Economics 6003
Quantitative Economics
Vector Autoregressions (VARs)

John Bluedorn

Lecturer, Economics Division

Monday/Thursday, 16/19 March 2009

What is a VAR?

- A VAR is the vector generalization of a single equation AR.
  - n linear equations for n variables, where a variable's own past and the current and past values of the other variables in the system are the explanatory variables.
  - Accordingly, there are also VMA (vector moving averages) and VARMAs (vector ARMAs).
- VARs have become a workhorse empirical tool for macroeconomists.
  - They are a parsimonious way of capturing the dynamic partial correlations among variables (descriptive role). They are also easy to estimate (equation-by-equation OLS or seemingly-unrelated regressions).
  - They are used as benchmarks for forecasting exercises.
  - When a mapping from the VAR to a structural model or identification is made (subject to the usual caveats), we can estimate the consequences of a variety of counterfactual experiments.

What is a VAR?

- For example, the behavioral equations (first-order) conditions of many macroeconomic models can be linearized, taking the form of a VAR or VARMA model. Such systems typically lead to a natural identification scheme.
  - Usual practice is to ignore the MA components of the VARMA. We will discuss possible problems with such an approach later.
- As we'll see, VAR modelling can generate a tremendous amount of statistical output using standard packages. It can be a bit overwhelming. We will consider how to winnow through the output and make things more manageable.

VAR diagnostics

- As in any model with lagged dependent variables, there are a variety of standard concerns:
  - number of lags to include. We want a parsimonious model, but we also do not want to neglect important dynamic patterns. Of course, it is also possible to specify that different lags appear for different variables. We will not consider such possibilities.
  - autocorrelation of the errors. With lagged dependent variables, autocorrelation in the errors indicates a failure of the standard orthogonality of the regressors and the error term.
  - normality of the errors. This is not essential, but it can be important if we use MLE or are worried about small sample performance.
  - stability of the implied VMA representation (the eigenvalues).

VAR lag order selection

- There are a variety of lag order selection criteria which may aid determination of the lag order for a VAR. They reward model fit and penalize model parameterization to varying degrees. See Hamilton (1994) or Greene's standard text for the exact formulas. Each criterion selects the model with the *lowest* value.
  - Akaike information criterion (AIC) $\rightarrow$ a criterion with a long history. However, it tends to choose very large lag orders. In fact, with positive probability, it will choose an infinite lag as the maximum lag considered approaches infinity.
  - Hannon-Quinn/Bayes/Schwarz information criteria (HQC, BIC, SC) $\rightarrow$ these are all pretty similar and usually generate similar lag selections. They are consistent, meaning that they will select the true lag order asymptotically.
  - Likelihood ratio (LR) and forecast prediction error (FPE) are other common criteria.

VAR lag order selection

- When using the canned Stata routine `varsoc`, you must specify a maximum lag to consider, as the criteria for multiple models will be generated. Stata helpfully highlights which lag order is chosen by each criterion.
- You will often find that the lag orders selected by the information criteria fail a number of other diagnostics. Why? The lag order selection criteria do not take into account other aspects of the model that may be important. This is a general issue with automatic model selection procedures.

VAR serial correlation tests

- Ideally, the error terms for the VAR should not show serial correlation. If they do, it is weak evidence that the lag order is not large enough.
- Why? Serial correlation leads to an endogeneity problem in the presence of lagged dependent variables. Parameter estimates are then inconsistent.
- In practice, we will usually find that there is some evidence of serial correlation at some lag in any given VAR (the usual 5% rule for spurious significance applies). Depending upon the severity and number of rejections of the null of *no* serial correlation, we may need to respecify the model and add lags.
- In Stata, the canned routine to use after calling the `var` estimation of a system is `varlmar`. It implements Breusch-Godfrey's Lagrange Multiplier test for serial correlation of the specified orders across equations.

VAR error normality tests

- It is not crucial, but normality of the errors does ensure that the models are well-behaved in small samples (or when you have burned up a lot of degrees of freedom on parameters).
- Stata implements a Jarque-Bera test of the skewness and kurtosis of the multivariate estimated errors, comparing them to those from a multivariate normal distribution. Of course, even if the model "passes", the errors need not be normal (since it does not test higher moments).
- The canned routine used post-`var` is `varnorm`.

VAR stability tests

- These are not really tests, but merely the estimated eigenvalues associated with the AR lag polynomial of the system.
- If one of them exceeds 1 in modulus, then some variable in the model may require differencing. This is related to nonstationarity issues, which we will discuss in more detail later in the course.
- In Stata, the stability conditions are checked via the `varstable` command post-`var` estimation. In practice, it is not unusual for several of the eigenvalues to be extremely close to one. Stata will also present a graphic of the unit circle with the eigenvalues indicated if you wish to have a visual representation.

VAR diagnostics

- It is generally useful to look at these diagnostics before diving into the more economically meaningful impulse responses and/or forecast error variance decompositions. They give you a feel for the data, indicating potential problem areas.
- In terms of driving changes in model specification, I personally rely upon the serial correlation tests. I figure that if the model fails these badly, then at the very least, endogeneity due to lagged dependent variables (let alone more elaborate varieties of endogeneity) is a problem.
- You will also often see Granger causality tests presented. I do not find these terribly convincing. It is easy to construct a model where a variable that is a function of expectations of another, truly exogenous variable, Granger-causes the exogenous variable. It is *not* what most people think of when they think of causality.

VAR identification

- If you look at empirical work which uses VARs, you will find that a huge amount of time is spent arguing for the particular identification scheme that is employed.
- Identification in a VAR is essentially identification in a system of simultaneous equations. In fact, we will see that the VAR identification can be interpreted as instrumental variables.
- We'll focus on identification via contemporaneous restrictions. Later in the course, we'll discuss alternative identification schemes in VARs that do not rely upon zero restrictions of contemporaneous covariances.
- The identification restrictions are related to an underlying *structural* model. This structural model is where the economic content of the exercise resides.

VAR identification

- Consider the following bivariate structural VAR(1):

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}$$

- In simpler matrix notation, this becomes:

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

where we define:

$$\begin{aligned} B &= \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \\ \Gamma_0 &= \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix} \\ E\left(\varepsilon_t \varepsilon_t^{'}\right) &= \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_z^2 \end{bmatrix} = \Phi, E\left(\varepsilon_{y,t}\varepsilon_{z,s}\right) = 0 \forall s, t, \\ E\left(\varepsilon_{i,t}\varepsilon_{i,s}\right) &= 0 \forall s \neq t, i \in \{y, z\}, E\left(\varepsilon_{i,t}\right) = 0 \forall t, i \in \{y, z\} \end{aligned}$$

VAR identification

- As defined, we can interpret ε_y and ε_z as *structural* shocks.
- We cannot estimate the structural VAR as written because of the presence of the contemporaneous connections between y and z.
- However, there is a *reduced form* VAR which corresponds to the structural VAR that *is* estimable:

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

where we have:

$$\begin{aligned} A_0 &= B^{-1}\Gamma_0, A_1 = B^{-1}\Gamma_1, \\ e_t &= B^{-1}\varepsilon_t, E\left(e_t e_t'\right) = \Omega \Rightarrow \Phi = B\Omega B' \end{aligned}$$

VAR identification

- Since the structural shocks are assumed to be uncorrelated and mean zero, then the reduced form errors are also uncorrelated and mean zero. Thus, OLS is consistent (for the reduced form parameters). In the absence of restrictions upon the reduced form, it is also efficient (with restrictions, GLS or the like is more efficient).
- Can we recover the structural parameters from the reduced form parameters? A bit of counting:
  - reduced form gives 2 intercepts plus 4 slopes plus 2 variances plus 1 covariance $= 9$
  - structural form has 2 intercepts plus 4 slopes plus 2 variances plus 2 contemporaneous B parameters $= 10$
- In the absence of additional information, we are stuck. There are more structural parameters than there are estimable, reduced form parameters.

VAR identification

- Why? The errors across equations are correlated in the reduced form, leading to an inability to decouple and thus identify the parameters of the structural equations. In fact, the errors represent linear combinations of the structural shocks.
- Identification proceeds by specifying a mapping from the structural shocks to their linear combination in the reduced form errors.
- The simplest restriction is to identify one of the reduced form errors with the structural shock. This is equivalent to allocating any contemporaneous cross-correlation between the two variables to only one of the structural shocks.
  - $B = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix}$, which implies $\varepsilon_{y,t} = e_{y,t}$ and $\varepsilon_{z,t} = e_{z,t} + b_{21}\varepsilon_{y,t}$.

VAR identification

- The zero restriction on an element of B reduces the number of structural parameters by one, allowing for the model to be just-identified with the reduced form model.
- Notice how the structure of B implies that that the matrix linking the structural shocks to the reduced form shocks is lower triangular. If the structural shocks had unit variance, it would be like the "square-root" of the Ω matrix. This is known as the *Cholesky* decomposition of the matrix Ω.
- The identification employed here is also known as a *recursive* identification, or a Wold causal chain.
- What if there are no reasonable contemporaneous restrictions? If we are only interested in the effects of a subset of the system upon the rest of the system, then sometimes block-recursiveness is enough. Otherwise, we are stuck.

Recursive identification and IV

- The recursive identification scheme has a nice interpretation as instrumental variables.
- In our simple bivariate scheme, y can act as its own instrument in the equation for z because of the restrictions.
- For more elaborate identification schemes in more complicated models, the correspondence with IV becomes more important conceptually. In general, we will need to use the restrictions to isolate variation in each equation that fulfills the instrument validity requirements for the other equations.
- This is not surprising, since a structural VAR is really just a simultaneous equations model.

Economic meaning and identification

- What does a recursive identification mean economically?
- It says that a structural shock to a variable farther down the chain only affects variables further up the chain with a lag. It is thus a restriction on the timing of responses of variables to shocks.
- Sometimes, such restrictions seem reasonable. For example, real output at the monthly frequency only responds to a monetary policy shock with a lag.
- However, if we were considering a model at the annual frequency, then it does not seem as reasonable.
- If we are considering the effect of stock market fluctuations on the exchange rate, then *no* recursive identification at common frequencies (e.g., daily, monthly, etc.) seems reasonable. These are fast-moving, information-rich variables which surely respond to each other contemporaneously.

VAR identification in larger systems

- As the number of variables increase, we need more and more restrictions on B.
- Suppose that we have a system of n variables. Then, B contains $n^2 - n$ elements plus the elements of $\Phi \Rightarrow n^2$ parameters.
- Ω contains $n^2 - \frac{n(n-1)}{2} = \frac{n^2+n}{2}$ elements.
- So, in order to pin everything down, we need $n^2 - \left(\frac{n^2+n}{2}\right) = \frac{n^2-n}{2}$ restrictions.
- Notice how the recursive identification gives us exactly that number.
- It is perfectly feasible to have more restrictions than necessary for identification. In that case, the model is deemed to be *over*-identified. A variety of additional statistical tests become available in this circumstance (we will talk about these in more detail later).

Order and rank conditions for identification

- The above discussion illustrates the **order condition** for identification $\Rightarrow$ the number of free elements in B and Φ can be no larger than the number of free elements in Ω. The order condition is necessary but not sufficient for identification.
- Why? If the restrictions occur in the wrong parts of the system, then it may be impossible to find a unique mapping from the structural shocks to their linear combinations in the reduced form errors.
- The **rank condition** is necessary and sufficient for identification. Essentially, it states that B is of sufficient rank. Most generally, it states a set of linear independence requirements equation-by-equation (identification can fail for some but not for others).

VAR identification

- A recursive identification scheme always fulfills the order and rank conditions. Moreover, the VAR is *globally* identified under a recursive scheme.
- We have restricted ourselves to recursive identification schemes. However, for richer models (more variables), it is perfectly feasible to consider different restrictions on the contemporaneous covariances of the VAR.
- Such approaches are referred to by the more general phrase as *structural* identification schemes. To be valid, they must fulfill the order and rank conditions. Unlike the recursive scheme, these can be evaluated only locally *after* VAR estimation.
- In Stata, they are executed using the `svar` command.

VAR impulse response functions

- In my opinion, impulse responses are the most useful outputs of a VAR analysis. If the identification assumptions are credible, then the impulse responses allow us to infer what the causal effect of a variable's structural shock upon another variable is.
- Moreover, they are inherently dynamic, which allows us to consider how the causal effect may vary over time. In the VAR, the impulse responses are complicated, non-linear functions of the identification constraints and the estimated AR coefficients.

VAR impulse response functions

- If we rewrite the VAR as a VMA, we can see this:

$$\begin{aligned} x_t &= \bar{A}_0 + \sum_{h=0}^{\infty} A_1^h e_{t-h} \Rightarrow \\ x_t &= \bar{A}_0 + \sum_{h=0}^{\infty} A_1^h B^{-1} \varepsilon_{t-h} \Rightarrow \\ \frac{\partial}{\partial \varepsilon_{t-j}} (x_t) &= A_1^j B^{-1} \varepsilon_{t-j} \Rightarrow \\ x_{t+j} - E_t (x_{t+j}) &= \sum_{h=0}^{j} A_1^h B^{-1} \varepsilon_{t+j-h} \end{aligned}$$

where $\bar{A}_0$ is the appropriate intercept for the VMA representation.

VAR impulse response functions

- The impulse response represents the difference between a world where the structural shock occurred and one where it did not.
- When stated in this manner, it becomes clear that the proper specification of the model (no omitted variables) and identification (appropriate decomposition of the simultaneity) are crucial to achieving good estimates of the dynamic causal effects of a variable.
- In Stata, impulse responses are estimated after `var` estimation and saved into a dataset via a suite of `irf` commands. The standard value chosen for the impulse is 1 standard deviation of the corresponding residual (why?).
- Since IRFs can be graphed, they can communicate an immense amount of information in a parsimonious manner.

VAR forecast error variance decompositions

- The VMA for a structurally identified VAR also allows us to consider how much of the forecast error variance at a given horizon $E\left[x_{t+j} - E_t\left(x_{t+j}\right)\right]^2$ is accounted for by a particular structural shock (e.g., ε_z for y). This is possible because of the uncorrelated nature of the structural shocks.
- Notice how the forecast error variance decomposition is closely related to the IRF. If we are only interested in the effects of subset of the variables, then block-recursiveness is enough.
- In my opinion, the interpretation of FEVDs requires an unpalatable assumption. They presume that the identified structural shock represents *all* of the interesting variability in the corresponding variable.

Summary

- VARs are an important empirical tool of modern macroeconomics and policy evaluation.
- There are some caveats of which to be aware in their application:
  - IRFs and FEVDs tell us something about the relationships among variables *during* the sample. As such, we must be careful in interpreting their counterfactual implications. In particular, if we wish to consider an impulse that is far larger than previously experienced, inference becomes much less reliable (in part, because of the Lucas critique).
- We'll consider an example recursive VAR by Christiano, Eichenbaum, and Evans (1999) next time.