we find

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

so the impact of ε_t onto y_{t+h} is $\Psi_h \varepsilon_t$. If we are interested in a unit variance shock of ε_{it} then we need to take into account the fact that ε_{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 η_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 ε_{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 ε_{1t} onto y_{t+h} is therefore

$$\left[\Psi_h R^{-1}\right]_{.,1} \qquad \text{first column of } \Psi_h R^{-1}$$

The impact onto variable l is then

$$\left[\Psi_n R^{-1}\right]_{l,i}.$$

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*

$$\begin{aligned} \mathcal{M}_t^1 &= \overline{sp}\{y_{1s}, s \leq t\} \\ \mathcal{M}_t^2 &= \overline{sp}\{y_{2s}, s \leq t\} \end{aligned}$$

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

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

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

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

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

$$\operatorname{var}\left(\varepsilon_t^1\right) = \operatorname{var}\left(\varepsilon_t^2\right)$$

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

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

To see this note that by Granger causality

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

which implies that

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

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

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

It has to be emphasized that this notion of causality is strongly related to the notion of sequentiality 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*

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

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_t^1 does not Granger cause y_t^2. Then

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

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

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

By stationarity this is equivalent to

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

The reverse implication follows from the fact that by (6.1) $y_{2t} - P_{\mathcal{M}_{t-1}^2}(y_{2t}) \perp \mathcal{M}_{t-1}^1 \cup \mathcal{M}_{t-1}^2$ 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 ε_{1t} and ε_{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

$$\begin{aligned} P_{\mathcal{M}_t^1 \cup \mathcal{M}_{t-1}^2}(y_{2t}) &= (\Phi_{22}(L) - I)\, y_{2t} + \Phi_{21}(L) y_{1t} \\ &= P_{\mathcal{M}_{t-1}^2}(y_{2t}) \end{aligned}$$

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

$$\Phi_{12}(L)\Psi_{11}(L) + \Phi_{22}(L)\Psi_{21}(L) = 0$$

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

$$P_{\mathcal{M}_t^2}(y_{1t}) = \Psi_{12}(L)\Phi_{22}(L) y_{2t}$$

since $\Psi_{11}(L)\varepsilon_{1t}$ is orthogonal to $\mathcal{M}_t^2$. It now follows immediately that

$$\begin{aligned} & \operatorname{cov}\left(y_{2t+j}, y_{1t} - \Psi_{12}(L)\Phi_{22}(L) y_{2t}\right) \\ = \; & \operatorname{cov}\left(y_{2,t+j}, \Psi_{11}(L)\varepsilon_{1t}\right) = 0 \quad \forall\, j > 0 \end{aligned}$$

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 $\hat{\Pi}_{21,1}...\hat{\Pi}_{21,p}$ in

$$y_{2t} = \hat{c}_2 + \hat{\Pi}_{21,1} y_{1t-1} + ... + \hat{\Pi}_{21,p} y_{1,t-p} + \hat{\Pi}_{22,1} y_{2t-1} ... \hat{\Pi}_{22,p} y_{2t-p} \qquad ((*))$$

are jointly significantly different from zero. For a bivariate system this can be carried out by a standard F-test. Calculate the unrestricted residuals $RSS_1 = \sum \hat{\varepsilon}_t^2$ and residuals from the restricted regression

$$y_{2t} = \tilde{c}_2 + \tilde{\Pi}_{22,1} y_{2,t-1} + ... + \tilde{\Pi}_{22,p} y_{2t-p}.$$

as $RSS_0 = \sum \tilde{\varepsilon}_t^2$. Then under normality

$$\frac{RSS_0 - RSS_1}{RSS_1}(T - 2p - 1)/p \sim F(p, T - 2p - 1)$$

An asymptotically equivalent test is $T(RSS_0 - RSS_1)/RSS_1 \xrightarrow{d} \chi_p^2$ 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

$$\sum_{s=0}^{\infty} A_s y_{t-s} = \sum_{s=0}^{\infty} B_s \varepsilon_{t-s} \tag{6.2}$$

We also assume that Y_t has an equivalent VAR(∞) representation

$$y_t = \sum \Pi_s y_{t-s} + u_t \tag{6.3}$$

If the number of elements in ε_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 ε_t in terms of lagged y_t then we can write

$$B(L)^{-1} A(L) y_t = B_0 \varepsilon_t \tag{6.4}$$

with

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

where $B(L)^{-1}$ has a polynomial expansion $\sum_0^\infty C_i L^i$ and

$$B(L)^{-1} A(L) = B(L)^{-1} \left[A_0 + (A(L) - A_0)\right].$$

Since $B(L)^{-1}$ satisfies $\left(I + \sum_{i=1}^\infty B_i B_0^{-1} L^i\right)\left(\sum_0^\infty C_i L^i\right) = I$ 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 $B(L)^{-1}(A(L) - A_0) = \sum_{s=1}^\infty \tilde{C}_s L^s$. Substitution from (6.3) then gives

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

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 ε_t by

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

Note that Π_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

$$\widehat{\Sigma} = \frac{1}{T} \sum \hat{u}_t \hat{u}_t^{'}$$

If we now impose the restriction that the policy disturbances ε_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 Ω 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

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

with $c_{11}(1) = 0$ such that ε_{st} has no long-run effect on ΔY_t. Also assume $E\varepsilon_t\varepsilon_t^{'} = I$.

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

$$\begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} = \begin{bmatrix} c_{11}(0) & c_{12}(0) \\ c_{21}(0) & c_{22}(0) \end{bmatrix} \begin{bmatrix} \varepsilon_{st} \\ \varepsilon_{dt} \end{bmatrix}$$

The goal is to estimate the structural residuals ε_t which can be done if we know the coefficients of the matrix $C_0 = C(0)$. From $E\eta_t\eta_t^{'} = \Sigma$ we have

$$\Sigma = C_0 C_0'$$

i.e.,

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

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

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

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

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