Vector Autoregressive Moving Average (VARMA) Models

Massimo Guidolin

February 2018

1 Foundations of Multivariate Time Series Analysis

1.1 Weak Stationarity of Multivariate Time Series

Weak Stationarity (multivariate case): Consider a N-dimensional time series $\mathbf{y_t} = [y_{1,t}, y_{2,t}, ..., y_{N,t}]'$. Formally, this is said to be weakly stationary if its first two unconditional moments are finite and constant through time, i.e.,

- $E[\mathbf{y_t}] \equiv \boldsymbol{\mu} < \infty \quad \forall t$
- $E[(\mathbf{y_t} - \boldsymbol{\mu})(\mathbf{y_t} - \boldsymbol{\mu})'] \equiv \mathbf{\Gamma_0} < \infty \quad \forall t$
- $E[(\mathbf{y_t} - \boldsymbol{\mu})(\mathbf{y_{t-h}} - \boldsymbol{\mu})'] \equiv \mathbf{\Gamma_{\boldsymbol{h}}} \quad \forall t, \forall h$

where
$\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_N]$;
$\mathbf{\Gamma_0} = N \times N$ covariance matrix where the ith diagonal element is the variance of $y_{i,t}$ and the (i,j)th element is the covariance between $y_{i,t}$ and $y_{j,t}$;
$\mathbf{\Gamma_{\boldsymbol{h}}}$ =cross-covariance matrix at lag h;
and the expectations are taken element-by-element over the joint distribution of $\mathbf{y_t}$.

1.2 Cross-Covariance and Cross-Correlation Matrices

Lag-0 correlation matrix of $\mathbf{y_t}$:

$$\boldsymbol{\rho_0} = \mathbf{D}^{-1}\mathbf{\Gamma_0}\mathbf{D}^{-1}$$

where $\mathbf{D}= N \times N$ diagonal matrix collecting on its main diagonal the standard deviations of $y_{i,t}$ for $i = 1, ..., N$ and

$$\rho_{i,j}(0) = \frac{Cov[y_{i,t}, y_{j,t}]}{\sigma_{i,t}\sigma_{j,t}}$$

- $\boldsymbol{\rho_0}$ is a symmetric matrix with unit diagonal elements because $\rho_{i,j}(0) = \rho_{j,i}(0)$, $-1 \leq \rho_{i,j} \leq 1$ and $\rho_{i,i} = 1$ for $1 \leq i$ and $j \leq N$.

Lag-h cross-covariance matrix of $\mathbf{y_t}$:

$$\boldsymbol{\Gamma_h} = E[(\mathbf{y_t} - \boldsymbol{\mu})(\mathbf{y_{t-h}} - \boldsymbol{\mu})']$$

where $\boldsymbol{\mu}$=mean vector of $\mathbf{y_t}$ and the (i,j)th element of $\boldsymbol{\Gamma_h}$=covariance between $y_{i,t}$ and $y_{j,t-h}$.

- It is time-invariant if the time-series is weakly stationary.

Lag-h cross-correlation matrix:

$$\boldsymbol{\rho_h} = \mathbf{D}^{-1}\boldsymbol{\Gamma_h}\mathbf{D}^{-1}$$

where $\mathbf{D}$ is the diagonal matrix of standard deviations of the individual series $y_{i,t}$ and

$$\rho_{i,j}(h) = \frac{Cov[y_{i,t}, y_{j,t-h}]}{\sigma_{i,t}\sigma_{j,t}}$$

- When $h > 0$, $\rho_{i,j}(h)$ measures the linear dependence of $y_{i,t}$ on $y_{j,t-h}$, while $\rho_{j,i}(h)$ measures the linear dependence of $y_{j,t}$ on $y_{i,t-h}$.
- $\rho_{i,i}(h)$ is the lag-h autocorrelation coefficient of $y_{i,t}$.
- $\rho_{j,i}(h) \neq \rho_{i,j}(h)$ for any $i \neq j$, therefore $\boldsymbol{\Gamma_h}$ and $\boldsymbol{\rho_h}$ do not need to be symmetric.
- Information summarized by the cross-correlation matrices:

Assumption on $\rho_{i,j}$	Relationship between $y_{i,t}$ and $y_{j,t}$
$\rho_{i,j}(0) \neq 0$	Contemporaneously linear
$\rho_{i,j}(h) = \rho_{j,i}(h) = 0$ *for all* $h \geq 0$	Not linearly correlated
$\rho_{i,j}(h) = 0$ *and* $\rho_{j,i}(h) = 0$ *for all* $h > 0$	Linearly uncoupled, but possibly contemporaneous and linear
$\rho_{i,j}(h) = 0$ *for all* $h > 0$, *but* $\rho_{j,i}(q) \neq 0$ *for at least some* $q > 0$	Unidirectional (linear), where $y_{i,t}$ does not depend on $y_{j,t}$, but $y_{j,t}$ linearly depends on (some) lagged values of $y_{i,t}$;
$\rho_{i,j}(h) \neq 0$ *for at least some* $h > 0$, *and* $\rho_{j,i}(q) \neq 0$ *for at least some* $q > 0$	Linear bi-directional feedback

1.3 Sample Cross-Covariance and Cross-Correlation Matrices

Sample cross-covariances matrix:

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{h}} = \frac{1}{T} \sum_{t=h+1}^{T} (\mathbf{y_t} - \bar{\mathbf{y}})(\mathbf{y_{t-h}} - \bar{\mathbf{y}}) \quad \text{with} \quad h \geq 0$$

where
$\bar{\mathbf{y}} = [\bar{y_1}, \bar{y_2}, ..., \bar{y_N}]'$
$\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{i,t}$ with $i = 1, ..., N$

Sample cross-correlation matrix:

$$\hat{\boldsymbol{\rho}}_{\boldsymbol{h}} = \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Gamma}}_{\boldsymbol{h}} \hat{\mathbf{D}}^{-1} \quad \text{with} \quad h \geq 0$$

where $\hat{\mathbf{D}}$ is the $N \times N$ diagonal matrix of the sample standard deviations of each of the component series.

1.4 Multivariate Portmanteau Tests

Multivariate version of the Ljung-Box statistic:

$$H_0 : \boldsymbol{\rho_1} = ... = \boldsymbol{\rho_m} = 0 \quad \text{vs} \quad H_1 : \boldsymbol{\rho_i} \neq 0 \quad \text{for some} \quad i \in \{1, ..., m\}$$

$$Q(m) = T^2 \sum_{h=1}^{m} \frac{1}{T-h} tr(\hat{\Gamma}_h \hat{\Gamma}_0^{-1} \hat{\Gamma}_h \hat{\Gamma}_0^{-1})$$

where
T = sample size
N = dimension of $\mathbf{y_t}$
m = maximum lag length we want to test
$tr(\mathbf{A})$ = trace of some matrix $\mathbf{A}$, defined as the sum of the diagonal elements of $\mathbf{A}$.

- Under H_0, $Q(m) \stackrel{a}{\sim} \chi^2(N^2 m)$.
- When T is small, the χ^2 approximation to the distribution of the test statistic may be misleading.
- When T is small, the nominal size of the portmanteau test tends to be lower than the significance level chosen and the test has low power against many alternatives. Adjusted versions of the Q statistic can be used:
  - Hosking's statistic:

$$Q^*(m) = T(T+2) \sum_{h=1}^{m} \frac{1}{T-h} tr(\hat{\Gamma}_h \hat{\Gamma}_0^{-1} \hat{\Gamma}_h \hat{\Gamma}_0^{-1})$$

  - Li and McLeod's statistic:

$$Q^{**}(m) = T \sum_{h=1}^{m} \frac{1}{T-h} tr(\hat{\Gamma}_h \hat{\Gamma}_0^{-1} \hat{\Gamma}_h \hat{\Gamma}_0^{-1}) + \frac{N^2 m(m+1)}{2T}$$

1.5 Multivariate White Noise Process

Multivariate White Noise: Let $\mathbf{z}_{i,t} = [z_{1,t}, z_{2,t}, ..., z_{N,t}]$ be a $N \times 1$ vector of random variables. This multivariate time series is said to be a multivariate white noise if it is a stationary vector with zero mean, and if the values of $\mathbf{z_t}$ at different times are uncorrelated, i.e., $\mathbf{\Gamma}_h$ is an $N \times N$ of zeros at all $h \neq 0$.

- Each component of $\mathbf{z}_t$ simply behaves like a univariate white noise.
- The individual white noises are uncoupled in a linear sense.
- Assuming that the values of $\mathbf{z}_t$ are uncorrelated does not necessarily imply that they are independent. Independence can be inferred by the lack of correlations at all leads and lags among the random variables that enter $\mathbf{z}_t$, when the random vector follows a multivariate normal distribution.

2 Introduction to VAR Analysis

2.1 From Structural to Reduced-Form VARs

Vector Autoregressive Model Var(p): A Vector Autoregressive model of order p is a process that can be represented as

$$\mathbf{y_t} = \mathbf{a_0} + \mathbf{A_1 y_{t-1}} + \mathbf{A_2 y_{t-2}} + ... + \mathbf{A_p y_{t-p}} + \mathbf{u_t} = \mathbf{a_0} + \sum_{j=1}^{p} \mathbf{A_j y_{t-j}} + \mathbf{u_t}$$

where
$\mathbf{y}_t$ = $N \times 1$ vector containing N endogenous variables
$\mathbf{a}_0$ = $N \times 1$ vector of constants
$\mathbf{A}_1$, $\mathbf{A}_2$, ..., $\mathbf{A}_p = p$ $N \times N$ matrices of autoregressive coefficients $\mathbf{u}_t$
$= N \times 1$ vector of serially uncorrelated, white noise disturbances.

Structural VAR or **VAR in primitive form**:

$$y_{1,t} = b_{1,0} - b_{1,2} y_{2,t} + \varphi_{1,1} y_{1,t-1} + \varphi_{1,2} y_{2,t-1} + \epsilon_{1,t}$$

$$y_{2,t} = b_{2,0} - b_{2,1} y_{1,t} + \varphi_{2,1} y_{1,t-1} + \varphi_{2,2} y_{2,t-1} + \epsilon_{2,t}$$

where
$y_{1,t}$ and $y_{2,t}$ are assumed to be stationary
$\epsilon_{1,t}$ and $\epsilon_{2,t}$ are uncorrelated white-noise disturbances with standard deviation σ_1 and σ_2, respectively.

In matrix notation

$$\begin{bmatrix} 1 & \mathrm{b}_{1,2} \\ \mathrm{b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{y}_{1,t} \\ \mathrm{y}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathrm{b}_{1,0} \\ \mathrm{b}_{2,0} \end{bmatrix} + \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,1} & \varphi_{2,2} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{1,t-1} \\ \mathrm{y}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

or in compact form

$$\mathbf{B y_t} = \mathbf{Q_0} + \mathbf{Q_1 y_{t-1}} + \boldsymbol{\epsilon_t}$$

- $y_{1,t}$ depends on its own lag and on both one lag and current value of $y_{2,t}$; $y_{2,t}$ depends on its own lag and on both one lag and current value of $y_{1,t}$.
- It captures contemporaneous feedback effects:
    1. $-b_{1,2}$ measures the contemporaneous effect of a unit change of $y_{2,t}$ on $y_{1,t}$;

2. $-b_{2,1}$ measures the contemporaneous effect of a unit change of $y_{1,t}$ on $y_{2,t}$.

- Each contemporaneous variable is correlated with its own error term, therefore the regressors are not uncorrelated with the error terms as required by OLS estimation techniques.
- When $-b_{1,2} \neq 0$, $y_{2,t}$ depends on $y_{1,t}$ and on $\epsilon_{1,t}$ and will be correlated with it. When $-b_{2,1} \neq 0$, $y_{1,t}$ depends on $y_{2,t}$ and on $\epsilon_{2,t}$.
- Contemporaneous terms cannot be used in forecasting.

Reduced-form VAR or **VAR in standard form**:

$$y_{1,t} = a_{1,0} + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + u_{1,t}$$

$$y_{2,t} = a_{2,0} + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + u_{2,t}$$

that is obtained from by pre-multiplying both sides of $\mathbf{By_t} = \mathbf{Q_0} + \mathbf{Q_1 y_{t-1}} + \boldsymbol{\epsilon_t}$ by $\mathbf{B}^{-1}$

$$\mathbf{y_t} = \mathbf{a_0} + \mathbf{A_1 y_{t-1}} + \mathbf{u_t}$$

where $\mathbf{a_0} = \mathbf{B^{-1}Q_0}$, $\mathbf{A_1} = \mathbf{B^{-1}Q_1}$, $\mathbf{u_t} = \mathbf{B_t^{-1}}\boldsymbol{\epsilon_t}$.

- It does not contain contemporaneous feedback terms.
- It can be estimated equation by equation using OLS.
- $u_{1,t}$ and $u_{2,t}$ are composites of $\epsilon_{1,t}$ and $\epsilon_{2,t}$: in fact

$$\mathbf{u_t} = \mathbf{B^{-1}}\boldsymbol{\epsilon_t} \quad \text{then}$$

$$u_{1,t} = \frac{\epsilon_{1,t} - b_{1,2}\epsilon_{2,t}}{1 - b_{1,2}b_{2,1}} \quad \text{and} \quad u_{2,t} = \frac{\epsilon_{2,t} - b_{2,1}\epsilon_{1,t}}{1 - b_{1,2}b_{2,1}}$$

.

Properties (derived by the white noise processes $\epsilon_{1,t}$, $\epsilon_{2,t}$):

1.

$$E[u_{1,t}] = E[\frac{\epsilon_{1,t} - b_{1,2}\epsilon_{2,t}}{1 - b_{1,2}b_{2,1}}] = 0$$

$$E[u_{2,t}] = E[\frac{\epsilon_{2,t} - b_{2,1}\epsilon_{1,t}}{1 - b_{1,2}b_{2,1}}] = 0$$

2.

$$Var[u_{1,t}] = \frac{Var[\epsilon_{1,t} - b_{1,2}\epsilon_{2,t}]}{(1 - b_{1,2}b_{2,1})^2} =$$

$$= \frac{Var[\epsilon_{1,t}] + b_{1,2}^2 Var[\epsilon_{2,t}] - 2b_{1,2}Cov[\epsilon_{1,t}, \epsilon_{2,t}]}{(1 - b_{1,2}b_{2,1})^2} = \frac{\sigma_{\epsilon,1}^2 + b_{1,2}^2\sigma_{\epsilon,2}^2}{(1 - b_{1,2}b_{2,1})^2}$$

$$Var[u_{2,t}] = \frac{\sigma_{\epsilon,2}^2 + b_{2,1}^2\sigma_{\epsilon,1}^2}{(1 - b_{1,2}b_{2,1})^2}$$

constant over time.

3.

$$Cov[u_{1,t}, u_{2,t}] = \frac{E[(\epsilon_{1,t} - b_{1,2}\epsilon_{2,t})(\epsilon_{2,t} - b_{2,1}\epsilon_{1,t})]}{(1 - b_{1,2}b_{2,1})^2} = \frac{-(b_{2,1}\sigma_{\epsilon,1}^2 + b_{1,2}\sigma_{\epsilon,2}^2)}{(1 - b_{1,2}b_{2,1})^2}$$

- $u_{1,t}$ and $u_{2,t}$ are serially uncorrelated, but are cross-correlated unless $b_{1,2} = b_{2,1} = 0$.

4.

$$\mathbf{\Sigma}_u = \begin{bmatrix} \text{Var}[\text{u}_{1,t}] & \text{Cov}[\text{u}_{1,t}, u_{2,t}] \\ \text{Cov}[\text{u}_{1,t}, u_{2,t}] & \text{Var}[\text{u}_{2,t}] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{bmatrix}$$

- In general, it is not possible to identify the structural parameters and errors from the OLS estimates of the parameters and the residuals of the standard form VAR, unless some restrictions are imposed on the primitive system.

Recursive Choleski triangularization: Impose a Choleski decomposition on the covariance matrix of the residuals of the VAR in its standard form, that is the restriction $b_{1,2} = 0$, so that

$$y_{1,t} = b_{1,0} + \varphi_{1,1}y_{1,t-1} + \varphi_{1,2}y_{2,t-1} + \epsilon_{1,t}$$

$$y_{2,t} = b_{2,0} - b_{2,1}y_{1,t} + \varphi_{2,1}y_{1,t-1} + \varphi_{2,2}y_{2,t-1} + \epsilon_{2,t}$$

then

$$u_{1,t} = \epsilon_{1,t} \quad \text{and} \quad u_{2,t} = \epsilon_{2,t} - b_{2,1}\epsilon_{1,t}$$

In matrix form, the restriction $b_{1,2} = 0$ means that $\mathbf{B^{-1}} = \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix}$, so that

$$\begin{bmatrix} \text{y}_{1,t} \\ \text{y}_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \text{b}_{1,0} \\ \text{b}_{2,0} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,1} & \varphi_{2,2} \end{bmatrix} \begin{bmatrix} \text{y}_{1,t-1} \\ \text{y}_{2,t-1} \end{bmatrix} +$$

$$+ \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$=\begin{bmatrix} \mathrm{b}_{1,0} \\ \mathrm{b}_{2,0} - b_{1,0}b_{2,1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,1} - b_{2,1}\varphi_{1,1} & \varphi_{2,2} - b_{2,1}\varphi_{1,2} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{1,t-1} \\ \mathrm{y}_{2,t-1} \end{bmatrix} +$$

$$+ \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} - b_{2,1}\epsilon_{1,t} \end{bmatrix}$$

so that

$$a_{1,0} = b_{1,0}, \quad a_{2,0} = b_{2,0} - b_{1,0}b_{2,1}, \quad a_{1,1} = \varphi_{1,1,}, \quad a_{1,2} = \varphi_{1,2},$$

$$a_{2,1} = \varphi_{2,1} - b_{2,1}\varphi_{1,1}, \quad a_{2,2} = \varphi_{2,2} - b_{2,1}\varphi_{1,2}, \quad u_{1,t} = \epsilon_{1,t}, \quad u_{2,t} = \epsilon_{2,t} - b_{2,1}\epsilon_{1,t}$$

It follows that

$$\sigma_1^2 \equiv Var[u_{1,t}] = \sigma_{\epsilon,1}^2$$

$$\sigma_2^2 \equiv Var[u_{2,t}] = \sigma_{\epsilon,2}^2 - b_{2,1}^2\sigma_{\epsilon,1}^2$$

$$Cov[u_{1,t}, u_{2,t}] = -b_{2,1}\sigma_{\epsilon,1}^2$$

- The restriction implies that the observed values of $u_{1,t}$ are completely attributed to pure (structural) shocks to $y_{1,t}$.

Choleski decomposition of the symmetric matrix: the covariance matrix of the residuals is forced to be equal to

$$\mathbf{\Sigma_u} = \mathbf{W\Sigma W'} = \mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}^{1/2})'$$

where $\mathbf{W} = \mathbf{B^{-1}}$, $\mathbf{\Sigma}$ is the diagonal covariance matrix of the structural innovations and $\mathbf{\Sigma}^{1/2}$ is the triangular "square root" of the covariance matrix $\mathbf{\Sigma_u}$

$$\mathbf{\Sigma}_u = \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{\epsilon,1}^2 & 0 \\ 0 & \sigma_{\epsilon,2}^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix}' =$$

$$= \begin{bmatrix} 1 & 0 \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{\epsilon,1}^2 & 0 \\ 0 & \sigma_{\epsilon,2}^2 \end{bmatrix} \begin{bmatrix} 1 & \text{-b}_{2,1} \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} \sigma_{\epsilon,1}^2 & \text{-b}_{2,1}\sigma_{\epsilon,2}^2 \\ \text{-b}_{2,1}\sigma_{\epsilon,1}^2 & \sigma_{\epsilon,2}^2 - b_{2,1}^2\sigma_{\epsilon,1}^2 \end{bmatrix}$$

We can go back from the estimated $\mathbf{\Sigma_u}$ to the original (and unobserved) diagonal matrix $\mathbf{\Sigma}$, and this is equivalent, after a little bit of algebra to

$$\mathbf{\Sigma} = \mathbf{W}^{-1}\mathbf{\Sigma}_u(\mathbf{W}')^{-1}$$

- In a N-variate VAR, we need to impose $(N^2 - N)/2$ in order to retrieve the N structural shocks from the residual of the OLS estimate.

  Example for a VAR(1) with three endogenous variables:
  We need to impose $(3^2-3)/2 = 3$ restrictions that is equivalent to pre-multiplying the structural VAR by the lower triangular matrix

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \text{-b}_{1,2} & 1 & 0 \\ \text{-b}_{1,3} & \text{-b}_{2,3} & 1 \end{bmatrix}$$

  so that

$$\mathbf{u_t} = \mathbf{B}^{-1}\boldsymbol{\epsilon}_t = \begin{bmatrix} 1 & 0 & 0 \\ \text{-b}_{2,1} & 1 & 0 \\ \text{-b}_{3,1} & \text{-b}_{3,2} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix} =$$

$$= \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} - b_{2,1}\epsilon_{1,t} \\ \epsilon_{3,t} - b_{3,1}\epsilon_{1,t} - b_{3,2}\epsilon_{2,t} \end{bmatrix}$$

- There are as many Choleski decompositions as all the possible orderings of the variables. Therefore, when we apply a Choleski triangular identification scheme to a VAR model we are introducing a number of (potentially arbitrary) assumptions on the contemporaneous relationships among the variables.

2.2 Stationarity Conditions and the Population Moments of a VAR(1) Process

Properties of a reduced-form, standard VAR(1) model:

(a)

$$E[\mathbf{y}_t] = \mathbf{a}_0 + \mathbf{A}_1 E[\mathbf{y}_{t-1}]$$

time invariant and thus

$$\boldsymbol{\mu} \equiv E[\mathbf{y}_t] = (\mathbf{I}_N - \mathbf{A}_1)^{-1}\mathbf{a}_0$$

where $(\mathbf{I}_N - \mathbf{A}_1)$ is a non-singular matrix and $\mathbf{I}_N$ is the $N \times N$ identity matrix

(b)

$$\boldsymbol{\mu}_{t|t-1} \equiv E[\mathbf{y}_t|\Im_{t-1}] = E[\mathbf{y}_t|\mathbf{y}_{t-1}] = \mathbf{a}_0 + \mathbf{A}_1\mathbf{y}_{t-1}$$

(c) Given that $\mathbf{a}_0 = (\mathbf{I}_N - \mathbf{A}_1)\boldsymbol{\mu}$, the VAR(1) model can be rewritten as

$$\mathbf{y}_t - \boldsymbol{\mu} = \mathbf{A}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{u}_t$$

so that if the mean-corrected time-series $\tilde{\mathbf{y}}_t \equiv \mathbf{y}_t - \boldsymbol{\mu}$ then

$$\tilde{\mathbf{y}}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{u}_t$$

and substituting $\mathbf{y}_{t-1} = \mathbf{A}_1\mathbf{y}_{t-2} + \mathbf{u}_{t-1}$, then $\mathbf{y}_{t-2} = \mathbf{A}_1\mathbf{y}_{t-3} + \mathbf{u}_{t-2}$ and keeping iterating, we obtain

$$\tilde{\mathbf{y}}_t = \mathbf{A}_1(\mathbf{A}_1\mathbf{y}_{t-2} + \mathbf{u}_{t-1}) + \mathbf{u}_t = \mathbf{A}_1^2\mathbf{y}_{t-2} + \mathbf{A}_1\mathbf{u}_{t-1} + \mathbf{u}_t =$$

$$= \mathbf{u}_t + \mathbf{A}_1\mathbf{u}_{t-1} + \mathbf{A}_1^2\mathbf{u}_{t-2} + \mathbf{A}_1^3\mathbf{u}_{t-3} + ... = \sum_{i=1}^{\infty} \mathbf{A}_1^i\mathbf{u}_{t-i} + \mathbf{u}_t$$

and

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \mathbf{A}_1^i\mathbf{u}_{t-i} + \mathbf{u}_t$$

Vector moving average (VMA) infinite representation of the VAR(1) model:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \boldsymbol{\Theta}_i\mathbf{u}_{t-i} + \mathbf{u}_t$$

that is derived by $\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \mathbf{A}_1^i\mathbf{u}_{t-i} + \mathbf{u}_t$, where we define $\boldsymbol{\Theta}_i \equiv \mathbf{A}_1^i$.

- It represents the multivariate extension of the Wold's representation theorem.

Properties:

(a) $\mathbf{u}_t$ is serially uncorrelated and it is also uncorrelated with the past values of $\mathbf{y}_t$

$$Cov[\mathbf{u}_t, \mathbf{y}_{t-1}] = \mathbf{0}$$

and $\mathbf{u}_t$ is called vector of innovations of the series at time t.

(b)

$$Cov[\mathbf{y}_t, \mathbf{u}_t] = \boldsymbol{\Sigma}_u$$

derived by post-multiplying $\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \mathbf{A}_1^i\mathbf{u}_{t-i} + \mathbf{u}_t$ by $\mathbf{u}_t'$, taking the expectation, and exploiting the fact that $\mathbf{u}_t$ is serially uncorrelated.

(c) $\mathbf{y}_t$ depends on the past innovations $\mathbf{u}_{t-j}$ with a coefficient matrix $\mathbf{A}_1^j$.

- $\mathbf{y}_t$ is stable if $det(\mathbf{I}_N - \mathbf{A}_1 z) \neq 0$ for $|z| \leq 1$.

(d)

$$Cov[\mathbf{y}_t] \equiv \Gamma_0 = \mathbf{\Sigma}_u + \mathbf{A}_1\mathbf{\Sigma}_u\mathbf{A}_1' + \mathbf{A}_1^2\mathbf{\Sigma}_u(\mathbf{A}_1^2)' + ... = \sum_{i=0}^{\infty}\mathbf{A}_1^i\mathbf{\Sigma}_u(\mathbf{A}_1^i)'$$

where $\mathbf{A}_1^0$ is a $N \times N$ $\mathbf{I}_N$, or alternatively

$$\Gamma_0 = \sum_{i=0}^{\infty}\mathbf{\Theta}_i\mathbf{\Sigma}_u\mathbf{\Theta}_i'$$

where $\mathbf{\Theta}_i =$ the coefficients of the moving average representations of the VAR, that can be derived from

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}(L)\mathbf{y}_t + \mathbf{u}_t$$

that can be rewritten as

$$\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{u}_t$$

where $\mathbf{A}(L) \equiv \mathbf{I}_N - \mathbf{A}(L)$. Let

$$\mathbf{\Theta}(L) \equiv \sum_{i=0}^{\infty}\mathbf{\Theta}_i\mathbf{L}^i$$

be an operator such that $\mathbf{\Theta}(L)\mathbf{A}(L) = \mathbf{I}_N$, then post-multiply $\mathbf{A}(L)\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{u}_t$ by $\mathbf{\Theta}(L)$ and obtain

$$\mathbf{y}_t = \mathbf{\Theta}(L)\boldsymbol{\mu} + \mathbf{\Theta}(L)\mathbf{u}_t$$

that is

$$\mathbf{y}_t = \sum_{i=0}^{\infty}\mathbf{\Theta}_i\boldsymbol{\mu} + \sum_{i=0}^{\infty}\mathbf{\Theta}_i\mathbf{u}_{t-i}$$

$\mathbf{\Theta}(L)$ is the inverse of $\mathbf{A}(L)$, then

$$\mathbf{\Theta}_i = \sum_{j=1}^{i}\mathbf{\Theta}_{i-j}\mathbf{A}_1 \quad \text{with} \quad \mathbf{\Theta}_0 = \mathbf{I}_N$$

Given that $Cov[\mathbf{u}_t, \mathbf{y}_{t-j}] = E[\mathbf{u}_t\mathbf{y}_{t-j}'] = 0$ for $j > 0$ if we post-multiply the expression of $\mathbf{y}_t$ by $\mathbf{y}_{t-h}$ we obtain

$$E(\mathbf{y}_t\mathbf{y}_{t+1-h}) = \mathbf{A}_1 E(\mathbf{y}_t\mathbf{y}_{t-h})' \quad \text{for} \quad h > 0$$

Therefore

$$\mathbf{\Gamma}_h = \mathbf{A}_1\mathbf{\Gamma}_{h-1} = \mathbf{A}_1^h\mathbf{\Gamma}_0 \quad \text{for} \quad h > 0$$

Finally, if we post-multiply $\mathbf{\Gamma}_h$ by $\mathbf{D}^{-1/2}$ we obtain

$$\boldsymbol{\rho}_h = \mathbf{D}^{-1/2}\mathbf{A}_1\mathbf{\Gamma}_{h-1}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{A}_1\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{\Gamma}_{h-1}\mathbf{D}^{1/2} =$$

$$\mathbf{\Psi}\boldsymbol{\rho}_{h-1} = \mathbf{\Psi}^h\boldsymbol{\rho}_0 \quad \text{for} \quad h > 0$$

where $\mathbf{\Psi} = \mathbf{D}^{-1/2}\mathbf{A}_1\mathbf{D}^{-1/2}$

(e)

$$Cov[\mathbf{y}_t|\Im_{t-1}] = Cov[\mathbf{y}_t|\mathbf{y}_{t-1}] = \mathbf{A}_1 Cov[\mathbf{y}_{t-1}|\mathbf{y}_{t-1}]\mathbf{A}_1' + \mathbf{\Sigma}_u = \mathbf{\Sigma}_u$$

because $Cov[\mathbf{y}_{t-1}|\mathbf{y}_{t-1}] = 0$.

- When the residuals are simultaneously uncorrelated (i.e., $\mathbf{\Sigma}_u$ is diagonal), then also $Cov[\mathbf{y}_t|\mathbf{y}_{t-1}]$ will be diagonal.

2.3 Generalization to a VAR(p) Model

Starting from the VAR(p) model equation

$$\mathbf{y_t} = \mathbf{a_0} + \mathbf{A_1 y_{t-1}} + \mathbf{A_2 y_{t-2}} + ... + \mathbf{A_p y_{t-p}} + \mathbf{u_t}$$

Rewrite using the lag operator

$$(\mathbf{I}_N - \mathbf{A}_1\mathbf{L} - ... - \mathbf{A}_p\mathbf{p})\mathbf{y}_t = \mathbf{a}_0 + \mathbf{u}_t$$

In compact form

$$\mathbf{A}(\mathbf{L})\mathbf{y}_t = \mathbf{a}_0 + \mathbf{u}_t$$

where $\mathbf{A}(\mathbf{L}) = \mathbf{I}_N - \mathbf{A}_1\mathbf{L} - ... - \mathbf{A}_p\mathbf{L}^p$.

Properties:

(a)

$$\boldsymbol{\mu} = E[\mathbf{y}_t] = (\mathbf{I}_N - \mathbf{A}_1 - ... - \mathbf{A}_p)^{-1}\mathbf{a}_0$$

provided that the inverse of the matrix $(\mathbf{I}_N - \mathbf{A}_1 - ... - \mathbf{A}_p)$ exists, and

$$\boldsymbol{\mu}_{t|t-1} \equiv E[\mathbf{y}_t|\mathbf{y}_{t-1}] = \mathbf{a}_0 + \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j}$$

(b)

$$Cov[\mathbf{y}_t, \mathbf{u}_t] = \mathbf{\Sigma}_u$$

(c)

$$Cov[\mathbf{y}_{t-h}, \mathbf{u}_t] = 0 \quad \text{for any} \quad h > 0$$

(d)

$$\mathbf{\Gamma}_h = \mathbf{A}_1\mathbf{\Gamma}_{h-1} + ... + \mathbf{A}_p\mathbf{\Gamma}_{h-p} \quad \text{for} \quad h > 0$$

(e)
$$\boldsymbol{\rho}_h = \boldsymbol{\Psi}_1 \boldsymbol{\rho}_{h-1} + ... + \boldsymbol{\Psi}_p \boldsymbol{\rho}_{h-p} \quad \text{for} \quad h > 0$$
where $\boldsymbol{\Psi}_i = \mathbf{D}^{-1/2}\mathbf{A}_i\mathbf{D}^{1/2}$.

Representation of VAR(p) as a Np-dimensional VAR(1)

$$\underset{(Np)\times 1}{\boldsymbol{\xi}_t} \equiv \begin{bmatrix} \mathbf{y}_t' \\ \mathbf{y}_{t-1}' \\ \\ \mathbf{y}_{t-p+1}' \end{bmatrix}, \underset{(Np\times Np)}{\mathbf{F}_1} \equiv \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & & \mathbf{A}_p \\ \mathbf{I}_N & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N & & \mathbf{0} \\ \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \underset{(Np)\times 1}{\mathbf{U}_t} \equiv [\mathbf{u}_t \quad \mathbf{0} \quad ... \quad \mathbf{0}]'$$

Then
$$\boldsymbol{\xi}_t = \mathbf{F}_1 \boldsymbol{\xi}_{t-1} + \mathbf{U}_t$$
where
$$E[\mathbf{U}_t\mathbf{U}_t'] = \begin{bmatrix} \boldsymbol{\Sigma}_u & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ & & ... & \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \end{bmatrix} \quad \text{and} \quad E[\mathbf{U}_t\mathbf{U}_{t-h}'] = 0 \quad \text{for} \quad h > 0$$

In VMA representation
$$\boldsymbol{\xi}_t = \mathbf{U}_t + \mathbf{F}_1\mathbf{U}_{t-1} + \mathbf{F}_1^2\mathbf{U}_{t-2} + ... = \sum_{i=1}^{\infty} \mathbf{F}_1^i \mathbf{U}_{t-i} + \mathbf{U}_t = \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i \mathbf{U}_{t-i} + \mathbf{U}_t$$
where $\boldsymbol{\Pi}_i \equiv \mathbf{J}\mathbf{F}_1^i\mathbf{J}'$ and $\mathbf{J} \equiv [\mathbf{I}_N, \mathbf{0}, ..., \mathbf{0}]'$.

- A VAR(p) model is stable (and thus stationary) as long as the eigenvalues of the companion matrix $\mathbf{F}_1$ are all less than one in modulus, which implies $det(\mathbf{I}_N - \mathbf{A}_1\mathbf{z} - ... - \mathbf{A}_p\mathbf{z}^p) \neq 0$ for $|z| \leq 1$.

2.4 Estimation of a VAR(p) Model

Multivariate LS estimator:

Starting from
$$\mathbf{Y} = \mathbf{B}\mathbf{Z} + \mathbf{U}$$
where $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t]$, $\mathbf{B} \equiv [\mathbf{a}_0, \mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_p]$, $\mathbf{U} \equiv [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_T]$, $\mathbf{Z} \equiv [\mathbf{Z}_0, \mathbf{Z}_1, \mathbf{Z}_{T-1}]$ with $\mathbf{Z}_t \equiv [\mathbf{1}', \mathbf{y}_{t-1}', \mathbf{y}_{t-2}', ..., \mathbf{y}_{t-p+1}']'$. Given that $\mathbf{y} \equiv vec(\mathbf{Y})$, $\boldsymbol{\beta} \equiv vec(\mathbf{B})$ and $\mathbf{u} \equiv vec(\mathbf{U})$ the multivariate LS estimator is
$$\hat{\boldsymbol{\beta}} = ((\mathbf{Z}\mathbf{Z}')^{-1} \otimes \boldsymbol{\Sigma}_u)(\mathbf{Z} \otimes \boldsymbol{\Sigma}_u^{-1})\mathbf{y} = ((\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z} \otimes \mathbf{I}_N)\mathbf{y}$$

that minimizes

$$S(\boldsymbol{\beta}) = \mathbf{u}'(\mathbf{I}_N \boldsymbol{\Sigma}_u)^{-1}\mathbf{u}$$

- When a reduced-form VAR is unconstrained, the GLS estimator is the same as the OLS estimator, $\hat{\mathbf{B}}$, and therefore an unconstrained VAR can be estimated equation by equation by OLS.

Asymptotic properties of the OLS estimator $\hat{\mathbf{B}}$ (under standard assumptions):

(a) Consistent and asymptotically normally distributed

$$\sqrt{T}vec(\hat{\mathbf{B}} - \mathbf{B}) \xrightarrow{D} N(0, \boldsymbol{\Sigma}_{\hat{\mathbf{B}}}) \quad \text{or} \quad vec(\hat{\mathbf{B}}) \overset{a}{\sim} N(vec(\mathbf{B}), \boldsymbol{\Sigma}_{\hat{\mathbf{B}}}/T)$$

where $\boldsymbol{\Sigma}_{\hat{\mathbf{B}}} = plim(\mathbf{Z}\mathbf{Z}'/T)^{-1} \otimes \boldsymbol{\Sigma}_u$.

(b)

$$\hat{\boldsymbol{\Sigma}}_u = \frac{1}{T - Np}\sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' \quad \text{or} \quad \tilde{\boldsymbol{\Sigma}}_u = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'$$

where $\hat{\mathbf{u}}_t = \mathbf{y}_t - \mathbf{B}\mathbf{Z}_{t-1}$.

Multivariate ML estimator:

Under the assumptions:

(a) Sample of T observations on $\mathbf{Y}$ and a pre-sample of p initial conditions $y_{-p+1}, y_{-p+2}, ..., y_0$.

(b) Stationary process and Gaussian multivariate white noise innovations.
$\Rightarrow \quad \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T]'$ is jointly normally distributed.

(c) Gaussian multivariate white noise (then innovations at different times are independent).

(d) Noise error terms are independent with $\boldsymbol{\Sigma}_u$, then the covariance matrix of $\mathbf{u}$ is $\boldsymbol{\Sigma}_U = \mathbf{I}_T \otimes \boldsymbol{\Sigma}_u$ and its normal density is
$f_u(\mathbf{u}) = (2\pi)^{-\frac{NT}{2}}|\mathbf{I}_T \otimes \boldsymbol{\Sigma}_u|^{-\frac{1}{2}} exp(-\frac{1}{2}\mathbf{u}'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_u^{-1})\mathbf{u})$.

(e) $f_y(\mathbf{y}) = (2\pi)^{-\frac{NT}{2}}|\mathbf{I}_T \otimes \boldsymbol{\Sigma}_u|^{-\frac{1}{2}} exp(-\frac{1}{2}(\mathbf{Y} - \mathbf{B}\mathbf{Z})'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_u^{-1})(\mathbf{Y} - \mathbf{B}\mathbf{Z}))$.

the ML estimator maximizes

$$\ell(\mathbf{B}, \boldsymbol{\Sigma}_u; \mathbf{Y}, \mathbf{Z}) = ln f_y(\mathbf{Y}) = -\frac{NT}{2}ln(2\pi) - \frac{T}{2}ln|\boldsymbol{\Sigma}_u| - \frac{1}{2}(\mathbf{Y} - \mathbf{B}\mathbf{Z})'(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_u^{-1})(\mathbf{Y} - \mathbf{B}\mathbf{Z}) =$$

$$= -\frac{NT}{2} ln(2\pi) - \frac{T}{2} ln|\mathbf{\Sigma}_u| - \frac{1}{2} tr(\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U})$$

- For an unconstrained VAR, the ML and OLS estimators are the same under the assumption of Gaussian innovations.

Average cross- vector product of the OLS residuals: the ML estimator of the matrix $\mathbf{\Sigma}_u$ is

$$\tilde{\mathbf{\Sigma}}_u = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'$$

Concentrated log-likelihood of the VAR(p) model: Substituting the expression for the matrix $\mathbf{\Sigma}_u$ that maximizes the likelihood, in the class of all symmetric positive definite matrices, we obtain

$$\ell(\mathbf{B}, \mathbf{\Sigma}_u; \mathbf{Y}, \mathbf{Z}) = -\frac{NT}{2} ln(2\pi) - \frac{T}{2} ln|\tilde{\mathbf{\Sigma}}_u| - \frac{1}{2} NT$$

- Optimizing $\ell(\mathbf{B}, \mathbf{\Sigma}_u; \mathbf{Y}, \mathbf{Z}) = -\frac{NT}{2} ln(2\pi) - \frac{T}{2} ln|\mathbf{\Sigma}_u| - \frac{1}{2} tr(\mathbf{U}'\mathbf{\Sigma}_u^{-1}\mathbf{U})$ in one pass or maximizing over $\tilde{\mathbf{\Sigma}}_u$ and $\ell(\mathbf{B}, \mathbf{\Sigma}_u; \mathbf{Y}, \mathbf{Z}) = -\frac{NT}{2} ln(2\pi) - \frac{T}{2} ln|\tilde{\mathbf{\Sigma}}_u| - \frac{1}{2} NT$ iterating between the two objects until convergence is achieved, will return identical results.

2.5 Specification of a VAR Model and Hypothesis Testing

- If the order of the VAR model increases increasing, the (absolute) size of the residuals decreases and both the fit of the model and its forecasting power increase.
- If the number of parameters increases, its in-sample accuracy increases, while its out-of-sample predictive power decreases.

Restricted, standard VAR: models in which the structure and number of lags included in each equation may vary across different equations.

Methods to select p:

(a) Multivariate information criteria:

i. (M)AIC: $ln|\tilde{\mathbf{\Sigma}}_u| + 2\frac{K}{T}$

ii. (M)SBC: $ln|\tilde{\mathbf{\Sigma}}_u| + \frac{K}{T}ln(T)$

iii. (M)HQIC: $ln|\tilde{\mathbf{\Sigma}}_u| + 2\frac{K}{T}ln(ln(T))$

(b) Final predictor error (FPE):

$$FPE(p) = [\frac{T + Np + 1}{T - Np + 1}]^N |\tilde{\mathbf{\Sigma}}_u|$$

where $|\tilde{\mathbf{\Sigma}}_u|$ =determinant of the estimated covariance matrix of the residuals from a given VAR(p) model.

(c) General-to-simple approach: use sequential Likelihood Ratio (LR) test:

$$H_0 : p_0 \quad \text{lags are sufficient}$$

$$LRT(p_0, p_1) = T(ln|\tilde{\mathbf{\Sigma}}_u^{p_0}| - |\tilde{\mathbf{\Sigma}}_u^{p_1}|) \overset{a}{\sim} \chi^2(N(p_1 - p_0))$$

where $\tilde{\mathbf{\Sigma}}_u^{p_0}$=determinant of the covariance matrix estimated under H_0:VAR includes p_0 lags and $\tilde{\mathbf{\Sigma}}_u^{p_1}$=determinant of the covariance matrix estimated under H_1:VAR includes p_1 lags.

Alternative statistics:

$$LRT'(p_0, p_1) = (T - Np - 1)(ln|\tilde{\Sigma}_u^{p_0}| - |\tilde{\Sigma}_u^{p_0}|) \overset{a}{\sim} \chi^2(N(p_1 - p_0))$$

- LR tests can only be used to perform a pairwise comparison of two VAR systems, one that is a restricted (nested inside) version of the bigger VAR. Then a simple-to-general approach is not possible.
- If the assumption that errors from each equation are normally distributed is not respected, the test is not valid.
- When the sample size is small, the test may be subject to substantial size distortions.

2.6 Forecasting with a VAR model

Forecasting method: Minimization of the mean squared forecast error (MSFE) using the loss function.

Assumption: $\mathbf{u}_t$ is an independent multivariate white noise, such that $\mathbf{u}_t$ and $\mathbf{u}_s$ are independent for $t \neq s \Rightarrow \quad \mathrm{E}_t[\mathbf{u}_{t+h}|\Im_t] = 0$ for $h > 0$.

Minimized time t MSFE prediction at h:

$$E_t[\mathbf{y}_{t+h}|\Im_t] = E_t[\mathbf{y}_{t+h}|\{\mathbf{y}_s|s \leq t\}] =$$
$$= \mathbf{a}_0 + \mathbf{A}_1 E_t[\mathbf{y}_{t+h-1}|\Im_t] + ... + \mathbf{A}_p E_t[\mathbf{y}_{t+h-p}|\Im_t]$$

Best linear predictor in terms of MSFE minimization:

$$E_t[\mathbf{y}_{t+h}|\Im_t] = \mathbf{a}_0 + \mathbf{A}_1 E_t[\mathbf{y}_{t+h-1}|\Im_t] + ... + \mathbf{A}_p E_t[\mathbf{y}_{t+h-p}|\Im_t]$$

Properties:

i. Unbiased predictor, i.e. $E[\mathbf{y}_{t+h} - E[\mathbf{y}_{t+h}|\Im_t]] = 0$.

ii. If $\mathbf{u}_t$ is an independent white noise vector, then $MSFE[E_t[\mathbf{y}_{t+h}]] = MSFE[E_t[\mathbf{y}_{t+h}|\mathbf{y}_t, \mathbf{y}_{t-1}, ...]]$.

3 Structural Analysis with VAR Models

3.1 Impulse Response Functions

- VAR models can be used to understand the dynamic relationships between the variables of interest.

Impulse Response Function: In the context of a VAR model, an impulse response function traces out the time path of the effects of an exogenous shock to one (or more) of the endogenous variables on some or all of the other variables in a VAR system.

Impact multipliers: coefficients of the matrix $\mathbf{\Phi}_i$.

Starting from the moving average representation of a VAR(1)

$$\mathbf{y}_t = \mathbf{a}_0 + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{u}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_1^i \mathbf{u}_{t-i} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{\Theta}_i \mathbf{u}_{t-i}$$

or, for example in the case of a VAR(1):

$$\begin{bmatrix} \mathrm{y}_{1,t} \\ \mathrm{y}_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \theta_{1,1(i)} & \theta_{1,2(i)} \\ \theta_{2,1(i)} & \theta_{2,2(i)} \end{bmatrix} \begin{bmatrix} \mathrm{u}_{1,t-i} \\ \mathrm{u}_{2,t-i} \end{bmatrix}$$

where

$$\begin{bmatrix} \mathrm{u}_{1,t} \\ \mathrm{u}_{2,t} \end{bmatrix} = \frac{1}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & \text{-b}_{1,2} \\ \text{-b}_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Then

$$\mathbf{\Phi}_i = \frac{\mathbf{A}_1^i}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & \text{-b}_{1,2} \\ \text{-b}_{2,1} & 1 \end{bmatrix} = \frac{\mathbf{\Theta}_i}{1 - b_{1,2}b_{2,1}} \begin{bmatrix} 1 & \text{-b}_{1,2} \\ \text{-b}_{2,1} & 1 \end{bmatrix}$$

and

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{\Phi}_i \boldsymbol{\epsilon}_{t-i}$$

For example, $\phi_{1,2(0)}$ is the instantaneous impact on $y_{1,t}$ of a one-unit change in $\epsilon_{2,t}$.

Cumulative response of the variable j to a shock to the variable k:

$$\sum_{i=0}^{H} \phi_{j,k,(i)}$$

For example, $\sum_{i=0}^{H} \phi_{1,2,(i)}$ is the cumulative effects of a one-unit shock (or impulse) to $\epsilon_{2,t}$ on the variable $y_{1,t}$ after H periods.

Long-run impact multipliers: impact multipliers when $H \to \infty$.

- The set of elements $\phi_{j,k(i)}$, with $i = 1, ..., H$ is the impulse response function of the jth variable of the system, up to the period H.
- VAR in its reduced form is under-identified by construction and therefore $\phi_{j,k(i)}$ cannot be computed from the OLS estimates of the VAR in its standard form without imposing adequate restrictions.
- Choleski decompositions provide a minimal set of restrictions concerning the simultaneous relationships among variables that can be used to identify the structural model, but this method forces potentially important identification asymmetry on the system.
- IRFs are constructed using estimated coefficient, thus will contain sampling error. Therefore, it is advisable to construct confidence intervals around them to account for the uncertainty that derives from parameter estimation.

- Bootstrapping techniques are usually used to compute confidence intervals, as these are more reliable and avoid the complex computation of exact expressions for the asymptotic variance of the IRF coefficients.

Bootstrapping methods:

i. Estimate each equation using OLS/MLE and construct $\{\mathbf{u}_t^b\}$ by randomly sampling with replacement from the estimated residuals.

ii. Use $\{\mathbf{u}_t^b\}$ and the estimated coefficients to construct a pseudo-vector of endogenous variable series $\{\mathbf{y}_t^b\}$.

iii. Discard the coefficients used to generate $\{\mathbf{y}_t^b\}$ and estimate new coefficients from $\{\mathbf{y}_t^b\}$. The impulse response functions are computed from the newly estimated coefficients and saved, also indexed by the bootstrap iteration b.

- An impulse response function is considered to be statistically significant if zero is not included in the bootstrapped confidence interval.

3.2 Variance Decompositions

Forecast error variance decomposition:

Starting from the VMA representation of the model,

$$\mathbf{u}_t(h) = \mathbf{y}_{t+h} - E_t[\mathbf{y}_{t+h}] = \sum_{i=0}^{h-1} \mathbf{\Phi}_i \boldsymbol{\epsilon}_{t+h-i}$$

then

$$u_{y1}(h) = y_{1,t+h} - E[y_{1,t+h|t}] = \phi_{1,1}(0)\epsilon_{1,t+h} + \phi_{1,1}(1)\epsilon_{1,t+h-1} + ... + \phi_{1,1}(h-1)\epsilon_{1,t+1} +$$
$$+\phi_{1,2}(0)\epsilon_{2,t+h} + \phi_{1,2}(1)\epsilon_{2,t+h-1} + ... + \phi_{1,2}(h-1)\epsilon_{2,t+1}$$

and

$$\sigma_{y1}^2(h) = \sigma_{y1}^2[\phi_{1,1}^2(0) + \phi_{1,1}^2(1) + ... + \phi_{1,1}^2(h-1)] + \sigma_{y2}^2[\phi_{1,2}^2(0) + \phi_{1,2}^2(1) + ... + \phi_{1,2}^2(h-1)]$$

- The variance of the forecast error increases as the forecast horizon h increases because all the coefficients $\phi_{j,k}^2$ are non-negative as they are squared.

Therefore, the h-step-ahead forecast error variance can be decomposed in

i. proportion due to the shocks in $\{\epsilon_{1,t}\}$

$$\frac{\sigma_{y1}^2[\phi_{1,1}^2(0) + \phi_{1,1}^2(1) + ... + \phi_{1,1}^2(h-1)]}{\sigma_{y1}^2(h)}$$

ii. proportion of forecast error variance due to the shocks in the sequence $\{\epsilon_{2,t}\}$

$$\frac{\sigma_{y2}^2[\phi_{1,2}^2(0) + \phi_{1,2}^2(1) + ... + \phi_{1,2}^2(h-1)]}{\sigma_{y1}^2(h)}$$

- Variance decompositions determine how much of the h-step-ahead forecast error variance of a given variable is explained by innovations to each explanatory variable for $h = 1, 2, ...$.
- They require identification, therefore Choleski decompositions (or other restriction schemes) are typically imposed.

Innovation accounting: approach where the forecast error variance decomposition and the impulse response function are combined to uncover the dynamic interrelationships among the endogenous variables.

3.3 Granger Causality

Granger causality: Let $\Im_t$ be the information set containing all the relevant information available up to and including time t. In addition, let $\mathbf{y}_t(h|\Im_t)$ be the optimal (minimum MSFE) h-step-ahead prediction of the process $\{\mathbf{y}_t\}$ at the forecast origin t, based on $\Im_t$. The vector time series process $\{\mathbf{x}_t\}$ is said to (Granger-) cause $\{\mathbf{y}_t\}$ in a Granger sense if and only if

$$MSFE_{yt}(h|\Im_t) < MSFE_{yt}(h|\Im_t\{\mathbf{x}_s|s \leq t\})$$

Feedback system: represented by the joint process $\{\mathbf{x}_t', \mathbf{y}_t'\}'$ when $\{\mathbf{x}_t\}$ causes $\{\mathbf{y}_t\}$ and $\{\mathbf{y}_t\}$ causes $\{\mathbf{x}_t\}$.

- We only consider the information in the past and present values of the process under examination, rather than the entire $\Im_t$ because $\Im_t$ of all the existent relevant information is rarely available.

Granger Causality - Restricted: Let $\mathbf{y}_t(h|\{\mathbf{x}_s, \mathbf{y}_s|s \leq t\})$ be the optimal linear (minimum MSFE) h-step-ahead prediction function of the process $\{\mathbf{y}_t\}$ at the forecast origin t, based on the information $\{\mathbf{x}_s, \mathbf{y}_s|s \leq t\}$. The process $\{\mathbf{x}_t\}$ is said to Granger cause $\{\mathbf{y}_t\}$ if

$$MSFE(E[\mathbf{y}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...]) \leq MSFE(E[\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...])$$

- Difference between Granger causality and exogeneity: for $\mathbf{y}_t$ to be exogenous it is required that it is not affected by the contemporaneous value of $\mathbf{x}_t$, while Granger causality refers to the effects of the past values of $\{\mathbf{x}_t\}$ on the current value of $\mathbf{y}_t$.

- The lack of causality can be assessed by looking at the representation of the VAR in its standard form.

- The lack of Granger causality can be verified using a standard F-test of the restriction $a_{1,2(1)} = a_{1,2(2)} = ... = a_{1,2(p)} = 0$, where, for instance in the case of $N = 2$

$$\begin{bmatrix} \mathrm{y}_{1,t} \\ \mathrm{y}_{2,t} \end{bmatrix} = \begin{bmatrix} \mathrm{a}_{1,0} \\ \mathrm{a}_{2,0} \end{bmatrix} + \begin{bmatrix} \mathrm{a}_{1,1(1)} & \mathrm{a}_{1,2(1)} \\ \mathrm{a}_{2,1(1)} & \mathrm{a}_{2,2(1)} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{1,t-1} \\ \mathrm{y}_{2,t-1} \end{bmatrix} + ... +$$

$$+ \begin{bmatrix} \mathrm{a}_{1,1(p)} & \mathrm{a}_{1,2(p)} \\ \mathrm{a}_{2,1(p)} & \mathrm{a}_{2,2(p)} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{1,t-p} \\ \mathrm{y}_{2,t-p} \end{bmatrix} + \begin{bmatrix} \mathrm{u}_{1,t} \\ \mathrm{u}_{2,t} \end{bmatrix}$$

Block-causality tests: verify whether one variable, $y_{n,t}$, Granger causes any other variables in the system, that is, whether taking into account the lagged value of $y_{n,t}$ helps forecasting any of the other variables in the VAR.They consist of likelihood ratio tests

$$(T - m)(ln|\tilde{\Sigma}_u^R| - |\tilde{\Sigma}_u^U|)$$

where $\tilde{\Sigma}_u^R$ is the covariance matrix of the residuals from a model that has been restricted to have all the coefficients of the lags of the variable $y_{n,t} = 0$ and $\tilde{\Sigma}_u^U$ is the residual covariance matrix of the unrestricted model.