Multivariate Time Series

Consider n time series variables $\{y_{1t}\}, \ldots, \{y_{nt}\}$. A *multivariate time series* is the $(n \times 1)$ vector time series $\{\mathbf{Y}_t\}$ where the i^{th} row of $\{\mathbf{Y}_t\}$ is $\{y_{it}\}$. That is, for any time t, $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$.

Multivariate time series analysis is used when one wants to model and explain the interactions and co-movements among a group of time series variables:

- Consumption and income
- Stock prices and dividends
- Forward and spot exchange rates
- interest rates, money growth, income, inflation

Stock and Watson state that macroeconometricians do four things with multivariate time series

1. Describe and summarize macroeconomic data

2. Make macroeconomic forecasts

3. Quantify what we do or do not know about the true structure of the macroeconomy

4. Advise macroeconomic policymakers

Stationary and Ergodic Multivariate Time Series

A multivariate time series $\mathbf{Y}_t$ is covariance stationary and ergodic if all of its component time series are stationary and ergodic.

$$E[\mathbf{Y}_t] = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$$
$$\text{var}(\mathbf{Y}_t) = \Gamma_0 = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})']$$
$$= \begin{pmatrix} \text{var}(y_{1t}) & \text{cov}(y_{1t}, y_{2t}) & \cdots & \text{cov}(y_{1t}, y_{nt}) \\ \text{cov}(y_{2t}, y_{1t}) & \text{var}(y_{2t}) & \cdots & \text{cov}(y_{2t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(y_{nt}, y_{1t}) & \text{cov}(y_{nt}, y_{2t}) & \cdots & \text{var}(y_{nt}) \end{pmatrix}$$

The correlation matrix of $\mathbf{Y}_t$ is the $(n \times n)$ matrix

$$\text{corr}(\mathbf{Y}_t) = \mathbf{R}_0 = \mathbf{D}^{-1}\mathbf{\Gamma}_0\mathbf{D}^{-1}$$

where $\mathbf{D}$ is an $(n \times n)$ diagonal matrix with j^{th} diagonal element $(\gamma_{jj}^0)^{1/2}$ =var$(y_{jt})^{1/2}$.

The parameters μ, Γ_0 and $\mathbf{R}_0$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ using the sample moments

$$
\begin{aligned}
\bar{\mathbf{Y}} &= \frac{1}{T}\sum_{t=1}^{T} \mathbf{Y}_t \xrightarrow{p} E[\mathbf{Y}_t] = \boldsymbol{\mu} \\
\hat{\mathbf{\Gamma}}_0 &= \frac{1}{T}\sum_{t=1}^{T} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})' \xrightarrow{p} \text{var}(\mathbf{Y}_t) = \mathbf{\Gamma}_0 \\
\hat{\mathbf{R}}_0 &= \hat{\mathbf{D}}^{-1}\hat{\mathbf{\Gamma}}_0\hat{\mathbf{D}}^{-1} \xrightarrow{p} \text{corr}(\mathbf{Y}_t) = \mathbf{R}_0
\end{aligned}
$$

where $\hat{\mathbf{D}}$ is the $(n \times n)$ diagonal matrix with the sample standard deviations of y_{jt} along the diagonal. The Ergodic Theorem justifies convergence of the sample moments to their population counterparts.

Cross Covariance and Correlation Matrices

With a multivariate time series $\mathbf{Y}_t$ each component has autocovariances and autocorrelations but there are also cross lead-lag covariances and correlations between all possible pairs of components. The autocovariances and autocorrelations of y_{jt} for $j = 1, \ldots, n$ are defined as

$$\begin{aligned} \gamma_{jj}^{k} &= \text{cov}(y_{jt}, y_{jt-k}), \\ \rho_{jj}^{k} &= \text{corr}(y_{jt}, y_{jt-k}) = \frac{\gamma_{jj}^{k}}{\gamma_{jj}^{0}} \end{aligned}$$

and these are symmetric in k: $\gamma_{jj}^{k} = \gamma_{jj}^{-k}$, $\rho_{jj}^{k} = \rho_{jj}^{-k}$.

The *cross lag covariances* and *cross lag correlations* between y_{it} and y_{jt} are defined as

$$\begin{aligned}\gamma_{ij}^{k} &= \text{cov}(y_{it}, y_{jt-k}),\\ \rho_{ij}^{k} &= \text{corr}(y_{jt}, y_{jt-k}) = \frac{\gamma_{ij}^{k}}{\sqrt{\gamma_{ii}^{0}\gamma_{jj}^{0}}}\end{aligned}$$

and they are not necessarily symmetric in k. In general,

$$\begin{aligned}\gamma_{ij}^{k} &= \text{cov}(y_{it}, y_{jt-k}) \neq \text{cov}(y_{it}, y_{jt+k})\\ &= \text{cov}(y_{jt}, y_{it-k}) = \gamma_{ij}^{-k}\end{aligned}$$

Defn:

- If $\gamma_{ij}^{k} \neq 0$ for some $k > 0$ then y_{jt} is said to *lead* y_{it}.

- If $\gamma_{ij}^{-k} \neq 0$ for some $k > 0$ then y_{it} is said to *lead* y_{jt}.

- It is possible that y_{it} leads y_{jt} and vice-versa. In this case, there is said to be *feedback* between the two series.

All of the lag k cross covariances and correlations are summarized in the $(n \times n)$ lag k cross covariance and lag k cross correlation matrices

$$\mathbf{\Gamma}_k = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-k} - \boldsymbol{\mu})'] =$$

$$\begin{pmatrix} \mathrm{cov}(y_{1t}, y_{1t-k}) & \mathrm{cov}(y_{1t}, y_{2t-k}) & \cdots & \mathrm{cov}(y_{1t}, y_{nt-k}) \\ \mathrm{cov}(y_{2t}, y_{1t-k}) & \mathrm{cov}(y_{2t}, y_{2t-k}) & \cdots & \mathrm{cov}(y_{2t}, y_{nt-k}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(y_{nt}, y_{1t-k}) & \mathrm{cov}(y_{nt}, y_{2t-k}) & \cdots & \mathrm{cov}(y_{nt}, y_{nt-k}) \end{pmatrix}$$

$$\mathbf{R}_k = \mathbf{D}^{-1}\mathbf{\Gamma}_k\mathbf{D}^{-1}$$

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are not symmetric in k but it is easy to show that $\mathbf{\Gamma}_{-k} = \mathbf{\Gamma}_k'$ and $\mathbf{R}_{-k} = \mathbf{R}_k'$.

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ using

$$\begin{aligned} \hat{\mathbf{\Gamma}}_k &= \frac{1}{T} \sum_{t=k+1}^{T} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})' \\ \hat{\mathbf{R}}_k &= \hat{\mathbf{D}}^{-1}\hat{\mathbf{\Gamma}}_k\hat{\mathbf{D}}^{-1} \end{aligned}$$

Multivariate Wold Representation

Any $(n \times 1)$ covariance stationary multivariate time series $\mathbf{Y}_t$ has a Wold or linear process representation of the form

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k} \\ \boldsymbol{\Psi}_0 &= \mathbf{I}_n \\ \boldsymbol{\varepsilon}_t &\sim \mathsf{WN}(0, \boldsymbol{\Sigma}) \end{aligned}$$

$\boldsymbol{\Psi}_k$ is an $(n \times n)$ matrix with (i,j)th element ψ_{ij}^k. In lag operator notation, the Wold form is

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\Psi}(L) \boldsymbol{\varepsilon}_t \\ \boldsymbol{\Psi}(L) &= \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k L^k \end{aligned}$$

elements of $\boldsymbol{\Psi}(L)$ are 1-summable

The moments of $\mathbf{Y}_t$ are given by

$$E[\mathbf{Y}_t] = \boldsymbol{\mu}, \text{ var}(\mathbf{Y}_t) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_k'$$

Long Run Variance

Let $\mathbf{Y}_t$ be an $(n \times 1)$ stationary and ergodic multivariate time series with $E[\mathbf{Y}_t] = \boldsymbol{\mu}$. Anderson's central limit theorem for stationary and ergodic process states

$$\sqrt{T}(\bar{\mathbf{Y}} - \boldsymbol{\mu}) \xrightarrow{d} N\left(\mathbf{0}, \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j\right)$$

or

$$\bar{\mathbf{Y}} \overset{A}{\sim} N\left(\boldsymbol{\mu}, \frac{\mathbf{1}}{T} \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j\right)$$

Hence, the *long-run variance* of $\mathbf{Y}_t$ is T times the asymptotic variance of $\bar{\mathbf{Y}}$:

$$\mathsf{LRV}(\mathbf{Y}_t) = T \cdot \mathsf{avar}(\bar{\mathbf{Y}}) = \sum_{j=-\infty}^{\infty} \Gamma_j$$

Since $\mathbf{\Gamma}_{-j} = \mathbf{\Gamma}_j'$, LRV$(\mathbf{Y}_t)$ may be alternatively expressed as

$$\text{LRV}(\mathbf{Y}_t) = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty}(\mathbf{\Gamma}_j + \mathbf{\Gamma}_j')$$

Using the Wold representation of $\mathbf{Y}_t$ it can be shown that

$$\text{LRV}(\mathbf{Y}_t) = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)'$$

where $\mathbf{\Psi}(1) = \sum_{k=0}^{\infty} \mathbf{\Psi}_k$.

Non-parametric Estimate of the Long-Run Variance

A consistent estimate of $\mathsf{LRV}(\mathbf{Y}_t)$ may be computed using non-parametric methods. A popular estimator is the Newey-West weighted autocovariance estimator

$$\widehat{\mathsf{LRV}}_{\mathsf{NW}}(\mathbf{Y}_t) = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{M_T} w_{j,T} \cdot \left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'\right)$$

where $w_{j,T}$ are weights which sum to unity and M_T is a truncation lag parameter that satisfies $M_T = O(T^{1/3})$. Usually, the Bartlett weights are used

$$w_{j,T}^{Bartlett} = 1 - \frac{j}{M_T + 1}$$

Vector Autoregression Models

The *vector autoregression* (VAR) *model* is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series.

- Made famous in Chris Sims's paper "Macroeconomics and Reality," *ECTA* 1980.

- It is a natural extension of the univariate autoregressive model to dynamic multivariate time series.

- Has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting.

- It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models.

- Used for structural inference and policy analysis. In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model are summarized. These causal impacts are usually summarized with impulse response functions and forecast error variance

The Stationary Vector Autoregression Model

Let $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ denote an $(n \times 1)$ vector of time series variables. The basic p-lag *vector autoregressive* (VAR(p)) model has the form

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \mathbf{\Pi}_2 \mathbf{Y}_{t-2} + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t \\ & \boldsymbol{\varepsilon}_t \sim \text{WN}\ (\mathbf{0}, \mathbf{\Sigma}) \end{aligned}$$

Example: Bivariate VAR(2)

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi^1_{11} & \pi^1_{12} \\ \pi^1_{21} & \pi^1_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi^2_{11} & \pi^2_{12} \\ \pi^2_{21} & \pi^2_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

or

$$\begin{aligned} y_{1t} &= c_1 + \pi^1_{11} y_{1t-1} + \pi^1_{12} y_{2t-1} \\ &\quad + \pi^2_{11} y_{1t-2} + \pi^2_{12} y_{2t-2} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi^1_{21} y_{1t-1} + \pi^1_{22} y_{2t-1} \\ &\quad + \pi^2_{21} y_{1t-1} + \pi^2_{22} y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

where $\text{cov}(\varepsilon_{1t}, \varepsilon_{2s}) = \sigma_{12}$ for $t = s$; 0 otherwise.

Remarks:

- Each equation has the same regressors – lagged values of y_{1t} and y_{2t}.

- Endogeneity is avoided by using lagged values of y_{1t} and y_{2t}.

- The VAR(p) model is just a *seemingly unrelated regression* (SUR) model with lagged variables and deterministic terms as common regressors.

In lag operator notation, the VAR(p) is written as

$$\begin{aligned}\mathbf{\Pi}(L)\mathbf{Y}_t &= \mathbf{c} + \varepsilon_t \\ \mathbf{\Pi}(L) &= \mathbf{I}_n - \mathbf{\Pi}_1 L - ... - \mathbf{\Pi}_p L^p\end{aligned}$$

The VAR(p) is stable if the roots of

$$\det\left(\mathbf{I}_n - \mathbf{\Pi}_1 z - \cdots - \mathbf{\Pi}_p z^p\right) = 0$$

lie outside the complex unit circle (have modulus greater than one), or, equivalently, if the eigenvalues of the companion matrix

$$\mathbf{F} = \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_n \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_n & \mathbf{0} \end{pmatrix}$$

have modulus less than one. A stable VAR(p) process is stationary and ergodic with time invariant means, variances, and autocovariances.

Example: Stability of bivariate VAR(1) model

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{\Pi}\mathbf{Y}_{t-1} + \varepsilon_t \\
\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}
\end{aligned}
$$

Then det $(\mathbf{I}_n - \mathbf{\Pi}z) = 0$ becomes

$$
(1 - \pi_{11}z)(1 - \pi_{22}z) - \pi_{12}\pi_{21}z^2 = 0
$$

- Stability condition involves cross terms π_{12} and π_{21}

- If $\pi_{12} = \pi_{21} = 0$ (diagonal VAR) then bivariate stability condition reduces to univariate stability conditions for each equation.

If $\mathbf{Y}_t$ is covariance stationary, then the unconditional mean is given by

$$\boldsymbol{\mu} = (\mathbf{I}_n - \mathbf{\Pi}_1 - \cdots - \mathbf{\Pi}_p)^{-1}\mathbf{c}$$

The *mean-adjusted* form of the VAR(p) is then

$$\mathbf{Y}_t - \boldsymbol{\mu} = \mathbf{\Pi}_1(\mathbf{Y}_{t-1} - \boldsymbol{\mu}) + \mathbf{\Pi}_2(\mathbf{Y}_{t-2} - \boldsymbol{\mu}) + \cdots$$
$$+\mathbf{\Pi}_p(\mathbf{Y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t$$

The basic VAR(p) model may be too restrictive to represent sufficiently the main characteristics of the data. The general form of the VAR(p) model with deterministic terms and exogenous variables is given by

$$\mathbf{Y}_t = \mathbf{\Pi}_1\mathbf{Y}_{t-1} + \mathbf{\Pi}_2\mathbf{Y}_{t-2} + \cdots + \mathbf{\Pi}_p\mathbf{Y}_{t-p}$$
$$+\mathbf{\Phi}\mathbf{D}_t + \mathbf{G}\mathbf{X}_t + \boldsymbol{\varepsilon}_t$$
$$\mathbf{D}_t = \text{ deterministic terms}$$
$$\mathbf{X}_t = \text{ exogenous variables } (E[\mathbf{X}_t\boldsymbol{\varepsilon}_t] = \mathbf{0})$$

Wold Representation

Consider the stationary VAR(p) model

$$\begin{aligned}\mathbf{\Pi}(L)\mathbf{Y}_t &= \mathbf{c}+\boldsymbol{\varepsilon}_t \\ \mathbf{\Pi}(L) &= \mathbf{I}_n - \mathbf{\Pi}_1 L - \ldots - \mathbf{\Pi}_p L^p\end{aligned}$$

Since $\mathbf{Y}_t$ is stationary, $\mathbf{\Pi}(L)^{-1}$ exists so that

$$\mathbf{Y}_t = \mathbf{\Pi}(L)^{-1}\mathbf{c} + \mathbf{\Pi}(L)^{-1}\boldsymbol{\varepsilon}_t$$
$$= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \mathbf{\Psi}_k \varepsilon_{t-k}$$
$$\mathbf{\Psi}_0 = \mathbf{I}_n$$
$$\lim_{k\to\infty} \mathbf{\Psi}_k = 0$$

Note that

$$\mathbf{\Pi}(L)^{-1} = \mathbf{\Psi}(L) = \sum_{k=0}^{\infty} \mathbf{\Psi}_k L^k$$

The Wold coefficients $\boldsymbol{\Psi}_k$ may be determined from the VAR coefficients $\boldsymbol{\Pi}_k$ by solving

$$\boldsymbol{\Pi}(L)\boldsymbol{\Psi}(L) = \mathbf{I}_n$$

which implies

$$\begin{aligned}
\boldsymbol{\Psi}_1 &= \boldsymbol{\Pi}_1 \\
\boldsymbol{\Psi}_2 &= \boldsymbol{\Pi}_1\boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2 \\
&\vdots \\
\boldsymbol{\Psi}_s &= \boldsymbol{\Pi}_1\boldsymbol{\Psi}_{s-1} + \cdots + \boldsymbol{\Pi}_p\boldsymbol{\Psi}_{s-p}
\end{aligned}$$

Since $\boldsymbol{\Pi}(L)^{-1} = \boldsymbol{\Psi}(L)$, the long-run variance for $\mathbf{Y}_t$ has the form

$$\begin{gathered}
\mathsf{LRV}_{\mathsf{VAR}}(\mathbf{Y}_t) = \boldsymbol{\Psi}(1)\boldsymbol{\Sigma}\boldsymbol{\Psi}(1)' \\
= \boldsymbol{\Pi}(1)^{-1}\boldsymbol{\Sigma}\boldsymbol{\Pi}(1)^{-1\prime} \\
= (\mathbf{I}_n - \boldsymbol{\Pi}_1 - ... - \boldsymbol{\Pi}_p)^{-1}\,\boldsymbol{\Sigma}\,(\mathbf{I}_n - \boldsymbol{\Pi}_1 - ... - \boldsymbol{\Pi}_p)^{-1\prime}
\end{gathered}$$

Estimation

Assume that the VAR(p) model is covariance stationary, and there are no restrictions on the parameters of the model. In SUR notation, each equation in the VAR(p) may be written as

$$\mathbf{y}_i = \mathbf{Z}\boldsymbol{\pi}_i + \mathbf{e}_i, \; i = 1, \ldots, n$$

- $\mathbf{y}_i$ is a $(T \times 1)$ vector of observations on the i^{th} equation

- $\mathbf{Z}$ is a $(T \times k)$ matrix with t^{th} row given by $\mathbf{Z}_t' = (1, \mathbf{Y}_{t-1}', \ldots, \mathbf{Y}_{t-p}')$

- $k = np + 1$

- $\boldsymbol{\pi}_i$ is a $(k \times 1)$ vector of parameters and e_i is a $(T \times 1)$ error with covariance matrix $\sigma_i^2 \mathbf{I}_T$.

Since the VAR(p) is in the form of a SUR model where each equation has the same explanatory variables, each equation may be estimated separately by ordinary least squares without losing efficiency relative to generalized least squares. Let

$$\hat{\mathbf{\Pi}} = [\hat{\boldsymbol{\pi}}_1, \ldots, \hat{\boldsymbol{\pi}}_n]$$

denote the $(k \times n)$ matrix of least squares coefficients for the n equations.Let

$$\text{vec}(\hat{\mathbf{\Pi}}) = \begin{pmatrix} \hat{\boldsymbol{\pi}}_1 \\ \vdots \\ \hat{\boldsymbol{\pi}}_n \end{pmatrix}$$

Under standard assumptions regarding the behavior of stationary and ergodic VAR models (see Hamilton (1994) vec($\hat{\mathbf{\Pi}}$) is consistent and asymptotically normally distributed with asymptotic covariance matrix

$$\widehat{\text{avar}}(\text{vec}(\hat{\mathbf{\Pi}})) = \hat{\mathbf{\Sigma}} \otimes (\mathbf{Z}'\mathbf{Z})^{-1}$$

where

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= \frac{1}{T-k} \sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' \\ \hat{\boldsymbol{\varepsilon}}_t &= \mathbf{Y}_t - \hat{\mathbf{\Pi}}' \mathbf{Z}_t \end{aligned}$$

Lag Length Selection

The lag length for the VAR(p) model may be determined using model selection criteria. The general approach is to fit VAR(p) models with orders $p = 0, ..., p_{max}$ and choose the value of p which minimizes some model selection criteria

$$\begin{aligned}
\text{MSC}(p) &= \ln|\tilde{\mathbf{\Sigma}}(p)| + c_T \cdot \varphi(n,p) \\
\tilde{\mathbf{\Sigma}}(p) &= T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t' \\
c_T &= \text{ function of sample size} \\
\varphi(n,p) &= \text{penalty function}
\end{aligned}$$

The three most common information criteria are the Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ):

$$\begin{aligned}
\text{AIC}(p) &= \ln|\tilde{\mathbf{\Sigma}}(p)| + \frac{2}{T}pn^2 \\
\text{BIC}(p) &= \ln|\tilde{\mathbf{\Sigma}}(p)| + \frac{\ln T}{T}pn^2 \\
\text{HQ}(p) &= \ln|\tilde{\mathbf{\Sigma}}(p)| + \frac{2\ln\ln T}{T}pn^2
\end{aligned}$$

Remarks:

- AIC criterion asymptotically overestimates the order with positive probability,
- BIC and HQ criteria estimate the order consistently under fairly general conditions if the true order p is less than or equal to p_{max}.

Granger Causality

One of the main uses of VAR models is forecasting. The following intuitive notion of a variable's forecasting ability is due to Granger (1969).

- If a variable, or group of variables, y_1 is found to be helpful for predicting another variable, or group of variables, y_2 then y_1 is said to *Granger-cause* y_2; otherwise it is said to *fail to Granger-cause* y_2.

- Formally, y_1 fails to Granger-cause y_2 if for all $s > 0$ the MSE of a forecast of $y_{2,t+s}$ based on $(y_{2,t}, y_{2,t-1}, \ldots)$ is the same as the MSE of a forecast of $y_{2,t+s}$ based on $(y_{2,t}, y_{2,t-1}, \ldots)$ and $(y_{1,t}, y_{1,t-1}, \ldots)$.

- The notion of Granger causality does not imply true causality. It only implies forecasting ability.

Example: Bivariate VAR Model

In a bivariate VAR(p) , y_2 fails to Granger-cause y_1 if all of the p VAR coefficient matrices $\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_p$ are lower triangular:

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & 0 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \cdots \\ &+ \begin{pmatrix} \pi_{11}^p & 0 \\ \pi_{21}^p & \pi_{22}^p \end{pmatrix} \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \end{aligned}$$

Similarly, y_1 fails to Granger-cause y_2 if all of the coefficients on lagged values of y_1 are zero in the equation for y_2. Notice that if y_2 fails to Granger-cause y_1 and y_1 fails to Granger-cause y_2, then the VAR coefficient matrices $\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_p$ are diagonal.

The p linear coefficient restrictions implied by Granger non-causality may be tested using the Wald statistic

$$\begin{aligned}\text{Wald} &= (\mathbf{R}\cdot\text{vec}(\hat{\mathbf{\Pi}}) - \mathbf{r})' \left\{\mathbf{R}\left[\widehat{\text{avar}}(\text{vec}(\hat{\mathbf{\Pi}}))\right]\mathbf{R}'\right\}^{-1} \\ &\quad \times(\mathbf{R}\cdot\text{vec}(\hat{\mathbf{\Pi}}) - \mathbf{r})\end{aligned}$$

Remark: In the bivariate model, testing $H_0 : y_2$ does not Granger-cause y_1 reduces to a testing $H_0 : \pi_{12}^1 = \cdots = \pi_{12}^p = 0$ from the linear regression

$$\begin{aligned}y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \cdots + \pi_{11}^p y_{1t-p} \\ &\quad + \pi_{12}^1 y_{2t-1} + \cdots + \pi_{12}^p y_{2t-p} + \varepsilon_{1t}\end{aligned}$$

The test statistic is a simple F-statistic.

Example: Trivariate VAR Model

In a trivariate VAR(p) , y_2 and y_3 fail to Granger-cause y_1 if $\pi_{12}^{j} = \pi_{13}^{j} = 0$ for all j:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \pi_{11}^{1} & 0 & 0 \\ \pi_{21}^{1} & \pi_{22}^{1} & \pi_{23}^{1} \\ \pi_{31}^{1} & \pi_{32}^{1} & \pi_{33}^{1} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \pi_{11}^{p} & 0 & 0 \\ \pi_{21}^{p} & \pi_{22}^{p} & \pi_{32}^{p} \\ \pi_{31}^{p} & \pi_{32}^{p} & \pi_{33}^{p} \end{pmatrix} \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ y_{3t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

Note: One can also use a simple F-statistic from a linear regression in this situation:

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^{1} y_{1t-1} + \cdots + \pi_{11}^{p} y_{1t-p} \\ &\quad + \pi_{12}^{1} y_{2t-1} + \cdots + \pi_{12}^{p} y_{2t-p} \\ &\quad + \pi_{13}^{1} y_{3t-1} + \cdots + \pi_{13}^{p} y_{3t-p} + \varepsilon_{1t} \end{aligned}$$

Example: Trivariate VAR Model

In a trivariate VAR(p) , y_3 fails to Granger-cause y_1 and y_2 if all of the p VAR coefficient matrices $\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_p$ are lower triangular:

$$\begin{aligned}\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} &= \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & 0 & 0 \\ \pi_{21}^1 & \pi_{22}^1 & 0 \\ \pi_{31}^1 & \pi_{32}^1 & \pi_{33}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \cdots \\ &+ \begin{pmatrix} \pi_{11}^p & 0 & 0 \\ \pi_{21}^p & \pi_{22}^p & 0 \\ \pi_{31}^p & \pi_{32}^p & \pi_{33}^p \end{pmatrix} \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ y_{3t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}\end{aligned}$$

Note: We cannot use a simple F-statistic in this case. We must use the general Wald statistic.

Forecasting Algorithms

Forecasting from a VAR(p) is a straightforward extension of forecasting from an AR(p). The multivariate Wold form is

$$\begin{aligned}
\mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\
\mathbf{Y}_{t+h} &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+h-1} + \cdots \\
&\quad + \boldsymbol{\Psi}_{h-1} \boldsymbol{\varepsilon}_{t+1} + \boldsymbol{\Psi}_h \boldsymbol{\varepsilon}_t + \cdots \\
&\quad \boldsymbol{\varepsilon}_t \sim \text{WN}(\mathbf{0}, \boldsymbol{\Sigma})
\end{aligned}$$

Note that

$$\begin{aligned}
E[\mathbf{Y}_t] &= \boldsymbol{\mu} \\
\text{var}(\mathbf{Y}_t) &= E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})'] \\
&= E\left[\left(\sum_{k=1}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}\right)\left(\sum_{k=1}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}\right)'\right] \\
&= \sum_{k=1}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_k'
\end{aligned}$$

The minimum MSE linear forecast of $\mathbf{Y}_{t+h}$ based on I_t is

$$\mathbf{Y}_{t+h|t} = \boldsymbol{\mu} + \boldsymbol{\Psi}_h \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_{h+1} \boldsymbol{\varepsilon}_{t-1} + \cdots$$

The forecast error is

$$\begin{aligned} \boldsymbol{\varepsilon}_{t+h|t} &= \mathbf{Y}_{t+h} - \mathbf{Y}_{t+h|t} \\ &= \boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+h-1} + \cdots + \boldsymbol{\Psi}_{h-1} \boldsymbol{\varepsilon}_{t+1} \end{aligned}$$

The forecast error MSE is

$$\begin{aligned} \text{MSE}(\boldsymbol{\varepsilon}_{t+h|t}) &= E[\boldsymbol{\varepsilon}_{t+h|t} \boldsymbol{\varepsilon}'_{t+h|t}] \\ &= \boldsymbol{\Sigma} + \boldsymbol{\Psi}_1 \boldsymbol{\Sigma} \boldsymbol{\Psi}'_1 + \cdots + \boldsymbol{\Psi}_{h-1} \boldsymbol{\Sigma} \boldsymbol{\Psi}'_{h-1} \\ &= \sum_{s=1}^{h-1} \boldsymbol{\Psi}_s \boldsymbol{\Sigma} \boldsymbol{\Psi}'_s \end{aligned}$$

Chain-rule of Forecasting

The best linear predictor, in terms of minimum mean squared error (MSE), of $\mathbf{Y}_{t+1}$ or 1-step forecast based on information available at time T is

$$\mathbf{Y}_{T+1|T} = \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_T + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{T-p+1}$$

Forecasts for longer horizons h (h-step forecasts) may be obtained using the *chain-rule of forecasting* as

$$\begin{aligned} \mathbf{Y}_{T+h|T} &= \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_{T+h-1|T} + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{T+h-p|T} \\ \mathbf{Y}_{T+j|T} &= \mathbf{Y}_{T+j} \text{ for } j \leq 0. \end{aligned}$$

Note: Chain-rule may be derived from state-space framework (assume $\mathbf{c} = 0$)

$$\begin{aligned} \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p+1} \end{pmatrix} &= \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_p \\ \mathbf{I}_n & \mathbf{0} & \vdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-p+1} \end{pmatrix} \\ &\quad + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \\ \boldsymbol{\xi}_t &= \mathbf{F} \boldsymbol{\xi}_{t-1} + \mathbf{v}_t \end{aligned}$$

Then

$$
\begin{aligned}
\boldsymbol{\xi}_{t+1|t} &= \mathbf{F}\boldsymbol{\xi}_t \\
\boldsymbol{\xi}_{t+2|t} &= \mathbf{F}\boldsymbol{\xi}_{t+1|t} = \mathbf{F}^2\boldsymbol{\xi}_t \\
&\vdots \\
\boldsymbol{\xi}_{t+h|t} &= \mathbf{F}\boldsymbol{\xi}_{t+h-1|t} = \mathbf{F}^h\boldsymbol{\xi}_t
\end{aligned}
$$