Vector Autoregression

School of Economics, University College Dublin

Spring 2018

Consider VAR model with 2 variables, 1 lag

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

In matrix notation

$$
\begin{aligned}
Y &= \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \\
A &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\
e_t &= \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}
\end{aligned} \tag{2}
$$

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

VAR expresses variables as function of what happened i) yesterday $(t-1)$and ii) today's (t) shocks

- What happened yesterday depends on yesterday's shocks and what happened the day before

Ergo, today's values are cumulation of the effect of all shocks from the past

- This is useful for deriving predictions about the properties of VAR model

VAR is represented as **Vector Moving Average** (VMA):

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

Shocks: Define initial shock as

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

and all the error terms are 0 afterwards

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

using VMA representation

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

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: Suppose we have information on Y_t and want to forecast what will happen in Y_{t+1}; model for next period will be

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

Given $E_t e_{t+1} = 0$, an unbiased forecast at time t is AY_t

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

i.e. $A^2 Y_t$ is an unbiased forecast of Y_{t+2} and $A^n Y_t$ of Y_{t+n}

- Once estimated, VAR easily used for forecasts

Model we discussed so far is very simple and is just a subset of all possible VAR models, for instance

- It doesn't have a constant
- Only contains one lagged value

NB- Can add third variable as constant taking value 1: the equation for the constant term will just state that it equals its own lagged values; this formulation actually incorporates models with constant terms.

Two-lag system: Model will be more complicated in terms of estimations, but concerning notation it can still be represented using first-order

$$y_{1t} = a_{11}y_{1,t-1} + a_{12}y_{1,t-2} + a_{13}y_{2,t-1} + a_{14}y_{2,t-2} + \epsilon_{1t} \quad (11)$$

$$y_{2t} = a_{21}y_{1,t-1} + a_{22}y_{1,t-2} + a_{23}y_{2,t-1} + a_{24}y_{2,t-2} + \epsilon_{2t} \quad (12)$$

In matrix form

$$Z_t = AZ_{t-1} + e_t \quad (13)$$

Reduced-form:

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

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

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

This leaves the question of how we should interpret the model and shocks to the model. One interpretation is that e_{t1} is a shock that only affects y_{t1} on impact and e_{t2} only affects y_{t2}

- If we are interested in inflation and output, our VAR model could calculate the dynamic effect of a shock to inflation and a shock to output
- From a theoretical point of view it could be that the true shocks generating inflation and output are aggregate supply and demand shocks: they can have both an effect on inflation and output.

So how do we identify structural shocks?

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

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

$$e_t = C_{\epsilon t} \tag{16}$$

Two VMA representations

$$Y_t = e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0 \tag{17}$$
$$= C_{\epsilon t} + AC_{\epsilon t-1} + A^2 C_{\epsilon t-2} + A^3 C_{\epsilon t-3} + ... + A^t C_{\epsilon 0} \tag{18}$$

Can interpret the model as

- One with shocks e_t and IRFs given by A^n
- As a model with structural shocks ϵ_t and IRFs are given by $A^n C$: Could do this for any C, we just don't know the structural shocks

Another way to see how reduced-form shocks are different from structural shocks is if there are contemporaneous interactions between variables: consider following VAR model

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

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

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

For reduced-form model

$$Y_t = DY_{t-1} + e_t \tag{22}$$

Reduced-form coefficients and shocks are

$$D = A^{-1}B \tag{23}$$

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

In contrast, for the structural model the impulse response to structural shocks from n periods is given by

$$D^n A^{-1} \tag{25}$$

which is true for any arbitrary A matrix.

$$\begin{aligned} Y_t &= e_t + De_{t-1} + D^2 e_{t-2} + D^3 e_{t-3} + ... \\ &= A^{-1}\epsilon_t + DA^{-1}\epsilon_{t-1} + D^2 A^{-1}\epsilon_{t-2} + D^3 A^{-1}\epsilon_{t-3} + ... \end{aligned} \tag{26}$$

No problem with using reduced-form VAR to forecast; only run into trouble when you start to ask 'what-if' questions, for instance

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

Given that the error series in reduced-form VARs are usually correlated this question 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?*

Interesting questions about the structure of the economy often

concern the impact of different types of shocks that are uncorrelated.

- Structural identification that explains how the reduced-form shocks are actually combinations of uncorrelated structural shocks is far more likely to give clear and interesting answers

Structural VAR

$$AY_t = BY_{t-1} + C\epsilon_t \tag{27}$$

n^2 parameters in A; n^2 parameters in B; n^2 parameters in C

- An an additional $\frac{n(n+1)}{2}$ parameters in $\sum$ describing the variance pattern in the covariance underlying the shock terms

Most general form of the structural VAR (SVAR) is a model with $3n^2 + \frac{n(n+1)}{2}$ parameters.

Estimating reduced-form VAR

$$Y_t = DY_{t-1} + e_t \tag{28}$$

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

- Parameters in D, estimated covariance matrix for the reduced-form errors

For information on structural shocks need to impose $2n^2$ *a priori* theoretical restrictions on the structural VAR

- Can be expressed as $n^2 + \frac{n(n+1)}{2}$ equations in $n^2 + \frac{n(n+1)}{2}$ unknowns, so it has an unique solution

This will leave

- $n^2 + \frac{n(n+1)}{2}$ known reduced form parameters
- $n^2 + \frac{n(n+1)}{2}$ structural parameters that we are interested in

For example asserting that the reduced-form VAR is equal to the SVAR means imposing the $2n^2$ restrictions that $A = C = I$

NB- SVARs identify their shocks as coming from distinct independent sources and thus assume that they are uncorrelated. This method (used in the original Sims paper) uses simple regression techniques to construct a set of uncorrelated structural shocks directly from the reduced-form shocks. The error series in reduced-form VARs are usually correlated with each other. One way to view these correlations is that the reduced-form errors are combinations of a set of statistically independent structural errors. This method sets $A = I$ and constructs a C matrix so that the structural shocks will be uncorrelated.

Identification: Start with a reduced-form VAR with three variables and errors e_{1t} , e_{2t}, e_{3t}: take one of the variables and assert that this is the first structural shock

$$\epsilon_{1t} = e_{1t} \tag{29}$$

Run following two regression involving reduced-form shocks

$$e_{2t} = c_{21} e_{1t} + \epsilon_{2t} \tag{30}$$

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

This produces

$$Ge_t = \epsilon_t \tag{32}$$

Can invert Ge_t to create C and give

$$e_t = C\epsilon_t \tag{33}$$

Identification: Done

- Recall that in OLS the error terms are uncorrelated with RHS variables: here by construction, we have that ϵ_t are uncorrelated with each other

Cholesky decomposition: posits causal chain of shocks

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 issues

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: can use idea that certain shocks have an effect on only some variable at time t: let $C = I$ and estimate A and B using OLS.

$$\begin{aligned}
y_{1t} &= b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + b_{13}y_{3,t-1} + \epsilon_{1t} \\
y_{2t} &= b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + b_{23}y_{3,t-1} - a_{21}y_{1t} + \epsilon_{2t} \\
y_{3t} &= b_{31}y_{1,t-1} + b_{32}y_{2,t-1} + b_{33}y_{3,t-1} - a_{31}y_{1t} - a_{32}y_{2t} + \epsilon_{3t}
\end{aligned} \tag{34}$$

This method delivers shocks and impulse responses that are

identical to the Cholesky decomposition.

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

OLS estimation: 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

$$y_t = \rho y_{t-1} + \epsilon_t \tag{35}$$

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

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \\
&= \rho + \frac{\sum_{t=2}^{T} \epsilon_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \\
&= \rho + \sum_{t=2}^{T} \left(\frac{y_{t-1}}{\sum_{t=2}^{T}} \right) \epsilon_t
\end{aligned} \tag{36}$$

For $\rho > 0$, positive shock to ϵ_t raises value y_t for t and $t + n$

- All of which are in $\sum_{t=1}^{T} y_{t-1}^2$

ϵ_t is independent of y_{t-1}

$$\mathbb{E}(y_{t-1}\epsilon_t) = 0 \tag{37}$$

However, $\epsilon_t$is not independent of the sum $\sum_{t=2}^{T} y_{t-1}^2$

- Since y_t is a function of ϵ_t.

This entails a negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so $\mathbb{E}\,\hat{\rho} < \rho$

- Argument generalises to VAR models.

The size of the bias will depend on two factors

1. The size of ρ: The bigger this is, the stronger the correlation of the shock with future values and thus the bigger the bias.
2. Sample size T: The larger this is, the smaller the fraction of the observations sample that will be highly correlated with the shock and thus the smaller the bias.

Bootstrapping: Use estimated error terms to simulate the underlying sampling distribution of the OLS estimators when the data generating process is given by a VAR with the estimated parameters

1. Use OLS to estimate model $Z_t = AZ_{t-1} + \epsilon_t$, save errors $\hat{\epsilon}_t$
2. Randomly sample from errors to create new error series ϵ_t^* and simulated data series by recursion $Z_t^* = \hat{A} Z_{t-1}^* + \epsilon_t^*$
3. Estimate a VAR model fitting the simulated data and save the different sets of OLS estimated coefficients $\hat{A}^*$
4. Compute median for each $\hat{A}^*$ as $\bar{A}$ and compare to $\hat{A}$ to get estimate of OLS bias
5. Construct new estimates $A^{boot} = \hat{A} - (\bar{A} - \hat{A})$

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

$$f(y_1, y_2, \dots\dots, Y_n|\theta) \tag{38}$$

MLE estimates might be biased but they are i) consistent and ii)

asymptotically efficient

- MLEs cannot be obtained using analytical methods, so numerical methods are used to estimate the set of coefficients that maximise the likelihood function

ML estimates are given by maximising joint likelihood: multiplying together the likelihood of each of the observations

$$f(y_1, ..., y_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_i - \mu)^2)}{2\sigma^2}\right] \tag{39}$$

$$log\ f(y_1, ..., y_n|\mu, \sigma) = -\frac{n}{2} log 2\pi - nlog\sigma + \sum_{i=1}^{n} \left[\frac{-(y_i - \mu)^2}{2\sigma^2}\right] \tag{40}$$

Can also write this as

$$\prod_{i=1}^{n} f(y_i|\theta) \tag{41}$$

$$\sum_{i=1}^{n} ln\ f(y_i|\theta) \tag{42}$$

Consider $AR(1)$ model with $\epsilon_t \sim N(0, \sigma^2)$

$$y_t = \rho y_{t-1} + \epsilon_t \tag{43}$$

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

$$f(y_2, ..., y_n|\theta, y1) = \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right] \tag{44}$$

The log-likelihood function is

$$logf(y_2, ..., y_n|\theta, y1) = -\frac{n}{2}log2\pi - nlog\sigma + \sum_{i=1}^{n}\left[\frac{-(y_i - \rho y_{i-1})^2}{2\sigma^2}\right] \tag{45}$$

$$= -\frac{n}{2}log2\pi - nlog\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \rho y_{i-1})^2$$

OLS will provide the MLE estimate for ρ

Using Cholesky decomposition to estimate a VAR with n variables and k lags means that the number of parameters equals

$$n^2k + \frac{n(n-1)}{2} \tag{46}$$

That's 12 parameters for $n = 3, k = 1$; 231 parameters for $n = 6, k = 6$. Two issues to consider here

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{47}$$

$$P(A|B) \propto P(B|A)P(A) \tag{48}$$

Can use **Bayes' Law.** For example, suppose you have prior knowledge that A is a very unlikely event, e.g. Canada invading the U.S.A. Then even if you observe something, call it B, that is likely to occur if A is true, e.g. a television broadcast of a Canadian invasion, you should probably still place a pretty low weight on A being true.

For the set of variables Z and parameters θ

$$P(\theta = \theta^* | Z = D) \propto P(Z = D | \theta = \theta^*) P(\theta = \theta^*) \qquad (49)$$

Calculate probability that parameters θ take on a particular value θ^* given the observed data D as a function of two other probabilities

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

Rewrite relationship in form of PDF given that coefficients and data are continuous

$$f_\theta(\theta^*|D) \propto f_Z(D|\theta^*)f_\theta(\theta^*) \tag{50}$$

Model has three important components

1. Likelihood function
2. Parameters
3. Prior

Likelihood function: $f_Z(D|\theta^*)$

- For each possible value of θ^* it tells you the probability of a given dataset occurring if the true coefficients $\theta = \theta^*$
- The likelihood functions can be calculated once you have made assumptions about the distributional form of the error process.

Prior: specified as distribution $f_\theta(\theta^*)$

- Summarises the researcher's pre-existing knowledge about the parameters θ

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

$$f_\theta(\theta^*|D) \tag{51}$$

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

To get best estimator we can use the mean of the posterior distribution

$$\hat{\theta} = \int_{-\infty}^{\infty} x f_{\theta}(x|D) dx \tag{52}$$

This estimator is a weighted average of

- The maximum likelihood estimator
- The mean of the prior distribution

Weights depend on the covariances of the likelihood and prior functions

- The more confidence the researcher specifies in the prior, the more weight will be placed on the prior mean in the estimator.
- 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 of recursive VAR require knowledge on how variables react instantaneous to certain shocks: often involves guesswork

- 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

$$Z - t = BZ_{t-1} + C\epsilon_t \tag{53}$$

$$E(\epsilon_t \epsilon_t') = \begin{pmatrix} E(\epsilon_1^2) & E(\epsilon_1 \epsilon_2) \\ E(\epsilon_1 \epsilon_2) & E(\epsilon_2^2) \end{pmatrix} = I \tag{54}$$

Structural shocks are uncorrelated and have unit variance.

Covariance matrix of reduced-form errors is

$$\sum = E(e_t e_t') = E\{(C\epsilon_t)(C\epsilon_t)'\} = CE(\epsilon_t \epsilon_t')C' = CC' \tag{55}$$

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

Long-run effects SVAR

$$Z_t = (\Delta y_t, \Delta x_t)' \tag{56}$$

Long-run effect of shock on y_t is the sum of effects on $\Delta y_t, \Delta y_{t+1}$, etc.

- Long-run effect is sum of impulse responses

Impulse response for model

$$Z_t = BZ_{t-1} + C\epsilon_t \tag{57}$$

is

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

Long run level effects are given by

$$D = (I + B + B^2 + B^3 + ...)C \tag{58}$$

With B's eigenvalues within unit circle

$$I + B + B^2 + B^3 + ... = (I - B)^{-1} \tag{59}$$

This becomes

$$D = (I - B)^{-1} C \tag{60}$$

Blacnhard-Quah method:

$$DD' = (I - B)^{-1} CC' \left((I - B)^{-1}\right)' \tag{61}$$

Given

$$CC' = \sum \tag{62}$$

Which can be estimated, so we get

$$DD' = (I - B)^{-1} \sum \left((I - B)^{-1}\right)' \tag{63}$$

We are going to make a restriction concerning the long-run effects described in D which we assume to be 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.

$$D = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix} \tag{64}$$

Cholesky factor; DD' is a symmetrix matrix which means that there is a unique lower-triangular matrix D so that DD' equals the symmetrix matrix

- entry i, j equals entry j, i

Can calculate D using know matrix

$$(I - B)^{-1} \sum \left((I - B)^{-1}\right)' \tag{65}$$

Recall that $D = (I - B)^{-1} C$, which means C, defining the structural shocks, can be calculated as

$$C = (I - B) D \tag{66}$$