Academic Year 2019-2020

Time Series Econometrics

Fabrizio Iacone

Chapter 12: Vector Autoregressions

Topics:
Vector White Noise, Vector ARMA,
Inference for VARs,
Granger- causality,
Impulse Response Function
Structuralised IRF
Forecast error variance decomposition

We are interested in the **vector time series**

$$Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \ldots \\ Y_{n,t} \end{bmatrix}$$

This is a vector of dimension n.

As we did for ARMA models, we we begin with a **vector white noise**.

Denote

$$\epsilon_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \ldots \\ \epsilon_{p,t} \end{bmatrix}$$

then by saying that ϵ is a vector white noise we mean that

$$E(\epsilon_t) = 0$$
$$E(\epsilon_t \ \epsilon_t') = \Omega$$
$$E(\epsilon_t \ \epsilon_\tau') = 0 \text{ if } t \neq \tau$$

where Ω is a symmetric, positive definite matrix.

Then, for $n \times n$ matrices Φ_1, Φ_2, ..., Φ_p and Θ_1, Θ_2, ..., Θ_n, we define Vector ARMA (VARMA)

$$Y_t - \Phi_1 Y_{t-1} - ... - \Phi_p Y_{t-p} = \epsilon_t + \Theta_1 \epsilon_{t-1} + ... + \Theta_q \epsilon_{t-q},$$

Special cases will include
Vector Autoregressions (VAR)

$$Y_t - \Phi_1 Y_{t-1} - ... - \Phi_p Y_{t-p} = \epsilon_t$$

and vector moving averages

$$Y_t = \epsilon_t + \Theta_1 \epsilon_{t-1} + ... + \Theta_q \epsilon_{t-q}$$

Stationarity and **invertibility** follow adapting conditions from the scalar time series case.

Adapting arguments as we saw for scalar time series, we can also write any stationary VARMA as infinite moving average

$$Y_t = \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + ...$$

Example of notation.

For the generic matrix Φ_j denote element in row a and column b as $\phi_{ab}^{(j)}$. For example, for $n = 2$,

$$Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix}, \ \epsilon_t = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \ \Phi_1 = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix}$$

so the VAR(1) $Y_t = \Phi_1 Y_{t-1} + \epsilon_t$ is

$$\begin{aligned} Y_{1,t} &= \phi_{11}^{(1)} Y_{1,t-1} + \phi_{12}^{(1)} Y_{2,t-1} + \epsilon_{1,t} \\ Y_{2,t} &= \phi_{21}^{(1)} Y_{1,t-1} + \phi_{22}^{(1)} Y_{2,t-1} + \epsilon_{2,t} \end{aligned}$$

with infinite Moving Average representation

$$\begin{aligned} Y_{1,t} &= \epsilon_{1,t} + \psi_{11}^{(1)} \epsilon_{1,t-1} + \psi_{12}^{(1)} \epsilon_{2,t-1} \\ &\quad + \psi_{11}^{(2)} \epsilon_{1,t-2} + \psi_{12}^{(2)} \epsilon_{2,t-2} + \ldots \end{aligned}$$

$$\begin{aligned} Y_{2,t} &= \epsilon_{2,t} + \psi_{21}^{(1)} \epsilon_{1,t-1} + \psi_{22}^{(1)} \epsilon_{2,t-1} \\ &\quad + \psi_{21}^{(2)} \epsilon_{1,t-2} + \psi_{22}^{(2)} \epsilon_{2,t-2} + \ldots \end{aligned}$$

Note: Hamilton has a slightly different notation, as he calls $Y_{1,t}$ and $Y_{2,t}$ as y_t and x_t, respectively.

Inference for VARs

★ **Estimation** is done using the same techniques that we saw for scalar time series. In particular, Conditional ML can be done using OLS (when no restrictions are imposed on elements of the matrices Φ_j)

★ Estimates are **consistent** and **asymptotically normal** (rate of convergence is $\sqrt{T}$, as in the scalar case).

★ **Model selection** can be done using information criteria as in the scalar case. However, since the VARs models are nested, it is also possible to test exclusion of lags as exclusion restrictions.

★ **Model validation** can be done using the Portmanteau test (a multivariate version of it). However, in view of the cost in terms of degrees of freedom (which depends on the number of parameters) the Lagrange Multiplier (LM) test for no autocorrelation of the residuals might be preferred.

Example: using the Likelihood Ratio to select the lags

Since the VARs models are nested, it is also possible to test exclusion of lags as exclusion restrictions. This is usually done by means of a Likelihood Ratio test.

For example, estimating a VAR(2) we have estimates $\widehat{\Phi}_1$ and $\widehat{\Phi}_2$.
We could test the hypothesis that one lag is sufficient by testing $H_0 : \{\Phi_2 = 0\}$.
There are $n \times n$ parameters in Φ_2, so this corresponds to testing n^2 hypotheses). The Likelihood Ratio test is particularly convenient to test this joint hypothesis, because the OLS estimates $\widehat{\Phi}_1$ and $\widehat{\Phi}_2$ are asymptotically equivalent to the ML ones, and the Concentrated likelihood can be computed directly from the miminised Residuals.

Granger Causality

An interesting application of inference on coefficients in the VAR representation is the Granger-causality test.

In a bivariate ($n = 2$) case, by saying that for example, $y_{2,t}$ Granger-causes $y_{1,t}$, we mean that past values of $y_{2,t}$ help predicting $y_{1,t}$ (the reverse might be equally considered).

If $Y_{2,t}$ does not Granger-causes $Y_{1,t}$, the matrices Φ_j and Ψ_j are lower triangular.

For example, for a VAR(2),

$$\begin{aligned} Y_{1,t} &= \phi_{11}^{(1)} Y_{1,t-1} + \phi_{11}^{(2)} Y_{1,t-2} + \epsilon_{1,t} \\ Y_{2,t} &= \phi_{21}^{(1)} Y_{1,t-1} + \phi_{22}^{(1)} Y_{2,t-1} \\ &\quad + \phi_{21}^{(2)} Y_{1,t-2} + \phi_{22}^{(2)} Y_{2,t-2} + \epsilon_{2,t} \end{aligned}$$

Thus we test for Granger-causality by testing $H_0 : \{\phi_{12}^{(1)} = 0, \phi_{12}^{(2)} = 0\}$. This is easy to do, as the distribution of $\widehat{\Phi}_1$, $\widehat{\Phi}_2$ is easy to use.

Granger Causality test.
Estimate the restricted regression

$$Y_{1,t} = \phi_{11}^{(1)} Y_{1,t-1} + \phi_{11}^{(2)} Y_{1,t-2} + e_t$$

and the unrestricted regression

$$\begin{aligned} Y_{1,t} &= \phi_{11}^{(1)} Y_{1,t-1} + \phi_{11}^{(2)} Y_{1,t-2} \\ &\quad + \phi_{12}^{(2)} Y_{2,t-1} + \phi_{12}^{(2)} Y_{2,t-2} + u_t \end{aligned}$$

Denote $\hat{e}_t$ and $\hat{u}_t$ as the residuals of these two regressions, and

$$RRSS = \sum_{t=2+1}^{T} \hat{e}_t^2, \; RRSU = \sum_{t=2+1}^{T} \hat{u}_t^2$$

Then, under H_0,

$$T \frac{RRSS - RRSU}{RSSU} \rightarrow_d \chi_2^2$$

For a VAR(p) restricted and unrestricted regressions are run on p lags, and $T \frac{RRSS - RRSU}{RSSU} \rightarrow_d \chi_p^2$.

★ other approached to testing for Granger causality are possible.

★ This is not really a causality test (Granger-causality is not really causality). Even if we find, for example, that $Y_{2,t}$ Granger-causes $Y_{1,t+j}$, this may well be that it is actually $Y_{1,t+j}$ causing the move in $Y_{2,t}$. One typical example is with price of shares: suppose that the expected future dividends increase: prices of shares should take this into account immediately. As long as the expectations are on average correct, we then observe that on average dividends in the future do indeed increase. Thus, we observed higher prices today anticipate higher dividends in the future. However, it is not the prices that caused the dividends, but rather the opposite. This is often the case when Rational Expectations are considered.

Impulse Response Function: an Identification Problem

In the scalar case we defined the Impulse Response Function as the plot of ψ_j against j.
In the multivariate case we could consider the plot of Ψ_j against j.
However, the interpretation is not always clear.

Consider $n = 2$ again, then for example

$$\psi_{11}^{(s)} = \frac{\partial Y_{1,t+s}}{\partial \epsilon_{1,t}}$$

In analogy with the interpretation in a multivariate regression, $\psi_{11}^{(s)}$ then gives the change in prediction of $Y_{1,t+s}$ due to a change in $\epsilon_{1,t}$ holding everything else constant.
However, **since $\epsilon_{1,t}$ is correlated with $\epsilon_{2,t}$, the joint effect of these should be considered**, to appreciate how our prediction of $Y_{1,t+s}$ is affected by a shock in $\epsilon_{1,t}$.

Consider a nonsingular, $n \times n$ matrix H, and

$$u_t = H\epsilon_t$$

Then, u_t is a vector white noise process and

$$Y_t = H^{-1}\ H\epsilon_t + \Psi_1 H^{-1}\ H\epsilon_{t-1} + \Psi_2 H^{-1}\ H\epsilon_{t-2} + ...$$

So letting

$$J_s = \Psi_s H^{-1}$$

we have

$$Y_t = J_0 u_t + J_1 u_{t-1} + J_2 u_{t-2} + ...$$

This shows that **there are alternative representations based on Vector White Noise processes different from** ϵ_t

The covariance matrix of u_t is $E(u_t\ u_t') = H\Omega H'$. We may choose H so that $E(u_t\ u_t')$ is diagonal:

Cholesky decomoposition.
For any real and symmetric matrix Ω there is a diagonal matrix D and a lower triangular matrix A with 1 along the main diagonal, such that

$$\Omega = ADA'$$

Thus, on setting $H = A^{-1}$, we get $E(u_t\ u_t') = D$ diagonal.
We decomposed the innovations ϵ_t in orthogonal components: each element $u_{j,t}$ characterises the "new" information for $Y_{j,t}$. We can use this to see how predictions $Y_{k,t+s}$ respond to the new information.
A plot of $\Psi_s A$ gives the **structuralised Impulse Response Function**.

★ We could also decompose $D = D^{1/2} D^{1/2}$. For $P = AD^{1/2}$ a plot of $\Psi_s P$ gives the structuralised Impulse Response Function for shocks measuring a standard deviation of u_t.

★ In the $n = 2$ case,

$$\begin{bmatrix} 1 & 0 \\ a_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

so

$$\begin{aligned} u_{1,t} &= \epsilon_{1,t} \\ u_{2,t} &= \epsilon_{2,t} - a_{21} u_{1,t} \end{aligned}$$

The identification depends on the ordering of the variables.
Had we set $Y_t = (Y_{2,t}, Y_{1,t})'$ we would have a different matrix, say B, to orthogonalize the innovations, and

$$\begin{aligned} u_{2,t} &= \epsilon_{2,t} \\ u_{1,t} &= \epsilon_{1,t} - b_{21} u_{2,t} \end{aligned}$$

Thus, with the first choice the correlated part of the shock is given to $u_{1,t}$, with the second choice the correlated part of the shock is given to $u_{2,t}$.
The choice of the ordering has great consequencies and should be informed by what we know of economic theory.

★ Other forms of identification are also possible, these may not even require setting up a triangular matrix.

Forecast Error Variance Decomposition

Forecast error

$$Y_{t+s} - \widehat{Y}_{t+s|t} = \epsilon_{t+s} + \Psi_1 \epsilon_{t+s-1} + ... + \Psi_{s-1} \epsilon_{t+1}$$

MSE forecast

$$MSE(\widehat{Y}_{t+s|t}) = \Omega + \Psi_1 \Omega \Psi_1' + ... + \Psi_{s-1} \Omega \Psi_{s-1}'$$

We can decompose Ω using the Cholesky decomoposition or other techniques, and see which innovations contributed most to the forecast variance.

Estimation: IRF and FEVD

★ Matrix $\widehat{\Omega}$ may be estimated as a sample variance covariance from the residuals of the estimated VAR. Matrices $\widehat{A}$ and $\widehat{D}$ can be computed directly from $\widehat{\Omega}$. Estimated standard errors for the IRF and FEXD can also be computed.