Multivariate Time Series Models

Outline:

- Stationary multivariate time series
- Vector autoregressive models
- Estimation and testing of VAR models
- Impulse response functions
- Granger Causality.
- Structural VAR
- Forecasting VAR models

Multivariate Time Series Models

- Multivariate time series models allow for investigation of dynamic relationships between a set of variables, without imposing endogeneity or exogeneity restrictions.
- In the univariate case the Autoregressive Moving Average (ARMA) Model was introduced.
- The generalisation of this type of model to the multivariate context is called Vector Autoregressive Moving Average (VARMA) Model.

Multivariate Time Series Models

- A special case of the VARMA models is the Vector autoregression (VAR) model. The latter is an econometric model used to capture the evolution and the interdependencies between multiple time series. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model.
- Based on this feature, Christopher Sims in 1980 advocated the use of VAR models as a theory-free method to estimate macroeconomic relations. For this reason Sims was awarded half of the *Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2011*.
- In this module only the VAR model will be described.

Stationary multivariate time series

Let $X_t = (X_{1,t}, ..., X_{k,t})'$ be a k dimensional vector time series. We can define the following quantities:

- Mean: $\mu_t = \mathrm{E}(X_t)$ [$k \times 1$ vector]. That is

$$\mu_t = \begin{bmatrix} \mathrm{E}[X_{1,t}] \\ \vdots \\ \mathrm{E}[X_{k,t}] \end{bmatrix}$$

- Variance matrix $\Gamma_{t,t} = \mathrm{E}[(X_t - \mu_t)(X_t - \mu_t)']$ [$k \times k$ matrix]. That is

$$\Gamma_{t,t} = \begin{bmatrix} \mathrm{var}(X_{1,t}) & \mathrm{cov}(X_{1,t}, X_{2,t}) & \cdots & \mathrm{cov}(X_{1,t}, X_{k,t}) \\ \mathrm{cov}(X_{1,t}, X_{2t}) & \mathrm{var}(X_{2,t}) & \cdots & \mathrm{cov}(X_{2,t}, X_{k,t}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_{1,t}, X_{k,t}) & \mathrm{cov}(X_{2,t}, X_{k,t}) & \cdots & \mathrm{var}(X_{k,t}) \end{bmatrix}$$

Stationary multivariate time series

- Autocovariance matrix $\Gamma_{t,t-\ell} = \mathrm{E}[(X_t - \mu_t)(X_{t-\ell} - \mu_{t-\ell})']$ $[k \times k$ matrix]. That is

$$\Gamma_{t,t-\ell} = \begin{bmatrix} \mathrm{cov}(X_{1t}, X_{1,t-\ell}) & \mathrm{cov}(X_{1t}, X_{2,t-\ell}) & \cdots & \mathrm{cov}(X_{1t}, X_{k,t-\ell}) \\ \mathrm{cov}(X_{2,t}, X_{1,t-\ell}) & \mathrm{cov}(X_{2,t}, X_{2,t-\ell}) & \cdots & \mathrm{cov}(X_{2,t}, X_{k,t-\ell}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_{k,t}, X_{1,t-\ell}) & \mathrm{cov}(X_{k,t}, X_{2,t-\ell}) & \cdots & \mathrm{cov}(X_{k,t}, X_{k,t-\ell}) \end{bmatrix}$$

(note that it is not symmetric)

Definition

X_t is *weakly stationary* if for all t and ℓ :

$$\begin{aligned} \mu_t &= \mu, \\ \Gamma_{t,t-\ell} &= \Gamma_\ell = \Gamma'_{-\ell} \end{aligned}$$

Stationary multivariate time series

Remarks:

- Γ_0 is symmetric positive definite matrix.
- The diagonal elements of Γ_ℓ are the usual (univariate) autocovariances:

$$\Gamma_{ii}(\ell) = \mathrm{E}[(X_{it} - \mu_i)(X_{i,t-\ell} - \mu_i)],$$

  where $\mu_i = \mathrm{E}(X_{it})$.
- The off-diagonal elements of $\Gamma_\ell$are the cross-autocovariance, eg.

$$\begin{aligned} \Gamma_{ij}(\ell) &= \mathrm{E}[(X_{it} - \mu_i)(X_{j,t-\ell} - \mu_j)] \\ &= \mathrm{E}[(X_{jt} - \mu_j)(X_{i,t+\ell} - \mu_i)] \\ &= \Gamma_{ji}(-\ell). \end{aligned}$$

Example: $k = 2$. The mean vector is

$$\mu_t = \begin{bmatrix} \mathrm{E}[X_{1,t}] \\ \mathrm{E}[X_{2,t}] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

The Variance matrix is

$$\begin{aligned} \Gamma_0 &= \begin{bmatrix} \mathrm{var}(X_{1,t}) & \mathrm{cov}(X_{1,t}, X_{2,t}) \\ \mathrm{cov}(X_{1,t}, X_{2t}) & \mathrm{var}(X_{2,t}) \end{bmatrix} \\ &= \begin{bmatrix} \Gamma_{11}(0) & \Gamma_{12}(0) \\ \Gamma_{12}(0) & \Gamma_{22}(0) \end{bmatrix}. \end{aligned}$$

The autocovariance of order ℓ is

$$\begin{aligned}
\Gamma_\ell &= \begin{bmatrix} \Gamma_{11}(\ell) & \Gamma_{12}(\ell) \\ \Gamma_{21}(\ell) & \Gamma_{22}(\ell) \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{cov}(X_{1t}, X_{1,t-\ell}) & \mathrm{cov}(X_{1t}, X_{2,t-\ell}) \\ \mathrm{cov}(X_{2,t}, X_{1,t-\ell}) & \mathrm{cov}(X_{2,t}, X_{2,t-\ell}) \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{cov}(X_{1t}, X_{1,t+\ell}) & \mathrm{cov}(X_{2,t}, X_{1,t+\ell}) \\ \mathrm{cov}(X_{1,t}, X_{2,t+\ell}) & \mathrm{cov}(X_{2,t}, X_{2,t+\ell}) \end{bmatrix}
\end{aligned}$$

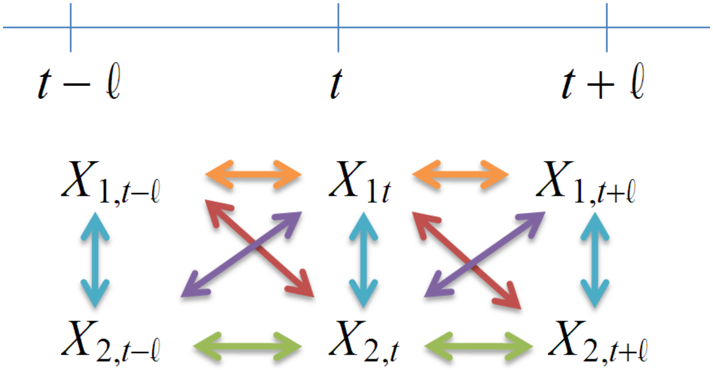
where the last line follows from stationarity.
Note that

$$\begin{aligned}
\Gamma_{-\ell} &= \begin{bmatrix} \Gamma_{11}(-\ell) & \Gamma_{12}(-\ell) \\ \Gamma_{21}(-\ell) & \Gamma_{22}(-\ell) \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{cov}(X_{1t}, X_{1,t+\ell}) & \mathrm{cov}(X_{1t}, X_{2,t+\ell}) \\ \mathrm{cov}(X_{2,t}, X_{1,t+\ell}) & \mathrm{cov}(X_{2,t}, X_{2,t+\ell}) \end{bmatrix}
\end{aligned}$$

and consequently $\Gamma_\ell = \Gamma'_{-\ell}$.

Stationary multivariate time series

Stationarity:



Arrows of the same colour mean that the covariances are identical.

Stationary multivariate time series

- We define the cross correlations as

$$\rho_{ij}(\ell) = Corr[X_{it}, X_{j,t-\ell}] = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}.$$

We can collect these cross-correlations in the cross-correlation matrix

$$\begin{aligned} \rho_\ell &= D^{-1}\Gamma_\ell D^{-1}, \\ D &= \text{diag}\{\sqrt{\Gamma_{11}(0)}, ..., \sqrt{\Gamma_{kk}(0)}\} \end{aligned}$$

Stationarity implies that $\rho_\ell = \rho'_{-\ell}$.

Diagonal elements of ρ_ℓ define the ACF of X_{kt}

Definition

Multivariate White Noise ε_t is a stationary process with

1. $E(\varepsilon_t) = 0$ (a $k \times 1$ vector of zeros)
2. $var(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \Omega$ (a constant variance-covariance matrix)
3. $cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s') = 0$ for $s \neq t$ (uncorrelated).

Stationary multivariate time series

- Sample cross-covariance:

$$\hat{\Gamma}_\ell = \frac{1}{T}\sum_{t=\ell+1}^{T}(X_t - \bar{X})(X_{t-\ell} - \bar{X})', \ell \geq 0$$

  where $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$.
- Cross-correlation matrices

$$\hat{\rho}_\ell = \hat{D}^{-1}\hat{\Gamma}_\ell\hat{D}^{-1},$$

  where $\hat{D} = \text{diag}\{\sqrt{\hat{\Gamma}_{11}(0)}, ..., \sqrt{\hat{\Gamma}_{kk}(0)}\}$.
- Under the Assumption of multivariate *i.i.d.*(hence $\rho_\ell = 0$ for all $\ell \neq 0$) we have

$$\sqrt{T}\hat{\rho}_{ij}(\ell) \xrightarrow{D} N(0,1).$$

- Multivariate $Q-$statistic

$$Q_k(m) = T^2\sum_{\ell=1}^{m}\frac{1}{T-\ell}tr(\hat{\Gamma}'_\ell\hat{\Gamma}_0^{-1}\hat{\Gamma}_\ell\hat{\Gamma}_0^{-1}).$$

  Under H_0 : X_t is i.i.d. $Q_k(m) \xrightarrow{D} \chi^2(k^2m)$

Multivariate Wold decomposition theorem

Theorem

(Multivariate Wold decomposition theorem) If the k-variate X_t time series process is weakly stationary, then it has the representation

$$X_t = \sum_{s=0}^{\infty} \Lambda_s \varepsilon_{t-s} + W_t$$

where the $k \times k$ matrices Λ_s are such that $\Lambda_0 = I_k$, $\sum_{s=1}^{\infty} \Lambda_s \Lambda_s'$ converges, the process ε_t is a k variate white noise process and $W_t \in R^k$ is a linear deterministic process, that this there exists a k vector c_0 and $k \times k$ matrices C_s such that $W_t = c_0 + \sum_{s=0}^{\infty} C_s W_{t-s}$, and $E[\varepsilon_t W_{t-m}] = 0$, for $m = 0, \pm 1, \pm 2, \ldots$

Remarks: Usually we ignore the determinist process W_t (or assume that it is a constant) and try to approximate $\sum_{s=0}^{\infty} \Lambda_s \varepsilon_{t-s}$.

Multivariate Polynomials in L

- Similarly to the univariate case we can define a (finite or infinite order) multivariate polynomial in L or a filter according to:

$$A(L) = A_0 + A_1 L + A_2 L^2 + \dots$$

where the matrices $A_j, j = 0, 1, ...$ are not necessarily square.

Inversion of Polynomials in L

Let $H(L)$ be a finite order polynomial in L. $H(L) = I - \sum_{i=1}^{p} H_i L^i$.We define its inverse as $H(L)^{-1}$ to be the multivariate polynomial in L if

$$H(L)^{-1} H(L) = I$$

- $H(L)^{-1}$ will correspond to a series of the form $\sum_{i=0}^{\infty} B_i L^i$.

Example: Suppose $H(L) = I - \Pi L$.Note that

$$(1 + \Pi L + \Pi^2 L^2 + \dots)(I - \Pi L) = I$$

so $H(L)^{-1} = \sum_{i=0}^{+\infty} \Pi^i L^i$.

- The coefficients of the infinite-order polynomial $H(L)^{-1} = \sum_{i=0}^{\infty} B_i L^i$ are absolutely summable if $\sum_{i=0}^{\infty} |b_{k\ell i}| < \infty$ for all k, ℓ, where $b_{k\ell i}$ is the element (k, ℓ) of the matrix B_i.
    - As in the univariate case the conditions that ensure that an inverse has absolutely summable coefficients play a crucial role in establishing necessary conditions for a multivariate time series model to be stationary.
    - Necessary and sufficient conditions for an inverse to meet the absolute summability condition:
        - $H(L)$ has an *absolutely summable inverse* if the roots of the characteristic equation
          $$\left| I\lambda^p - \sum_{\ell=1}^{p} H_\ell \lambda^{p-\ell} \right| = 0$$
          are *inside* the unit circle, where $|A|$ corresponds to the determinant of A.
        - Equivalently $H(L)$ has an *absolutely summable inverse* if all values of z satisfying
          $$\left| I - \sum_{\ell=1}^{p} H_\ell z^\ell \right| = 0$$
          are *outside* the unit circle.

Vector Autoregressive models

Let us consider $k = 3$ for simplicity. The vector autoregressive model of order 1, $VAR(1)$ is defined as

$$
\begin{aligned}
X_{1,t} &= \phi_{10} + \Phi_{11}(1)X_{1,t-1} + \Phi_{12}(1)X_{2,t-1} + \Phi_{13}(1)X_{3,t-1} + \varepsilon_{1t}, \\
X_{2,t} &= \phi_{20} + \Phi_{21}(1)X_{1,t-1} + \Phi_{22}(1)X_{2,t-1} + \Phi_{23}(1)X_{3,t-1} + \varepsilon_{2t}, \\
X_{3,t} &= \phi_{30} + \Phi_{31}(1)X_{1,t-1} + \Phi_{32}(1)X_{2,t-1} + \Phi_{33}(1)X_{3,t-1} + \varepsilon_{3t}
\end{aligned}
$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ is a multivariate white noise with $var(\varepsilon_t) = \Omega$;

This model can be written in matrix form as

$$X_t = \phi_0 + \Phi_1 X_{t-1} + \varepsilon_t$$

where

$$X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{bmatrix}, \phi_0 = \begin{bmatrix} \phi_{10} \\ \phi_{20} \\ \phi_{20} \end{bmatrix}$$

and

$$\Phi_1 = \begin{bmatrix} \Phi_{11}(1) & \Phi_{12}(1) & \Phi_{13}(1) \\ \Phi_{21}(1) & \Phi_{22}(1) & \Phi_{23}(1) \\ \Phi_{31}(1) & \Phi_{32}(1) & \Phi_{33}(1) \end{bmatrix}$$

- **Example:** $VAR(1)$ process:

$$\begin{bmatrix} GNP_t \\ M2_t \\ IR_t \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0 & 0.4 & 0.1 \\ 0.9 & 0 & 0.8 \end{bmatrix} \begin{bmatrix} GNP_{t-1} \\ M2_{t-1} \\ IR_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

where GNP_t is the Gross National Product, $M2_t$ is money supply, and IR_t is interest rate.

Vector Autoregressive models

For any k the Vector autoregressive model of order p - $VAR(p)$ model - is a system of regression equations

$$\begin{aligned} X_{1,t} &= \phi_{10} + \textstyle\sum_{\ell=1}^{p} \sum_{j=1}^{k} \Phi_{1j}(\ell) X_{j,t-\ell} + \varepsilon_{1t} \\ X_{2,t} &= \phi_{20} + \textstyle\sum_{\ell=1}^{p} \sum_{j=1}^{k} \Phi_{2j}(\ell) X_{j,t-\ell} + \varepsilon_{2t} \\ &\vdots \\ X_{k,t} &= \phi_{k0} + \textstyle\sum_{\ell=1}^{p} \sum_{j=1}^{k} \Phi_{kj}(\ell) X_{j,t-\ell} + \varepsilon_{kt} \end{aligned}$$

or in matrix notation

$$X_t = \phi_0 + \textstyle\sum_{\ell=1}^{p} \Phi_\ell X_{t-\ell} + \varepsilon_t,$$

where

- $X_t = (X_{1,t}, X_{2,t}, ..., X_{k,t})'$.
- $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{kt})'$ is a multivariate white noise with $var(\varepsilon_t) = \Omega$;
- $\phi_0 = (\phi_{10}, \phi_{20}, ..., \phi_{k0})'$ is a vector of intercepts;
- $\Phi_\ell = [\Phi_{ij}(\ell)]$ are $k \times k$ coefficient matrices.

Stationarity of VAR(1)

Consider the $VAR(1)$ process

$$X_t = \phi_0 + \Phi_1 X_{t-1} + \varepsilon_t$$

Stationary condition: All eigenvalues λ^* of Φ,i.e. all roots of $|\lambda I_k - \Phi_1| = 0$, should lie *inside* the unit circle. $|.|$ is the determinant of the , matrix.
Equivalent condition: roots z^* of the characteristic equation $|I_k - \Phi_1 z| = 0$ should lie *outside* the unit circle ($z = 1/\lambda$).
Remark: $\sum_{j=0}^{\infty} \Phi_1^j$ is only convergent under the stationary condition.

Stationarity of VAR(1)

Using the Lag operator notation we can write the model as

$$\Phi(L)X_t = \phi_0 + \varepsilon_t$$

where $\Phi(L) = I_k - \Phi_1 L$ is a matrix lag polynomial.
Under this stationarity condition $\Phi(L)$ has an *absolutely summable inverse:*

$$\begin{aligned}\Phi(L)^{-1} &= (I_k - \Phi_1 L)^{-1} \\ &= \textstyle\sum_{j=0}^{\infty} \Phi_1^j L^j.\end{aligned}$$

Thus

$$\begin{aligned}X_t &= \Phi(L)^{-1}[\phi_0 + \varepsilon_t] \\ &= \textstyle\sum_{j=0}^{\infty} \Phi_1^j L^j [\phi_0 + \varepsilon_t] \\ &= \mu + \textstyle\sum_{j=0}^{\infty} \Phi_1^j \varepsilon_{t-j},\end{aligned}$$

where $\mu = (I_k - \Phi_1)^{-1}\phi_0$.

Stationarity of VAR(1)

Under stationarity condition

$$\begin{aligned}
\mu &= \mathrm{E}(X_t) = \textstyle\sum_{j=0}^{\infty} \Phi_1^j \phi_0 = (I_k - \Phi_1)^{-1} \phi_0, \\
\Gamma_0 &= \mathrm{var}(X_t) = \textstyle\sum_{j=0}^{\infty} \Phi_1^j \Omega (\Phi_1^j)', \\
\Gamma_\ell &= \mathrm{cov}(X_t, X_{t-\ell}) = \Phi_1^\ell \Gamma_0, \\
\rho_\ell &= corr(X_t, X_{t-\ell}) = A^\ell \rho_0,
\end{aligned}$$

where $\rho_0 = D^{-1}\Gamma_0 D^{-1}$, $A = D^{-1}\Phi D$ where $D = \mathrm{diag}\{\sqrt{\Gamma_{11}(0)}, ..., \sqrt{\Gamma_{kk}(0)}\}$.

- **Example:** $VAR(1)$ process:

$$\begin{bmatrix} GNP_t \\ M2_t \\ IR_t \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0 & 0.4 & 0.1 \\ 0.9 & 0 & 0.8 \end{bmatrix} \begin{bmatrix} GNP_{t-1} \\ M2_{t-1} \\ IR_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix},$$

where GNP_t is the Gross National Product, $M2_t$ is money supply, and IR_t is interest rate.

$$\begin{aligned} \left| \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0 & 0.4 & 0.1 \\ 0.9 & 0 & 0.8 \end{bmatrix} \right| &= 0 \\ \lambda^3 - 1.9\lambda^2 + 1.16\lambda - 0.233 &= 0 \end{aligned}$$

Roots:

$\lambda_1 = 0.89395,\ \lambda_2 = 0.50303 + 0.087213i,\ \lambda_3 = 0.50303 - 0.087213i.$

Thus

$$|\lambda_1| = 0.89395, |\lambda_2| = |\lambda_3| = 0.51053$$

Hence the process is stationary

Stationarity of VAR(p)

- Consider now general $VAR(p)$ model:

$$X_t = \phi_0 + \sum_{\ell=1}^{p} \Phi_\ell X_{t-\ell} + \varepsilon_t,$$

or

$$\begin{aligned} \Phi(L)X_t &= \phi_0 + \varepsilon_t, \\ \Phi(L) &= I_k - \sum_{\ell=1}^{p} \Phi_\ell L^\ell \end{aligned}$$

- A $VAR(p)$ process is stationary if the roots of

$$\left| I_k \lambda^p - \sum_{\ell=1}^{p} \Phi_\ell \lambda^{p-\ell} \right| = 0$$

are *inside* the unit circle.

- Equivalently the $VAR(p)$ process is stationary if all values of z satisfying

$$\begin{aligned} \left| I_k - \sum_{\ell=1}^{p} \Phi_\ell z^\ell \right| &= 0 \\ |\Phi(z)| &= 0 \end{aligned}$$

are *outside* the unit circle.

VMA representation of a VAR(p) process

- If all roots of $|\Phi(z)| = 0$ lie outside the unit circle, stationarity implies that $\Phi(L)$ has an absolutely summable inverses and the $VAR(p)$ process has the Vector Moving Average representation (VMA):

$$
\begin{aligned}
X_t &= \Phi(L)^{-1}(\phi_0 + \varepsilon_t) \\
&= \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}
\end{aligned}
$$

  where $\Phi(L)^{-1} = \Psi(L) = \sum_{j=0}^{\infty} \Psi_j L^j$ and $\sum_{j=0}^{\infty} \Psi_j$ and $\sum_{j=0}^{\infty} \Psi_j \Psi_j'$ converge.

- **Example:** Recall that if $p = 1$ and $\phi_0 = 0$ then $X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}$ where

$$
\Psi_0 = I_k, \Psi_j = \Phi_1^j, j \geq 1.
$$

- For any p we have

$$
\begin{aligned}
\mu &= E(X_t) = \Psi(1)\phi_0, \\
\Gamma_\ell &= cov(X_t, X_{t-\ell}) = \sum_{j=0}^{\infty} \Psi_{j+\ell} \Omega \Psi_j', \ell \geq 0
\end{aligned}
$$

Impulse response functions

Consider the MA representation of the $VAR(p)$ process

$$X_t = c + \sum_{r=0}^{\infty} \Psi_r \varepsilon_{t-r}$$

where $c = \sum_{r=0}^{\infty} \Psi_r \phi_0$.
Notice that

$$\frac{\partial X_{t+\ell}}{\partial \varepsilon_t'} = \Psi_\ell \Rightarrow \frac{\partial X_{i,t+\ell}}{\partial \varepsilon_{jt}} = \Psi_{ij}(\ell)$$

where $\Psi_{ij}(\ell)$ is the element in row i and column j of Ψ_ℓ.

- A plot of $\Psi_{ij}(\ell)$ against ℓ is the *impulse response function*.
- To see what is going on let us consider the case that $k = 2$, that is $X_t = (X_{1t}, X_{2t})'$ therefore the model becomes

$$\begin{aligned} X_{1t} &= c_1 + \sum_{r=0}^{\infty} \Psi_{11}(r)\varepsilon_{1,t-r} + \sum_{r=0}^{\infty} \Psi_{12}(r)\varepsilon_{2,t-r}, \\ X_{2t} &= c_2 + \sum_{r=0}^{\infty} \Psi_{21}(r)\varepsilon_{1,t-r} + \sum_{r=0}^{\infty} \Psi_{22}(r)\varepsilon_{2,t-r} \end{aligned}$$

where $c = (c_1, c_2)'$.

Impulse response functions

- In period $t + \ell$ we have

$$\begin{aligned} X_{1t+\ell} &= c_1 + \sum_{r=0}^{\infty} \Psi_{11}(r)\varepsilon_{1,t+\ell-r} + \sum_{r=0}^{\infty} \Psi_{12}(r)\varepsilon_{2,t+\ell-r}, \\ X_{2t+\ell} &= c_2 + \sum_{r=0}^{\infty} \Psi_{21}(r)\varepsilon_{1,t+\ell-r} + \sum_{r=0}^{\infty} \Psi_{22}(r)\varepsilon_{2,t+\ell-r}. \end{aligned}$$

- Hence

$$\frac{\partial X_{1,t+\ell}}{\partial \varepsilon_{1,t}} = \Psi_{11}(\ell), \quad \frac{\partial X_{1,t+\ell}}{\partial \varepsilon_{2,t}} = \Psi_{12}(\ell),$$

$$\frac{\partial X_{2,t+\ell}}{\partial \varepsilon_{1,t}} = \Psi_{21}(\ell), \quad \frac{\partial X_{2,t+\ell}}{\partial \varepsilon_{2,t}} = \Psi_{22}(\ell).$$

- The *impulse response function* describes the response of $X_{i,t+\ell}$ to a one-time unit change in ε_{jt}. where the units are those that ε_{jt} is measured.
- Usually we multiply $\Psi_{ij}(\ell)$ by the standard deviation of ε_{jt} so we obtain the response of $X_{i,t+\ell}$ to a one-time change in ε_{jt} of $var(\varepsilon_{jt})^{1/2}$ units.

Estimation

Let us assume that

$$X_t = \phi_0 + \sum_{\ell=1}^{p} \Phi_\ell X_{t-\ell} + \varepsilon_t, t = 1, ..., T$$

where $\varepsilon_t \sim i.i.d\ N(0, \Omega)$.

- We shall condition on the p first observations and derive the conditional likelihood function for $X_1, ..., X_T$.
- Let θ denote the vector of unknown parameters: ϕ_0, Φ_ℓ $(\ell = 1, ..., p)$ and Ω. The dimension of θ is $k + pk^2 + k(k+1)/2$.

Let:

- $Z_t = (1, X'_{t-1}, ..., X'_{t-p})'$ $((kp+1) \times 1)$
- $B' = [\phi_0, \Phi_1, ..., \Phi_p]$, $(k \times (kp+1))$.
- $Z_t^* = (X'_t, ..., X'_{t-p})', t \geq p+1$

Then the $VAR(p)$ model can be written more compactly as

$$X_t = B'Z_t + \varepsilon_t, t = p+1, .., T$$

Conditioning on the past values we obtain

$$X_t | Z_t^* \sim N(B'Z_t, \Omega)$$

Estimation

- Hence the conditional density of $X_t|Z_t^*$ is

$$f_{X_t|Z_t^*}(x_t|z_t^*,\theta) = (2\pi)^{-k/2}|\Omega|^{-1/2}\exp\{-(1/2) \times (x_t - B'z_t)'\Omega^{-1}(x_t - B'z_t)\}.$$

- Recall that the formula of the conditional log-likelihood is given by

$$\begin{aligned} \log \mathcal{L}(\theta) &= \sum_{t=p+1}^{T} \log f_{X_t|Z_t^*}(x_t|z_t^*,\theta) \\ &= -\frac{kT^*}{2}\log(2\pi) - \frac{T^*}{2}\log|\Omega| \\ &\quad -(1/2)\sum_{t=p+1}^{T}(x_t - B'z_t)'\Omega^{-1}(x_t - B'z_t) \end{aligned}$$

with $T^* = T - p$.

- There is a closed form solution for the conditional MLE:

$$\hat{B} = [\sum_{t=p+1}^{T} x_t z_t'][\sum_{t=p+1}^{T} z_t z_t']^{-1}$$

$$\hat{\Omega} = \frac{1}{T}\sum_{t=p+1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t', \text{ where } \hat{\varepsilon}_t = x_t - \hat{B}'z_t$$

- Note that the j row of $\hat{B}$ is given by

$$\hat{b}_j = [\sum_{t=p+1}^{T} z_t z_t']^{-1} [\sum_{t=p+1}^{T} x_{j,t} z_t']$$

(where $j = 1, .., k$)

- Conclusion the conditional MLE of $\hat{B}$ is obtained by **applying ordinary least squares separately to each equation**. One can show that for $\hat{b} = vec(\hat{B}) = (\hat{b}_1', ..., \hat{b}_j')'$
- **Remark:** The vec operator applied to a matrix A ($vec(A)$) creates a column vector from a matrix A by stacking the column vectors of A.
- One can show that

$$\sqrt{T}(\hat{b} - b) \xrightarrow{D} N(0, \Omega \otimes E(z_t z_t'))$$

where $\otimes$ denotes the Kronecker product.

Testing

The Likelihood ratio to test h restrictions $H_0 : r(\theta) = 0$ has the form

$$\mathcal{LR} = T(\log|\hat{\Omega}_r| - \log|\hat{\Omega}|)$$

where $\hat{\Omega}_r$ is the restricted MLE.
One can show that

$$\mathcal{LR} \xrightarrow{D} \chi^2(h)$$

where h is the dimension of $r(\theta)$.

Bivariate Granger causality

A scalar variable X *Granger-causes* another scalar variable Y if Y can be better predicted using the histories of both X and Y than it can using the history of Y alone.
Formally:

Definition

X *fails to Granger cause* Y if

$$MSE[\hat{E}(Y_{t+s}|Y_t, Y_{t-1}, ...)] = MSE[\hat{E}(Y_{t+s}|X_t, X_{t-1}, ..., Y_t, Y_{t-1}, ...)]$$

where MSE is the mean square error of prediction:
$MSE(\hat{E}(.)) = E[(Y_{t+s} - \hat{E}(.))^2]$.

- In a VAR model with $k = 2$ with $Z_t = (X_t, Y_t)'$:

$$Z_t = \phi_0 + \sum_{\ell=1}^{p} \Phi_\ell Z_{t-\ell} + \varepsilon_t,$$

  - If $\Phi_{12}(\ell) = 0$ for $\ell = 1, ..., p$, Y does not Granger Cause X.
  - If $\Phi_{21}(\ell) = 0$ for $\ell = 1, ..., p$, X does not Granger Cause Y.

Example: Consider the following $VAR(2)$ process

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.7 & 0 \\ 0.9 & 0.8 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},$$

where $(\varepsilon_t^1, \varepsilon_t^2)'$ is a vector of white noise processes.

- Y is not Granger Causal to X.
- X Granger causes Y.

Simple econometric tests for bivariate Granger Causality

- The tests based on the VAR methodology can be used to test Granger Causality.
- However, there is a simpler alternative way to test this based on a multivariate regression model:
  - Let
$$y_t = c + \sum_{i=1}^{p} [\alpha_i x_{t-i} + \beta_i y_{t-i}] + u_t$$
    where for $z_t = (x_{t-1}, ..., x_{t-p}, y_{t-1}, ..., y_{t-p})$ we have:
    - $E(u_t|z_t) = 0$ the regressors are *contemporaneously exogenous*
    - $var(u_t|z_t) = \sigma^2$ the regressors are *contemporaneously homoskedastic*,
    - $cov(u_t, u_s|z_t, z_s) = 0, s \neq t$ (*no autocorrelation*).
  - x fails to Granger cause y if
$$H_0 : \alpha_i = 0, \text{ for } i = 1, ..., p$$

Simple Econometric tests for bivariate Granger Causality

We can test this hypothesis in the following way under the above assumptions:

- Let RSS_1 be the residual sum of squares of the regression

$$y_t = c + \sum_{i=1}^{p} [\alpha_i x_{t-i} + \beta_i y_{t-i}] + u_t, t = 1, .., T$$

- Let RSS_0 be the residual sum of squares of the regression

$$y_t = c + \sum_{i=1}^{p} \beta_i y_{t-i} + u_t, t = 1, ..., T,$$

- Under H_0

$$S = \frac{T(RSS_0 - RSS_1)}{RSS_1} \xrightarrow{D} \chi^2(p)$$

  We can use this statistic to test H_0. Let c_α the $100 \times \alpha\%$ critical value. We reject H_0 if the actual value of S is bigger than c_α.

Specification testing in VAR models

The residuals $\hat{\varepsilon}_{it}$ can be used for usual (univariate) misspecification tests.
Stronger results are obtained from vector tests:

- Multivariate $Q-$statistic can be applied to residuals with asymptotic $\chi^2(k^2(m-p))$ distribution.
- One can also apply vector LM tests of serial correlation.

Lag Length selection can be based on the minimization of the information criteria:

$$
\begin{aligned}
AIC(p) &= -\frac{2}{T^*}\log\mathcal{L}(\hat{\theta}_p) + \frac{2k^2p}{T^*}, \text{ Akaike information criterion} \\
BIC(p) &= -\frac{2}{T^*}\log\mathcal{L}(\hat{\theta}_p) + \frac{k^2p\log(T^*)}{T^*}, \text{Schwarz Information criterion}
\end{aligned}
$$

where $\hat{\theta}_p$ is the conditional MLE estimator for the parameters of the $VAR(p)$ model and with $T^* = T - p$ (usual definitions).

Structural VAR

- Structural VAR (SVAR) allows contemporaneous relationships between elements of X_t :

$$B_0 X_t = c + B_1 X_{t-1} + B_2 X_{t-2} + ... + B_p X_{t-p} + U_t$$

where U_t is a multivariate white noise process with $E(U_t) = 0$ and $var(U_t) = D$ and B_i are $k \times k$ matrices $i = 0, ..., p$.

- If B_0 is invertible, then this model is equivalent to a reduced form VAR

$$X_t = \phi_0 + \sum_{\ell=1}^{p} \Phi_\ell X_{t-\ell} + \varepsilon_t,$$

where

$$\begin{aligned} \phi_0 &= B_0^{-1} c, \Phi_\ell = B_0^{-1} B_\ell, \\ \varepsilon_t &= B_0^{-1} U_t \end{aligned}$$

Thus $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \Omega = B_0^{-1} D [B_0^{-1}]'$.

- Can we derive the elements of the structural VAR uniquely from the reduced form VAR?

- Consider the number of elements in each model

SVAR		VAR	
c	k	ϕ_0	k
$B_0, ..., B_p$	$(1+p)k^2$	$\Phi_1, ..., \Phi_p$	pk^2
D	$k(k+1)/2$	Ω	$k(k+1)/2$

- The SVAR has k^2 more parameters than the VAR and so we need k^2 restrictions in order to identify the parameters of the SVAR.

Structural VAR

- Essentially a necessary condition for identification requires B_0 and D to have no more unknown elements than Ω which is $k(k+1)/2$. This condition is known the *order condition* for identification.
- Normalization restrictions: Assign the coefficient of 1 to X_{jt} in each equation (k).
- Covariance matrix restrictions: e.g. Specifying D to be diagonal (k).
- So under these restrictions B_0 **and** D have k^2 elements and Ω has $k(k+1)/2$ elements.
- We still need to impose $k(k-1)/2$ restrictions.
- A solution: *Cholesky Decomposition* - B_0 is lower triangular

$$B_0 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ b_{21} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ b_{k1} & b_{k2} & b_{k_3} & \cdots & 1 \end{bmatrix}$$

- **Remark:** The Cholesky decomposition does not have a direct economic interpretation.
- This approach is called Cholesky decomposition because it is based on a Cholesky type decomposition of a positive definite matrix: Any symmetric positive definite matrix A can be decomposed as $A = LGL'$ where G is a diagonal matrix a L is a lower triangular matrix with $1'$s in the diagonal.
- So basically we are applying this decomposition to $\Omega = LGL'$, with $G = D$ and $L = B_0^{-1}$.
- **Remark:** Other alternative is to impose some restrictions based on Economic Theory.

Impulse response functions in the structural model

Consider the MA representation of the $VAR(p)$ process

$$X_t = \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} \tag{1}$$

and recall that

$$\varepsilon_t = B_0^{-1} U_t, t = 1, ...$$

Replacing this in (1) we have

$$X_t = \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=0}^{\infty} \Psi_j B_0^{-1} U_{t-j}$$

Consequently

$$X_{t+\ell} = \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=0}^{\infty} \Psi_j B_0^{-1} U_{t+\ell-j}$$

and therefore

$$\frac{\partial X_{t+\ell}}{\partial U_t'} = \Psi_\ell B_0^{-1}$$

To simplify the notation write $A_\ell = \Psi_\ell B_0^{-1}$ and denote $A_{ij}(\ell)$ the element i,j of this matrix. Then

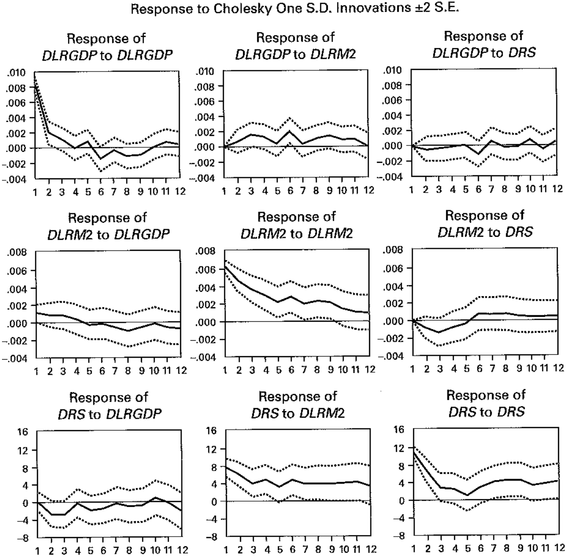
$$\frac{\partial X_{i,t+\ell}}{\partial U_{jt}} = A_{ij}(\ell)$$

Impulse response functions in the structural model

- A plot of $A_{ij}(\ell)$ against ℓ is the structural impulse response function (Enders denotes this function simply as impulse response function.)
- It describes the response of $X_{i,t+\ell}$ to a one-time unit change in U_{jt}. where the units are those that U_{jt} is measured.
- As before some researchers prefer to multiply $A_{ij}(\ell)$ by the standard deviation of u_{jt} so we obtain the response of $X_{i,t+\ell}$ to a one-time change in U_{jt} of $var(U_{jt})^{1/2}$ units.
- **Example:** Let DLRGDP- logarithmic change in real GDP for USA, DLRM2 - logarithmic change in real money supply, DRS - change in short term interest rate. Quarterly data: 1959:1-2001:1.

Impulse response functions in the structural model

Example: After estimating a VAR model we obtain:



Forecasting VAR models

- Consider a stationary $VAR(p)$ model:

$$X_t = \phi_0 + \sum_{i=1}^{p} \Phi_i X_{t-i} + \varepsilon_t.$$

- Suppose we are in period h and we want to forecast the observations in period $h + \ell$, $\ell > 0$.

Forecasting VAR models

Forecasting in stationary $VAR(p)$ models similar to univariate $AR(p)$:

- $\ell-$ step forecasts $X_h(\ell) = E_h[X_{h+\ell}], \ell > 0$ (assuming that ε_h is a martingale difference sequence: $E_h[\varepsilon_{h+\ell}] = 0$) defined recursively from

  $$X_h(\ell) = \phi_0 + \sum_{i=1}^{p} \Phi_i X_h(\ell - i)$$

  where $X_h(\ell - i) = X_{h+\ell-i}$ for $i \geq \ell$.
- From

  $$X_{h+\ell} = \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{h+\ell-j}$$
- Thus

$$\begin{aligned} X_h(\ell) &= E_h[X_{h+\ell}] \\ &= \sum_{j=0}^{\infty} \Psi_j \phi_0 + \sum_{j=\ell}^{\infty} \Psi_j \varepsilon_{h+\ell-j} \end{aligned}$$

Forecasting VAR models

- Hence we obtain the forecast error

$$\begin{aligned} e_h(\ell) &= X_{h+\ell} - X_h(\ell) \\ &= \sum_{j=0}^{\ell-1} \Psi_j \varepsilon_{h+\ell-j} \end{aligned}$$

  and as $var(\varepsilon_{h+\ell-j}) = \Omega$ and $\varepsilon_{h+\ell-j}$ is a multivariate White noise process the variance is

$$var(e_h(\ell)) = \sum_{j=0}^{\ell-1} \Psi_j \Omega \Psi_j'.$$