Time Series Analysis

Vector Autoregressive Model

Binzhi Chen

Department of Economics
University of Birmingham

March 4, 2020

Introduction

In previous lectures, we go through time series analysis based on one variable and its lags, which is univariate time series.

However, univariate analysis is insufficient to capture the complex dynamics among interrelated time series. Sometimes one variable will depend on its own lags and other variables' lags, which is multivariate time seires.

The vector autoregression (VAR) is the standard model used to model multiple stationary time-series.

VAR(P)

VAR is similar to AR process. A general VAR P_{th} order model (Which is also called reduced form VAR) can be represented as:

$$\mathbf{Y_t} = \mathbf{\Phi_0} + \mathbf{\Phi_1 Y_{t-1}} + \mathbf{\Phi_2 Y_{t-2}} + \cdots + \mathbf{\Phi_p Y_{t-p}} + \boldsymbol{\epsilon_t}$$

Where $\mathbf{Y_t}$ is a k by 1 vector stochastic process, $\mathbf{\Phi_0}$ is the vector of intercept, $\mathbf{\Phi_j}$ is a $k \times k$ vector of coefficient and $\boldsymbol{\epsilon_t}$ is the white noise.

The vector of white noise satisfies:

$$E(\boldsymbol{\epsilon_t}) = \mathbf{0}$$
$$E(\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t'}) = \mathbf{\Sigma}$$
$$E(\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_{t-s}'}) = \mathbf{0}$$

Where $\mathbf{\Sigma}$ is a symmetric positive define matrix, which implies that error terms could be correlated at the same time period.

VAR(P)

Simplest case is a bivariate VAR(1):

$$\mathbf{Y_t} = \mathbf{\Phi_0} + \mathbf{\Phi_1 Y_{t-1}} + \boldsymbol{\epsilon_t}$$

Or:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi^1_{11} & \phi^1_{12} \\ \phi^1_{21} & \phi^1_{22} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Or:

$$Y_{1t} = \phi_{10} + \phi^1_{11} Y_{1t-1} + \phi^1_{12} Y_{2t-1} + \epsilon_{1t}$$
$$Y_{2t} = \phi_{20} + \phi^1_{21} Y_{1t-1} + \phi^1_{22} Y_{2t-1} + \epsilon_{2t}$$

VAR(P)

Advantages of VAR Modelling:

1- Do not need to specify which variables are endogenous or exogenous - all are endogenous

2- Allows the value of a variable to depend on more than just its own lags or combinations of white noise terms, so more general than ARMA modelling

3- Provided that there are no contemporaneous terms on the right hand side of the equations, can simply use OLS separately on each equation

4- Forecasts are often better than traditional structural" models.

Disadvantages of VAR Modelling:

1- VAR's are a-theoretical (as are ARMA models)

VAR(P)

Recall lag operator can also be used in VAR process:

$$(\mathbf{I} - \mathbf{\Phi_1 L} - \mathbf{\Phi_2 L^2} - \cdots - \mathbf{\Phi_p L^p})\mathbf{Y_t} = \mathbf{\Phi_0} + \boldsymbol{\epsilon_t}$$

The stationary condition for VAR process are defined as follows,

the VAR process is stationary if the roots of:

$$\mathbf{det}(\mathbf{I_k} - \mathbf{\Phi_1 Z} - \mathbf{\Phi_2 Z^2} - \cdots - \mathbf{\Phi_p Z^p})$$

all lie outside the unit circle.

VAR(1)

For simplicity, we will examine the expectation and variance of VAR(1):

$$\mathbf{Y_t} = \mathbf{\Phi_0} + \mathbf{\Phi_1 Y_{t-1}} + \boldsymbol{\epsilon_t}$$

If we assume the process is stationary, we will get:

$$\begin{aligned}
(\mathbf{I} - \mathbf{\Phi_1 L})\mathbf{Y_t} &= \mathbf{\Phi_0} + \boldsymbol{\epsilon_t} \\
\mathbf{Y_t} &= (\mathbf{I} - \mathbf{\Phi_1 L})^{-1}\mathbf{\Phi_0} + (\mathbf{I} - \mathbf{\Phi_1 L})^{-1}\boldsymbol{\epsilon_t} \\
\mathbf{Y_t} &= (\mathbf{I} - \mathbf{\Phi_1 L})^{-1}\mathbf{\Phi_0} + \sum_{i=0}^{\infty} \mathbf{\Phi_1^i}\boldsymbol{\epsilon_{t-i}}
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{E}(\mathbf{Y_t}) &= \mathbf{E}[(\mathbf{I} - \mathbf{\Phi_1 L})^{-1}\mathbf{\Phi_0} + \sum_{i=0}^{\infty} \mathbf{\Phi_1^i}\boldsymbol{\epsilon_{t-i}}] \\
&= (\mathbf{I} - \mathbf{\Phi_1 L})^{-1}\mathbf{\Phi_0}
\end{aligned}$$

VAR(1)

The variance of VAR(1) process is:

$$\begin{aligned}
\mathbf{Var}(\mathbf{Y_t}) &= \mathbf{E}[(\mathbf{Y_t} - \mathbf{E}(\mathbf{Y_t}))(\mathbf{Y_t} - \mathbf{E}(\mathbf{Y_t}))'] \\
&= \mathbf{E}[(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i})(\sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i})'] \\
&= \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \mathbf{E}[\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}'](\mathbf{\Phi}_1')^i \\
&= \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \mathbf{\Sigma}(\mathbf{\Phi}_1')^i
\end{aligned}$$

Or you can express as :

$$vec(\mathbf{Var}(\mathbf{Y_t})) = (\mathbf{I_{k^2}} - \mathbf{\Phi_1} \otimes \mathbf{\Phi_1})^{-1} vec(\mathbf{\Sigma})$$

Where *vec* operator znd a Kronecker $\otimes$ product stack the elements of a matrix product into a single column.

VAR(1)

If $\mathbf{A}$ is an $(m \times n)$ matrix, then $vec(A)$ is an $(mn \times 1)$ column vector, obtained bt stacking the columns of A. For example:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Then,

$$vec(\mathbf{A}) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \end{bmatrix}$$

VAR(1)

Let $\mathbf{A}$ be an m by n matrix, and let $\mathbf{B}$ be a k by l matrix. The Kronecker product is defined as:

$$\mathbf{A} \otimes \mathbf{A} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nn}\mathbf{B} \end{bmatrix}$$

and has dimension mk by nl.

Then we will have:

$$vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})vec(\mathbf{B})$$

VAR Companion Form

Any stationary VAR(P) can be rewritten as a VAR(1):

$$\xi_t = \mathbf{F}\xi_{t-1} + \mathbf{v_t}$$

Where

$$\xi_t = \begin{bmatrix} \mathbf{Y_t} - \mu \\ \mathbf{Y_{t-1}} - \mu \\ \vdots \\ \mathbf{Y_{t-p+1}} - \mu \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{n-1} & \Phi_n \\ \mathbf{I} & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{I} & 0 \end{bmatrix} \quad \mathbf{v_t} = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Order Selection Method

Since a VAR process would create many parameters, it is essential to decide the lag length.

For example,If we have n equations for n variables and we have p lags of each of the variables in each equation, we have to estimate $(n + pn^2)$ parameters. e.g. $n = 3$, $p = 3$, parameters=30

In general, we use Likelihood Ratio Test and Information Cretia to select the best lag length.

Order Selection Method

Suppose that a bivariate VAR(8) estimated using quarterly data has 8 lags of the two variables in each equation(unrestricted model $p_1 = 8$), and we want to examine a restriction that the coefficients on lags 5 through 8 are jointly zero(restricted model $p_0 = 4$). This can be done using a likelihood ratio test where the test is the null hypothesis of the lag of restricted model against the alternative hypothesis of the lag of unrestricted model

Disadvantages: Conducting the LR test is cumbersome and requires a normality assumption for the disturbances.

Order Selection Method

Denote the variance-covariance matrix of residuals as Σ. The likelihood ratio test for this joint hypothesis is given by:

$$LR = (T - c)[log|\hat{\Sigma}_{\mathbf{0}}| - log|\hat{\Sigma}_{\mathbf{1}}|]$$

Where T is the sample size, $c = 1 + np_1$ is the number of parameters estimated per equation under the alternative.

$|\Sigma_{\mathbf{0}}|$ is the variance-covariance matrix of the residuals for the restricted model (with 4 lags),$|\Sigma_{\mathbf{1}}|$ is the variance-covariance matrix of the residuals for the restricted model (with 8 lags).

The test statistic is asymptotically distributed as a $\chi^2_{[n^2(p_1-p_0)]}$

Order Selection Method

Multivariate versions of the information criteria are required. These can be defined as:

$$\text{MAIC} = log(|\hat{\Sigma}|) + \frac{2k}{T}$$

$$\text{MSBIC} = log(|\hat{\Sigma}|) + \frac{k}{T} log(T)$$

$$\text{MHQIC} = log(|\hat{\Sigma}|) + \frac{2k}{T} log(log(T))$$

Where $k = n + pn^2$ is the total number of regressors in all equations

Standard VAR Form

Recall the *Reduced-form bivariate VAR(1)*:

$$\mathbf{Y_t} = \mathbf{\Phi_0} + \mathbf{\Phi_1 Y_{t-1}} + \boldsymbol{\epsilon_t}$$

The reason behind the name is because it does not show explicitly the concurrent dependence between the component series.

If necessary, an explicit expression involving the concurrent relationship can be deduced from the reduced form model by a simple linear transformation, which involves a contemporaneous term.

Since $\mathbf{\Sigma}$ is positive definite, according to Cholesky Decomposition, there exists a lower triangular matrix $\mathbf{L}$ with unit diagonal elements and a diagonal matrix $\mathbf{G}$ such that:

$$\mathbf{\Sigma} = \mathbf{LGL}'$$

Hence

$$\mathbf{L}^{-1}\mathbf{\Sigma}(\mathbf{L}')^{-1} = \mathbf{G}$$

Standard VAR Form

Define:

$$\mathbf{v_t} = \mathbf{L}^{-1}\epsilon_t$$

Thus, we have:

$$\mathbf{E}(\mathbf{v_t}) = \mathbf{L}^{-1}\mathbf{E}\epsilon_t = 0$$

$$\mathbf{Cov}(\mathbf{v_t}) = \mathbf{L}^{-1}\Sigma(\mathbf{L}')^{-1} = \mathbf{G}$$

Multiplying $\mathbf{L}^{-1}$ into the reduced-form bivariate VAR(1) we get:

$$\begin{aligned}\mathbf{L}^{-1}\mathbf{Y_t} &= \mathbf{L}^{-1}\mathbf{\Phi_0} + \mathbf{L}^{-1}\mathbf{\Phi_1}\mathbf{Y_{t-1}} + \mathbf{L}^{-1}\epsilon_t \\ &= \mathbf{\Phi_0^*} + \mathbf{\Phi_1^*}\mathbf{Y_{t-1}} + \mathbf{v_t}\end{aligned}$$

Standard VAR Form

To illustrate the transformation from a reduced-form model to structural equations, consider the bivariate VAR(1) model:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} + \begin{bmatrix} +0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \qquad \mathbf{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

The inverse of the lower triangular matrix is:

$$\mathbf{L^{-1}} = \begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix}$$

Standard VAR Form

Multiplying $\mathbf{L^{-1}}$ into the reduced-form bivariate VAR(1) we get:

$$\begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix} + \begin{bmatrix} +0.2 & 0.3 \\ -0.7 & 0.95 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

Where

$$\mathbf{G} = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Thus, we get:

$$Y_{2t} = 0.3 + 0.5\,Y_{1t} - 0.7\,Y_{1t-1} + 0.95\,Y_{2t-1} + v_{2t}$$

We can get the result fot Y_{1t} by rearranging the order of elements in $\mathbf{Y_t}$

Granger Causality

A scalar random variable x_t does not Granger cause y_t if:

$$E[y_t|x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2} \dots] = E[y_t|y_{t-1}, y_{t-2} \dots]$$

That is, x_t does not Granger cause y_t if the forecast of y_t is the same whether conditioned on past values of x_t or not.

Equivalently, we may say x_t is not linearly informative about future y_t. Or y_t is exogenous in the time series with respect to x_t

Granger Causality

In a bivariate VAR(2) process:

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & \phi_{12}^1 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11}^2 & \phi_{12}^2 \\ \phi_{21}^2 & \phi_{22}^2 \end{bmatrix} \begin{bmatrix} Y_{1t-2} \\ Y_{2t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

If ϕ_{12}^1 and ϕ_{12}^2 equal 0, then Y_{2t} is not Granger cause Y_{1t}

We may set:

$$H_0 : \phi_{12}^1 = \phi_{12}^2 = 0$$

as the null hypothesis.
The joint hypotheses can be tested within the F-test framework, which could also be referred to as Granger causality tests.

If both sets of lags are significant, there is "bi-directional causality"

Impulse Response Functions and Variance Decompositions

VAR models are often difficult to interpret:

Granger Causality could only show that "if" x_t has a forecast relationship with y_t. But it does not show "how much" x_t will have a forecast influence on y_t.

In this case, we will introduce two methods:
Impulse Response Functions and Variance Decompositions

Impulse Response Functions

Impulse responses trace out the responsiveness of the dependent variables in the VAR to shocks to the error term. 'A unit shock is applied to each variable and its effects are noted.

Consider for example a simple bivariate VAR(1):

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11}^1 & \phi_{12}^1 \\ \phi_{21}^1 & \phi_{22}^1 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

A change in ϵ_{1t} will immediately change Y_{1t}. It will change change Y_{2t} and also Y_{1t} during the next period.

We can examine how long and to what degree a shock to a given equation has on all of the variables in the system.

Impulse Response Functions

The main problem for this is that above, we assumed that the VAR error terms were statistically independent of one another.

This is generally not true, however, the error terms will typically be correlated to some degree.

Therefore, the notion of examining the effect of the innovations separately has little meaning, since they have a common component.

What is done is to "orthogonalise" the innovations.

In the bivariate VAR, this problem would be approached by attributing all of the effect of the common component to the rst of the two variables in the VAR.

In the general case where there are more variables, the situation is more complex but the interpretation is the same.

Variance Decompositions

Variance decompositions offer a slightly different method of examining VAR dynamics. They give the proportion of the movements in the dependent variables that are due to their "own" shocks, versus shocks to the other variables.

This is done by determining how much of the s-step ahead forecast error variance for each variable is explained innovations to each explanatory variable

The variance decomposition gives information about the relative importance of each shock to the variables in the VAR.