Vector Autoregression

School of Economics, University College Dublin

Spring 2018

Matrix notation

Consider VAR model with 2 variables, 1 lag

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + \epsilon_{2t} \end{aligned} \tag{1}$$

Can write in matrix notation as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{2}$$

$$Y_t = AY_{t-1} + e_t \tag{3}$$

Cumulation of shocks: VAR expresses variables as function of shocks

1. Yesterday $t-1$
2. Today t

Shock at $t-1$ depends on shock at $t-1$ and $t-2$, etc.

- Value at t is cumulation of the effect of all shocks from the past

The fact that the value at t depends on what happened at $t-n$ is useful for generating predictions for $t+1$

VAR can therefore be represented as **Vector Moving Average** (VMA):

$$
\begin{aligned}
Y_t &= e_t + AY_{t-1} \qquad (4)\\
&= e_t + A(e_{t-1} + AY_{t-2})\\
&= e_t + Ae_{t-1} + A^2(e_{t-2} + AY_{t-3})\\
&= e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0
\end{aligned}
$$

Shocks: Introduce an initial shock and let the error terms be 0 afterwards

$$e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{5}$$

$$e_t = 0 \, , t > 0 \tag{6}$$

Using VMA representation we get

$$Y_t = e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0 \tag{7}$$

Response after n periods will be

$$A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{8}$$

VAR's IRF is directly analogous to AR(1) IRF.

Forecasting: Given information on Y_t we want to forecast Y_{t+1}; can model Y_{t+1} as

$$Y_{t+1} = AY_t + e_{t+1} \tag{9}$$

Given $\mathbb{E}(e_{t+1}) = 0$, an unbiased forecast at time t is AY_t

$$E_t Y_{t+1} = AY_t \tag{10}$$

- Similarly, $A^2 Y_t$ is an unbiased forecast of Y_{t+2} and $A^n Y_t$ of Y_{t+n}

Once estimated, VAR easily used for forecasts

Semantics:

1. **Forecast:** probabilistic statement, usually over a longer time period
2. **Prediction:** definitive and specific statement

From *The Signal and the Noise* (Silver, 2012)

Two-lag system:
Using first-order representation

$$y_{1t} = a_{11}y_{1,t-1} + a_{12}y_{1,t-2} + a_{13}y_{2,t-1} + a_{14}y_{2,t-2} + \epsilon_{1t} \tag{11}$$
$$y_{2t} = a_{21}y_{1,t-1} + a_{22}y_{1,t-2} + a_{23}y_{2,t-1} + a_{24}y_{2,t-2} + \epsilon_{2t}$$

In matrix form

$$Z_t = AZ_{t-1} + e_t \tag{12}$$

Notation is similar to simpler model, estimation will just be more complex

Reduced-form:

$$Z_t = AZ_{t-1} + e_t$$

$$Z_t = \begin{pmatrix} y_{1t} \\ y_{1,t-1} \\ y_{2t} \\ y_{2,t-1} \end{pmatrix} ; A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 1 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 1 & 0 \end{pmatrix} ; e_t \quad = \begin{pmatrix} e_{1t} \\ 0 \\ e_{2t} \\ 0 \end{pmatrix} \quad (13)$$

The reduced-form VAR is a purely econometric model with no theoretical element.

Interpreting the model:

$$e_t = \begin{pmatrix} e_{1t} \\ 0 \\ e_{2t} \\ 0 \end{pmatrix} \tag{14}$$

1. e_{1t} is a shock that only affects y_{1t}
2. e_{2t} only affects y_{2t}

But what if the shocks has an effect on both y_{1t} and y_{2t}?

- e.g. aggregate supply and demand shocks affecting inflation and output

Need to use the reduced-form model to identify structural shocks.

Structural shocks: Suppose structural and reduced-form shocks are related

$$
\begin{aligned}
e_{1t} &= c_{11}\epsilon_{1t} + c_{12}\epsilon_{2t} \\
e_{2t} &= c_{21}\epsilon_{1t} + c_{22}\epsilon_{2t}
\end{aligned} \tag{15}
$$

$$
e_t = C\epsilon_t \tag{16}
$$

Can use two VMA representations

$$
\begin{aligned}
Y_t &= e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0 && (17) \\
&= C\epsilon_t + AC\epsilon_{t-1} + A^2 C\epsilon_{t-2} + A^3 C\epsilon_{t-3} + ... + A^t C\epsilon_0 && (18)
\end{aligned}
$$

Model can be interpreted as

1. Shocks e_t, IRFs given by A^n
2. Structural shocks ϵ_t, IRFs are given by $A^n C$
   - Can be done for any C; just don't know the structural shocks.

Reduced vs. Structural shocks:
Consider contemporaneous interactions between variables

$$\begin{aligned} y_{1t} &= a_{12} y_{2t} + b_{11} y_{1,t-1} + b_{12} y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= a_{21} y_{1t} + b_{21} y_{1,t-1} + b_{22} y_{2,t-1} + \epsilon_{1t} \end{aligned} \tag{19}$$

$$AY_t = BY_{t-1} + \epsilon_t \tag{20}$$

$$A = \begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \tag{21}$$

Can re-write SVAR as reduced-form model

$$AY_t = BY_{t-1} + \epsilon_t \tag{22}$$
$$Y_t = DY_{t-1} + e_t \tag{23}$$

Reduced-form model has following coefficients and shocks

$$D = A^{-1}B \tag{24}$$
$$e_t = A^{-1}\epsilon_t \tag{25}$$

Structural model, impulse response to structural shocks from n periods given by

$$D^n A^{-1} \tag{26}$$

Holds for any arbitrary A matrix

$$\begin{aligned} Y_t &= e_t + De_{t-1} + D^2 e_{t-2} + D^3 e_{t-3} + ... \\ &= A^{-1}\epsilon_t + DA^{-1}\epsilon_{t-1} + D^2 A^{-1}\epsilon_{t-2} + D^3 A^{-1}\epsilon_{t-3} + ... \end{aligned} \tag{27}$$

'What-if': for reduced-form VAR the question

> *What happens if there is a shock to the first variable in the VAR?*

Becomes

> *What will normally happen if there is a shock to the first variable, given that this is usually associated with a corresponding shock to the second variable?*

Due to correlation of error series in reduced-form VAR: often

interested in different shock types that are uncorrelated

- Structural identification how reduced-form shocks are actually combinations of uncorrelated structural shocks

Structural VAR

$$AY_t = BY_{t-1} + C\epsilon_t \tag{28}$$

Number of parameters in the model is

$$3n^2 + \frac{n(n+1)}{2} \tag{29}$$

1. n^2 parameters in A
2. n^2 parameters in B
3. n^2 parameters in C
4. $\frac{n(n+1)}{2}$ parameters in $\sum_\epsilon$

Estimation

$$Y_t = DY_{t-1} + e_t \tag{30}$$

Provides information on $n^2 + \frac{n(n+1)}{2}$ parameters

1. Parameters in D
2. Estimated covariance matrix for the reduced-form errors

Need to impose $2n^2$ *a priori* theoretical restrictions on the structural VAR.

Imposing $2n^2$ restrictions will leave

$$n^2 + \frac{n(n+1)}{2} \tag{31}$$

known reduced-form parameters and equal number of structural parameters that we like to know; can get an unique solution here.

- $n^2 + \frac{n(n+1)}{2}$ equations in $n^2 + \frac{n(n+1)}{2}$ unknowns

e.g. can assume that reduced-form VAR is equal to SVAR

$$A = C = I \tag{32}$$

Recursive SVAR: SVARs identify shocks as coming from distinct independent sources (uncorrelated).

- How to get uncorrelated structural shocks from correlated reduced form shocks?

Reduced-form errors are combinations of set of independent structural errors

$$
\begin{aligned}
Y_t &= DY_{t-1} + e_t \\
e_t &= A^{-1}\epsilon_t
\end{aligned}
$$

$$AY_t = BY_{t-1} + C\epsilon_t$$

1. Set $A = I$
2. Construct C such that structural shocks will be uncorrelated

Cholesky decomposition: Take reduced-form VAR with error series (e_{1t} , e_{2t}, e_{3t}); Assume that one variable is first structural shock

$$\epsilon_{1t} = e_{1t} \tag{33}$$

Run following two regression involving reduced-form shocks

$$e_{2t} = c_{21} e_{1t} + \epsilon_{2t} \tag{34}$$
$$e_{3t} = c_{31} e_{1t} + c_{32} e_{2t} + \epsilon_{3t} \tag{35}$$

This produces

$$G e_t = \epsilon_t \tag{36}$$

Invert Ge_t to create C and give

$$e_t = C \epsilon_t \tag{37}$$

Cholesky decomposition posits causal chain of shocks and creates a lower-triangular matrix

1. First shock affects all variables at time t
2. Second only affects two of them at time t
3. Last shock only affects one variable at time t

Two important issues with using Cholesky decomposition:

1. Restriction assumptions: variables are sticky and do not respond immediately to some shocks
2. Ordering: not unique meaning that there are $n!$ possible recursive orderings

Ordering: Some variables only having effect on some variables at time t

- Let $C = I$, estimate A and B using OLS.

$$
\begin{aligned}
y_{1t} &= b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + b_{13}y_{3,t-1} + \epsilon_{1t} \qquad (38)\\
y_{2t} &= b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + b_{23}y_{3,t-1} - a_{21}y_{1t} + \epsilon_{2t}\\
y_{3t} &= b_{31}y_{1,t-1} + b_{32}y_{2,t-1} + b_{33}y_{3,t-1} - a_{31}y_{1t} - a_{32}y_{2t} + \epsilon_{3t}
\end{aligned}
$$

Different combinations of A, B and C can deliver the same structural model.

OLS: VAR is set of linear equations

- n-variable and n-equation model where each variable is explained by its lagged value and the current and lagged value of the $n-1$ remaining variables

OLS would be obvious technique for estimating coefficients; however it will produce biased estimates

$$y_t = \rho y_{t-1} + \epsilon_t \tag{39}$$

For AR(1) model, the OLS estimator for sample size T is

$$\begin{aligned}\hat{\rho} &= \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \\ &= \rho + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon}{\sum_{t=2}^{T} y_{t-1}^2} = \rho + \left(\sum_{t=2}^{T} \frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}\right)\epsilon_t\end{aligned} \tag{40}$$

1. ϵ_t is independent of y_{t-1}

$$\mathbb{E}(y_{t-1}\epsilon_t) = 0 \tag{41}$$

2. $\rho > 0$: positive shock to ϵ_t will increase current and future values of y_t.

y_t is a function of ϵ_t: ϵ_t is not independent of $\sum_{t=2}^{T} y_{t-1}^2$

$$\mathbb{E}\,\hat{\rho} < \rho \tag{42}$$

Due to negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$

Size of bias depends on two factors

1. Size of ρ; larger correlation of ϵ_t with $y_t + n$: larger bias
2. Sample size T; larger T will reduce fraction on y that will be highly correlated with ϵ_t: smaller bias

In VAR context, bias is more severe if 'own' lag coefficients are larger and smaller sample size.

Bootstrapping: Use OLS to estimate model

$$Z_t = AZ_{t-1} + \epsilon_t \tag{43}$$

Save errors $\hat{\epsilon_t}$ and randomly sample error series ϵ_t^* and simulated data

$$Z_t^* = \hat{A}Z_{t-1}^* + \epsilon_t^* \tag{44}$$

Estimate VAR with simulated data; save coefficients $\hat{A}^*$

- Compute median for each $\hat{A}^*$: $\bar{A}$; compare to $\hat{A}$ to get estimate of OLS bias

Can construct new estimates

$$A^{boot} = \hat{A} - (\bar{A} - \hat{A}) \tag{45}$$

Maximum Likelihood Estimation: estimator that maximises the value of the likelihood function for the observed data, for parameter set θ

$$f(y_1, y_2, ..., Y_n|\theta) \tag{46}$$

Similar to OLS, ML estimates are biased but they are also

1. Consistent
2. Asymptotically efficient

Joint likelihood: ML estimates are given by multiplying likelihood of each observation; maximising joint likelihood

$$f(y_1, ..., y_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_i - \mu)^2)}{2\sigma^2}\right] \tag{47}$$

$$= \prod_{i=1}^{n} f(y_i|\theta)$$

$$log\ f(y_1, ..., y_n|\mu, \sigma) = -\frac{n}{2} log 2\pi - n log \sigma + \sum_{i=1}^{n}\left[\frac{-(y_i - \mu)^2}{2\sigma^2}\right] \tag{48}$$

$$= \sum_{i=1}^{n} ln\ f(y_i|\theta)$$

Consider AR(1) model

$$y_t = \rho y_{t-1} + \epsilon_t \tag{49}$$

$$\epsilon_t \sim N(0, \sigma^2) \tag{50}$$

For joint unconditional series $y_1, y_2, ..., y_n$ we assume

$$y_2 \sim N(\rho y_1, \sigma^2), y_3 \sim N(\rho y_2, \sigma^2), ... \tag{51}$$

For $\theta = (\rho, \sigma)$, conditional on the first observation the joint distribution can be written as

$$f(y_2, ..., y_n|\theta, y1) = \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right] \tag{52}$$

$$logf(y_2, ..., y_n|\theta, y1) = -\frac{n}{2} log 2\pi - nlog\sigma + \sum_{i=1}^{n}\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right] \tag{53}$$

$$= -\frac{n}{2} log 2\pi - nlog\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \rho y_{i-1})^2$$

Parameters: using Cholesky decomposition, for VAR with n variables and k lags, number of parameters equals

$$n^2 k + \frac{n(n-1)}{2} \tag{54}$$

- $n = 3, k = 1$: 12 parameters
- $n = 6, k = 6$: 231 parameters

Two issues to consider

1. Can limit the number of variables/lags used
   - Misspecification: poor inferences, bad forecasts
2. Many coefficients are probably zero (or close)
   - Overfitting: poor-quality estimates, bad forecasts

Bayesian modeling: Can incorporate additional information about coefficients to produce models that are not as highly sensitive to the features of the particular data sets we are using

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{55}$$

$$P(A|B) \propto P(B|A)P(A) \tag{56}$$

Can use **Bayes' Law** to incorporate prior knowledge.

For the set of variables Z and parameters θ

$$P(\theta = \theta^* | Z = X) \propto P(Z = X | \theta = \theta^*) P(\theta = \theta^*) \tag{57}$$

In English: the probability data parameters θ take on value θ^* given data X is a function of

1. The probability that $Z = X$ if $\theta = \theta^*$
2. The probability that $\theta = \theta^*$

Probability density function: Rewrite relationship given that data and coefficients are continuous

$$f_\theta(\theta^*|X) \propto f_Z(X|\theta^*)f_\theta(\theta^*) \tag{58}$$

Model has three important components

1. Likelihood function
2. Parameters
3. Prior

Likelihood function:

$$f_Z(X|\theta^*) \tag{59}$$

For each possible value of θ^* gives the probability of observed dataset if true coefficients

$$\theta = \theta^* \tag{60}$$

The likelihood function can be calculated once you have made assumptions about the distributional form of the error process.

Prior: Summarises the researcher's pre-existing knowledge about the parameters θ, specified as distribution

$$f_\theta(\theta^*) \tag{61}$$

Prior distribution is combined with the likelihood function to produce posterior distribution

$$f_\theta(\theta^*|X) \tag{62}$$

Specifies the probability of all possible coefficient values given both the observed data and the priors

Point estimate: For best estimator can use mean of posterior distribution

$$\hat{\theta} = \int_{-\infty}^{\infty} x f_{\theta}(x|X) dx \tag{63}$$

Estimator is weighted average of

1. The maximum likelihood estimator
2. The mean of the prior distribution

With normally distributed errors Bayesian estimators of VAR coefficients are weighted averages of OLS coefficients and the mean of the prior distribution

Long-run restrictions: Identifying assumptions for VAR requires knowledge on how variables react instantaneous to certain shocks

- Variables can be slow or information available with lag
- Economic theory of little help due to focus on long run
  - Positive aggregate demand shock will on the long-run have no effect on output and positive effect on price level

Alternative approach: use theoretically-inspired long-run restrictions to identify shocks and impulse responses.

$$Z_t = BZ_{t-1} + C\epsilon_t \tag{64}$$

Covariance matrix of structural shocks is

$$\mathbb{E}(\epsilon_t \epsilon_t') = \begin{pmatrix} \mathbb{E}(\epsilon_1^2) & \mathbb{E}(\epsilon_1 \epsilon_2) \\ \mathbb{E}(\epsilon_1 \epsilon_2) & \mathbb{E}(\epsilon_2^2) \end{pmatrix} = I \tag{65}$$

Structural shocks are uncorrelated and have unit variance.

Reduced-form:

$$\sum = \mathbb{E}(e_t e_t') = \mathbb{E}\{(C\epsilon_t)(C\epsilon_t)'\} = C\mathbb{E}(\epsilon_t \epsilon_t')C' = CC' \tag{66}$$

Observed covariance structure of the reduced-form shocks provide information on how they are related to uncorrelated structural shocks

Long-run effects SVAR

$$Z_t = (\Delta y_t, \Delta x_t)' \tag{67}$$

Long-run effect of shock on y_t is sum of effects on

$$\Delta y_t, \Delta y_{t+1}, \Delta y_{t+1}, ..., \Delta y_{t+n} \tag{68}$$

i.e. long-run effect is sum of impulse responses, meaning that for model

$$Z_t = BZ_{t-1} + C\epsilon_t \tag{69}$$

The impulse response is

1. C in t
2. BC in $t+1$
3. $B^n C$ after n periods

Long-run level effect

$$D = (I + B + B^2 + B^3 + ...)C \tag{70}$$

With B's eigenvalues within unit circle

$$I + B + B^2 + B^3 + ... = (I - B)^{-1} \tag{71}$$

This becomes

$$D = (I - B)^{-1}C \tag{72}$$

Blanchard-Quah method:

$$DD' = (I - B)^{-1} CC' \left((I - B)^{-1}\right)' \tag{73}$$

We defined the covariance matrix of reduced-form shocks as

$$CC' = \sum \tag{74}$$

This can be estimated, producing

$$DD' = (I - B)^{-1} \sum \left((I - B)^{-1}\right)' \tag{75}$$

Long-run effect restriction: Assume that D is lower-triangular

1. First shock has long-run effect on first variable
2. First and second shock have long run effect on second variable
3. etc.

$$D = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix} \tag{76}$$

Cholesky factor: All symmetric matrices have a unique lower-diagonal matrix D such that DD' equals the symmetric matrix

- Symmetric matrix means that entry i, j equals entry j, i

Calculate D using known matrix

$$(I - B)^{-1} \sum \left((I - B)^{-1}\right)' \tag{77}$$

Given

$$D = (I - B)^{-1} C \tag{78}$$

Matrix C defining structural shocks can be calculated as

$$C = (I - B) D \tag{79}$$

Galí (1999): Looks at change in labour productivity versus number of hours worked

- Based on Real Business Cycle (RBC) model which assumes that technology shocks drive business cycle
- In this case hours worked should increase in booms compared to recessions

Lower-diagonal assumption is that technology shock can affect productivity in long-run, but non-technology shock cannot

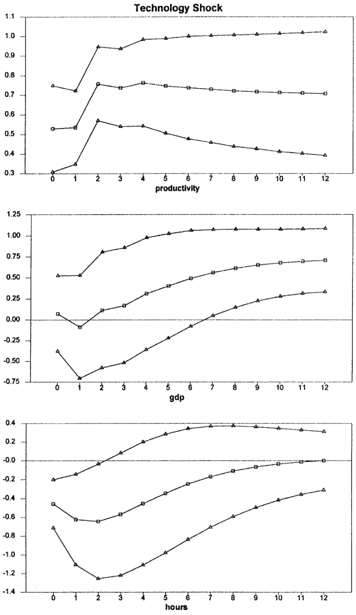

FIGURE 3. ESTIMATED IMPULSE RESPONSES FROM A BIVARIATE MODEL: U.S. DATA, DETRENDED HOURS (POINT ESTIMATES AND ±2 STANDARD ERROR CONFIDENCE INTERVALS)