Vector auto regression

School of Economics, University College Dublin

Spring 2017

Background on Vector Auto Regression

The VAR model was introduced by Christopher Sims in a 1980 paper titled "Macroeconomics and reality". Sims criticized

1. The widespread use of highly specified macro-models that made very strong identifying restrictions
   - In the sense that each equation in the model usually excluded most of the model's other variables from the right-hand-side
2. The very strong assumptions about the dynamic nature of these relationships.

As highlighted in previous lecture, AR models are useful in understanding the dynamics of individual variables, but they ignore the relationships between variables. Vector Auto Regression (VAR) models the dynamics between n different variables, allowing each variable to depend on the lagged values of all the variables. Such a model allows us to examine the impulse response of all n variables to all n shocks..

The VAR model was provided as an alternative, to allow one to model macroeconomic data accurately, without having to impose lots of incredible restrictions.:

"macro modelling without pretending to have too much a priori theory."

VAR basics

Consider the simplest VAR model with 2 variables and 1 lag

$$y_{1t} = a_1 1 y_{1,t-1} + a_1 2 y_{2,t-1} + \epsilon_{1t} \tag{1}$$

$$y_{2t} = a_2 1 y_{1,t-1} + a_2 2 y_{2,t-1} + \epsilon_{2t} \tag{2}$$

A compact way to express this VAR system is using matrix notation

$$Y = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

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

Note that this system can be compactly written as

$$Y_t = AY_{t-1} + e_t$$

In other words, the VAR expresses variables as a function of what happened yesterday and today's shocks. Following the same logic, what happened yesterday depends on yesterday's shock and what happened the day before. The representation of the VAR is a Vector Moving Average (VMA) which can be obtained as follows

I.e. today's values for the series are the cumulation of the effects of all the shocks from the past. This is useful for deriving predictions about the properties of VAR.

$$\begin{aligned} Y_t &= e_t + AY_{t-1} \\ &= e_t + A(e_{t-1} + AY_{t-2}) \\ &= e_t + Ae_{t-1} + A^2(e_{t-2} + AY_{t-3}) \\ &= e_t + Ae_{t-1} + A^2e_{t-2} + A^3e_{t-3} + ... + A^te_0 \end{aligned}$$

Impulse response functions

Suppose that there is an initial shock. This shock can be defined as

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

and all the error terms are 0 afterwards. Remember the VMA representation

$e_t = 0 \; for \; t > 0$

$$Y_t = e_t + Ae_{t-1} + A^2e_{t-2} + A^3e_{t-3} + ... + A^te_0$$

This means that the response after n periods will be

This means that the IRF for a VAR model is directly analagous to the IRF of the $AR(1)$ model that we looked at before.

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

Using VAR to forecast

Remember that one of the things macroeconomists do is forecast. VARs are often used to forecast, so let's say we have information on what happens in Y_t and want to forecast what will happen in Y_{t+1}. The model for the next period will be

Albeit with very mixed results. There is a joke that says that economists correctly predicted 5 out of the last 3 recessions.

$$Y_{t+1} = AY_t + e_{t+1}$$

Since $E_te_{t+1} = 0$, an unbiased forecast at time t is AY_t, i.e. $E_tY_{t+1} = AY_t$. This makes that once a VAR is estimated it is very easy to construct forecasts.

By the same reasoning A^2Y_t is an unbiased forecast of Y_{t+2} and A^nY_t of Y_{t+n}.

Generality of matrix formulation

The model we used might look like a small subset of all possible VARs. For instance

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

A third variable can be added which takes the constant value 1 for each period. The equation for the constant term will just state that it equals its own lagged values. So this formulation actually incorporates models with constant terms.

Adding an additional lag would make the model more complicated, in terms of estimating, but it can still be represented using first-order matrix formulation. Consider a two-lag system

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

Can define a vector for the variables and their lags called Z

$$Z_t = \begin{pmatrix} y_{1t} \\ y_{1,t-1} \\ y_{2t} \\ y_{2,t-1} \end{pmatrix}$$

This system can be represented in matrix form

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

where

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 1 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

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

This system is called a reduced form VAR model. It is a purely econometric model without any theoretical element.

How should we than interpret the 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}.

Let's say that 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. However, from a theoretical point of view it could be that the true shocks generating inflation and output are aggregate supply and demand shocks. As such, they can have both an effect on inflation and output. Question is, how to identify these structural shocks? Suppose that the structural and reduced form shocks are related

$$e_{1t} = c_{11}\epsilon_{1t} + c_{12}\epsilon_{2t}$$
$$e_{2t} = c_{21}\epsilon_{1t} + c_{22}\epsilon_{2t}$$

Can be written in matrix form

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

Can use two different VMA representations

Can interpret the model as one with shocks e_t and IRFs given by A^n or 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.

$$\begin{aligned} Y_t &= e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0 \\ &= C_{\epsilon t} + AC_{\epsilon t-1} + A^2 C_{\epsilon t-2} + A^3 C_{\epsilon t-3} + ... + A^t C_{\epsilon 0} \end{aligned}$$

Contemporaneous interactions

Another way to see how reduced-form shocks are different from structural shocks is if there are contemporaneous interactions between variables, which is likely. Consider following model

$$y_{1t} = a_{12}y_{2t} + b_{11}y_{1,t1} + b_{12}y_{2,t1} + \epsilon_{1t}$$
$$y_{2t} = a_{21}y_{2t} + b_{21}y_{1,t1} + b_{22}y_{2,t1} + \epsilon_{1t}$$

This can be written in matrix form

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

where

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

If we would estimate the reduced-form VAR model the reduced-form coefficients and shocks are

The reduced-form model is $Y_t = DY_{t-1} + e_t$

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

For the structural model, the impulse responses to the structural shocks from n periods are given by $D^n A^{-1}$, which is true for any arbitrary A matrix.

Again the data can be described by two decompositions

$$\begin{aligned} Y_t &= e_t + De_{t1} + D^2 e_{t2} + D^3 e_{t3} + ... \\ &= A^1\epsilon_t + DA^1\epsilon_{t-1} + D^2A^1\epsilon_{t-2} + D^3A^1\epsilon_{t-3} + ... \end{aligned}$$

Structural VARs

In its general formulation, the structural VAR is

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

The model is fully described by

- n^2 parameters in A
- n^2 parameters in B
- n^2 parameters in C
- $\frac{n(n+1)}{2}$ parameters in $\sum$

$\sum$ describes the variance pattern in the covariance underlying the shock terms.

I.e. the most general form of the structural VAR (SVAR) is a model with $3n^2 + \frac{n(n+1)}{2}$ parameters. Again let's consider a reduced-form VAR

$$Y_t = DY_{t-1} + e_t$$

This gives information on $n^2 + \frac{n(n+1)}{2}$ parameters. In order to obtain information on the structural shocks we need to impose $2n^2$ *a priori* theoretical restrictions on the structural VAR. This will leave $n^2 + \frac{n(n+1)}{2}$ known reduced form parameters and $n^2 + \frac{n(n+1)}{2}$ structural parameters that we are interested in.

The coefficients in D and the estimated covariance matrix for the reduced-form errors.

This 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. As an example, asserting that the reduced-form VAR is equal to the SVAR means imposing the $2n^2$ restrictions that $A = C = I$.

Whereas the reduced-form VAR is a purely econometric model, the SVAR imposes economic theory to sort out the contemporaneous relationships between the variables.

ONE POPULAR METHOD TO ESTIMATE SVAR MODELS IS TO DO IT RECURSIVELY. Note that 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. This method sets $A = I$ and constructs a C matrix so that the structural shocks will be uncorrelated.

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.

Cholesky decomposition

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: $\epsilon_{1t} = e_{1t}$. Then run the following two OLS regressions involving the reduced-form shocks

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

This gives matrix equation $Ge_t = \epsilon_t$ which can be inverted to create C and gives $e_t = C\epsilon_t$. Identification: Done.

Remember that error terms in OLS equations are uncorrelated with the right-hand-side variables in the regressions. Here, by construction, we have that ϵ_t are uncorrelated with each other. Note that the there are some strong assumptions for the restrictions. Usually the reasoning underlying them draws on arguments that certain variables are sticky and do not respond immediately to some shocks.

THE CHOLESKY DECOMPOSITION METHOD POSITS A SORT OF 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

A serious drawback of the method here is that the ordering is not unique: any of the variables can be listed anywhere in the chain. This entails $n!$ possible recursive orderings.

The idea that certain shocks have an effect on only some variable at time t can be re-stated as some variables only having an effect on some variables at time t. This approach sets $C = I$ and estimates A and B using OLS.

See how the first shock affects all variables, while the last shock only affects the last variable.

$$\begin{aligned}
y_{1t} &= b_{11}y_{1,t1} + b_{12}y_{2,t1} + b_{13}y_{3,t1} + \epsilon_{1t} \\
y_{2t} &= b_{21}y_{1,t1} + b_{22}y_{2,t1} + b_{23}y_{3,t1}a_{21}y_{1t} + \epsilon_{2t} \\
y_{3t} &= b_{31}y_{1,t1} + b_{32}y_{2,t1} + b_{33}y_{3,t1}a_{31}y_{1t}a_{32}y_{2t} + \epsilon_{3t}
\end{aligned}$$

This method delivers shocks and impulse responses that are identical to the Cholesky decomposition. Shows that different combinations of A, B and C can deliver the same structural model.

Estimating a VAR model

Note that a VAR model is a set of linear equations. OLS would be an obvious technique for estimating the coefficients however OLS will produce biased estimates. Here we discuss a number of issues related to estimating VAR models.

A VAR is a 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 estimates of VAR models are biased

Consider the following $AR(1)$ model

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

The OLS estimator for sample size T is

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

If ρ is positive, then a positive shock to ϵ_t raises current and future values of y_t, all of which are in $\sum_{t=2}^{T} y_{t-1}^2$. This entails a negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so $\mathbb{E}\,\hat{\rho} < \rho$ This argument generalises to VAR models.

ϵ_t is independent of y_{t-1} so $\mathbb{E}(y_{t-1}\epsilon_t) = 0$. However, $\epsilon_t$is not independent of the sum $\sum_{t=2}^{T} y_{t-1}^2$

The size of the bias will depend on two factors

More generally, the bias in OLS estimates of VAR coefficients will be larger the higher are the "own lag" coefficients and the smaller the sample size.

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.

Bootstrap adjustment

One possible solution when the OLS estimates are biased is to use bootstrap methods. These use the 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. These calculations can be used to apply an adjustment to the OLS bias. This follows roughly the following procedure;

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

The maximum likelihood estimator (MLE) is an estimator that maximises the value of the likelihood function for the observed data. Let θ be a potential set of parameters for a model and f be a function such that when the model generates parameters θ, the joint probability density function that generates the data is given by $f(y_1, y_2, ..., Y_n|\theta)$. MLE might be biased, but they are consistent and asymptotically efficient, i.e. they have the lowest possible asymptotic variance of all consistent estimators. In general, MLEs cannot be

In other words, given a value of θ, $f(y_1, y_2, ..., y_n|\theta)$ described the probability density of the sample $(y_1, y_2, ..., y_n)$ occurring. $f(y_1, y_2, ..., Y_n|\theta)$ is called the likelihood of this sample occurring when the true value of the parameters equals θ.

obtained using analytical methods, so numerical methods are used to estimate the set of coefficients that maximise the likelihood function.

SUPPOSE A SET OF OBSERVATIONS $(y_1, y_2, ..., y_n)$ were generated by a normal distribution with an unknown mean and standard deviation, μ and σ. Then the MLEs are the values of μ and σ that maximise the joint likelihood obtained by multiplying together the likelihood of each of the observations.

Note that this is the normal density.

$$f(y_1, y_2, ..., y_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp \left[\frac{-(y_i - \mu)^2)}{2\sigma^2} \right]$$

Often econometricians work with the log-likelihood which is

More generally we can also write the likelihood simply as $\prod_{i=1}^{n} f(y_i|\theta)$ where the log-likelihood is $\sum_{i=1}^{n} ln\ f(y_i|\theta)$.

$$lnf(y_1, y_2, ..., y_n|\mu, \sigma) = -\frac{n}{2} ln2\pi - nln\sigma + \sum_{i=1}^{n} \left[\frac{-(y_i - \mu)^2}{2\sigma^2} \right]$$

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

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

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

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

OLS will provide the MLE estimate for ρ. This generalises to VAR models.

Problem with estimation

One problem with estimate of VAR models is the number of parameters.

Estimating a Cholesky decomposition VAR with n variables and k lags involves the direct estimation of $n^2k + \frac{n(n-1)}{2}$ parameters. Couple of issues to consider here

For 3 variables with one lag this is already 12 parameters, for 6 variables with 6 lags it is 231 parameters.

- Many coefficients are probably (close to) zero
  - This can lead to overfitting
  - Result is poor-quality and estimates and bad forecasts
- Researchers could limit the number of variables/lags used
  - Could result in misspecification
  - Leaving out important variables or missing important dynamics
  - Result again will be poor inference and bad forecasts

A BAYESIAN APPROACH CAN HELP OVERCOME SOME OF THESE ISSUES. It does this for VAR model by incorporating additional information about coefficients to produce models that are not as highly sensitive to the features of the particular data sets we are using.

Bayesian methods

BAYES'S LAW is a well-known result from probability theory. It states that

$$P(A|B) \propto P(B|A)P(A)$$

This can also be written as

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

For example, suppose you have prior knowledge that A is a very unlikely event, e.g. an alien invasion. Then even if you observe something, call it B, that is likely to occur if A is true, e.g. a radio broadcast of an alien invasion, you should probably still place a pretty low weight on A being true.

In the context of econometric estimation, we can think of this as relating to variables Z and parameters θ

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

Here we calculate that the probability that the vector with parameters θ takes 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 $\theta = \theta^*$

COEFFICIENTS AND DATA ARE CONTINUOUS IN A VAR MODEL. Therefore, we need to we write the Bayes relationship in form of probability density functions

$$f_\theta(\theta^*|D) \propto f_Z(D|\theta^*)f_\theta(\theta^*)$$

- The function $f_Z(D|\theta^*)$ is the likelihood function
  - for each possible value of θ^* it tells you the probability of a given dataset occurring if the true coefficients $\theta = \theta^*$.
- Bayesian analysis specifies a prior distribution: $f_\theta(\theta^*)$
  - This summarises the researcher's pre-existing knowledge about the parameters θ
- This is combined with the likelihood function to produce a posterior distribution: $f_\theta(\theta^*|D)$
  - This specifies the probability of all possible coefficient values given both the observed data and the priors

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

An obvious way to derive a "best estimator" from the posterior distribution is to calculate the mean of the distribution

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

This estimator is a weighted average of

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

The 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, the maximum likelihood estimates are simply the OLS estimates, so Bayesian estimators of VAR coefficients are weighted averages of OLS coefficients and the mean of the prior distribution.