14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
October 4, 2007
revised October, 2012

Lecture 12-13

Structural VARs

1. Goals & Assumptions

This lecture is about applied Macroeconomics. This is not a course about macroeconomics, so we won't have too much to say about the correctness of various macroeconomic assumptions. However, we will try to clearly separate econometric issues from macroeconomic ones.

History

Sims (1980) "Macroeconomics and Reality" introduced VARs. It is largely a philosophical paper. The common practice at the time was to estimate very large macro-models with many equations and many restrictions. Sims argued that these restrictions were often unrealistic. Sims introduced VARs as an alternative.

Goals of VAR analysis

- Estimate causal relations – eg. how does monetary shock affect output?
- Perform Policy analysis – eg. how should fed adjust interest rate to get the prescribed change in GDP?
- Test economic theories – eg. test RBC against Neo-Keynesian models as in problem set. RBC implies that a permanent shock to labor productivity should cause hours to go up. Neo-Keynesian implies the opposite.
- First step in DSGE estimation of deep parameters – eg. estimate VAR in real data, and match these estimates with the theoretical ones to recover the structural parameters.

Plan of VAR analysis

1. Estimate by OLS $A(L)Y_t = e_t$, a $VAR(p)$
2. Invert VAR to get $MA(\infty)$,
$$Y_t = C(L)e_t$$
3. Impose identification assumption to find a matrix D such that $Du_t = e_t$ where u_t are orthonormal, *i.e.* $Eu_tu_t' = I_t$, serially uncorrelated, and have structural interpretation (a.k.a "can be labeled as shocks").
4. Compute impulse responses from $Y_t = \tilde{C}(L)u_t$, where $\tilde{C}_j = DC_j$

2. Assumptions behind VAR

- Stationarity (no regime-switching) – all parameters are stable. This is important and tricky for policy analysis. The Lucas critique says that if you change policy, people will change their expectations and their behavior, so the coefficients of your VAR might change.
- Linearity
- Finite-order VAR
  Some people argue that it is not important that the true model has a finite-order VAR. They say that you can always add more lags as your sample size increases. However, in practice, people typically use fairly few, often 4, lags in estimation. Given the size of available data, we can't use many more lags.
- Forecast errors, $\{e_t\}$, span the space of structural shocks – this is an economic assumption, which we can use theory to think about.

3. Identification

Now, let's start thinking about identification. From our OLS regression we get "forecast errors" e_t with variance-covariance matrix $Ee_te_t' = \Omega$, which is k by k symmetric and positive-definite and has $\frac{k(k+1)}{2}$ unique parameters. We wish to recover u_t which we can label as "structural shocks". They relate to e_t by linear transformation: $e_t = Du_t$. We usually automatically impose two type of assumptions: 1) structural shocks are orthogonal to each other; 2) normalization (say variances of shocks are equal to1). This implies that D must satisfy $DD' = \Omega$. D is a $k \times k$ matrix, so it has k^2 free parameters. Therefore, calculating the number of free parameters and the number of restrictions we will need $\frac{k(k-1)}{2}$ identification restrictions.

Identification can be achieved by two general types of restrictions, short and long run.

3.1 Short-run Restrictions

One type of short-run restrictions is called "world ordering" and based on ordering variables in VAR according their speed of reaction to different shocks.

Example 1. Suppose we have three variables in the vector Y_t: prices p_t, output y_t and money m_t for which OLS residuals (forecast errors)are e_t^p, e_t^y and e_t^m. We wish to find matrix D that would lead us to structural shocks (there will be three of them) u_t^p, u_t^y and u_t^m. You may label them, say, call u_t^m as "monetary shock" and u_t^y as "supply shock". In general,

$$\begin{bmatrix} e_t^p \\ e_t^y \\ e_t^m \end{bmatrix} = \begin{bmatrix} \star & \star & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix} \begin{bmatrix} u_t^p \\ u_t^y \\ u_t^m \end{bmatrix}$$

Let's assume that monetary shock can influence output only with lag (i.e. immediate response of output to monetary shock is zero):

$$\begin{bmatrix} e_t^p \\ e_t^y \\ e_t^m \end{bmatrix} = \begin{bmatrix} \star & \star & \star \\ \star & \star & 0 \\ \star & \star & \star \end{bmatrix} \begin{bmatrix} u_t^p \\ u_t^y \\ u_t^m \end{bmatrix}$$

Now, let's impose an assumption in Keynsian tradition that prices are sluggishly responding to all shocks but their own:

$$\begin{bmatrix} e_t^p \\ e_t^y \\ e_t^m \end{bmatrix} = \begin{bmatrix} \star & 0 & 0 \\ \star & \star & 0 \\ \star & \star & \star \end{bmatrix} \begin{bmatrix} u_t^p \\ u_t^y \\ u_t^m \end{bmatrix}$$

Three "exclusion" restrictions is what we need in this example for identification.

More generally, *"world ordering"* leads to the case of triangular D . One way of estimating D is to find D by Choleski decomposition of Ω.

3.1.1 Treating exclusion restrictions as instrumental variable.

Even though performing Choletski decomposition is easy, it is useful to think about different method for implementing structural restriction, namely, as GMM moment conditions (or IV regression in this case). Remember, what we have:

$$Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + e_t; \quad e_t = Du_t;$$

that is,

$$D^{-1}Y_t = D^{-1}A_1 Y_{t-1} + ... + D^{-1}A_p Y_{t-p} + u_t = B_1 Y_{t-1} + ... + B_p Y_{t-p} + u_t, \tag{1}$$

where $B_i = D^{-1}A_i$. We imposed the identification condition that $D = \begin{pmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{pmatrix}$ is a lower diagonal matrix. But then D^{-1} is also a lower diagonal matrix, assume that d_{ij} are elements of D^{-1}. Then equation (1) means that

$$\begin{pmatrix} d_{11}p_t \\ d_{12}p_t + d_{22}y_t \\ d_{13}p_t + d_{23}y_t + d_{33}m_t \end{pmatrix} = B_1 Y_{t-1} + ... + B_p Y_{t-p} + u_t,$$

with additional assumptions that u_t are not auto-correlated over time and different of u_t are nt correlated with each other either. There is also normalization assumption that $Var(u_t) = I_3$. This suggests an estimation plan:

Step 1. Run OLS of p_t on $Y_{t-1}, ..., Y_{t-p}$, get residuals $\hat{u}_t^p$, re-normalize by defining $d_{11} = 1/\sqrt{\frac{1}{T}\sum_t (\hat{u}_t^p)^2}$, and $u_t^p = d_{11}\hat{u}_t^p$(this imposes variance 1).

Step 2. We wish to run a regression of y_t on p_t and $Y_{t-1}, ..., Y_{t-p}$. Unfortunately OLS does not work in this case, since p_t is endogenous(!!!). But we have an instrument: u_t^p (think why this is a good instrument! is it correlated with p_t? is it uncorrelated with u_t^y?). So we run IV regression of y_t on p_t and $Y_{t-1}, ..., Y_{t-p}$ with instrument u_t^p for p_t. We get preliminary coefficient a on regressor p_t. Find the residuals from the IV regression and re-normalize everything to impose $Var(u_t^y) = 1$. Namely $d_{22} = 1/\sqrt{\frac{1}{T}\sum_t (\hat{u}_t^y)^2}$, $u_t^y = d_{22}\hat{u}_t^y$, $d_{12} = -ad_{22}$.

Step 3. We want to regress m_t on p_t, y_t and $Y_{t-1}, ..., Y_{t-p}$. Now we have two endogenous regressors: y_t and p_t and two instruments u_t^p and u_t^y. Proceed in a fashion similar to step 2.

As an output of the above procedure we have a series for u_t, matrix D^{-1} and $B_1, ..., B_p$. To get impulse responses we may just invert VAR as before, or go alternative way: we may regress Y_t on u_{t-j} and get $\tilde{C}_j$ as OLS coefficients (argue why it should work!).

3.1.2 Puzzles

Another way of imposing short-run restrictions is if one know some constant relations. For example, government spending is affected by output shock only through taxes (if you assume that all taxes are spent every period). If you know the marginal tax rate, you can impose this constant.

In the past, many people wrote papers that used VARs to identify various puzzles. Among known puzzles is so called *price puzzle*:

- Run a VAR with $Y_t = (i_t, \pi_t)'$, where i_t is the federal funds rate and π_t is inflation

- Identification: define a monetary shock as an unexpected change to federal funds rate, that is $u_t^m = e_t^i$, or

$$\begin{pmatrix} u_t^m \\ u_t^2 \end{pmatrix} = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \begin{pmatrix} e_t^i \\ e_t^\pi \end{pmatrix}$$

- What you can find after producing impulse responses is: the drop in federal fund rate (positive monetary shock, money easining) leads to inflation **fall** rather than rise! This is the opposite direction from what we'd think!

Sims explained this by saying that the Fed uses many other variables when choosing the federal funds rate. When these variables pointed to an increase in inflation, the Fed would raise the rate to fight the inflation (negative monetary shock). Then inflation increases (although presumably by less than it would have if the Fed had not increased the rate), and it looks like a contraction in the monetary supply caused an increase in inflation. Remember our discussion on rational forcasts and forward looking expectations? It may look like reverse causality.

A potential solution to this would be to add more variables to your VAR. However, the number of parameters to estimate is the number of lags times the squared number of variables. This will quickly exceed the length of available data. A more practical solution is factor analysis, which will be covered next lecture.

3.2 Long-run Restrictions

Introduced by Blanchard and Quah (1989). They did a VAR with two variables, first difference of output and unemployment, that is, $Y_t = (\Delta y_t, u_t)'$:

$$Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + e_t; \text{thus} \quad Y_t = \sum_{j=0}^{\infty} C_j e_{t-j}$$

They had two shocks, one that they called a supply shock and another that they called a demand shock. Let's denote the supply and demand shocks as v_t^s and v_t^d. That is

$$e_t = D \begin{pmatrix} v_t^s \\ v_t^d \end{pmatrix}; \begin{pmatrix} \Delta y_t \\ u_t \end{pmatrix} = \sum_{j=0}^{\infty} \widetilde{C}_j \begin{pmatrix} v_{t-j}^s \\ v_{t-j}^d \end{pmatrix},$$

where $\widetilde{C}_j = C_j D$. Identification comes from the assumption that *only supply shocks can have a permanent effect on output*. Additionally, both shocks only have temporary effects on unemployment (this is implicitly imposed by the fact that we estimate the VAR with unemployment in levels). Assume for simplicity that

$$\tilde{C}_j = \begin{bmatrix} \tilde{c}_j^{(1)} \\ \tilde{c}_j^{(2)} \end{bmatrix}.$$

Consider the impulse response of y_t to $\begin{bmatrix} v_{t-j}^s \\ v_{t-j}^d \end{bmatrix}$. We know $y_t = y_{t-1} + \Delta y_t = y_{t-j} + \sum_{i=0}^{j-1} \Delta y_{t-i}$, so

$$\frac{\partial y_t}{\partial \begin{bmatrix} v_{t-j}^s \\ v_{t-j}^d \end{bmatrix}} = \sum_{i=1}^{j} \tilde{c}_j^{(1)}$$

$$\rightarrow \sum_{i=1}^{\infty} \tilde{c}_j^{(1)}$$

Our identification restriction is that

$$\sum_{i=1}^{\infty} \tilde{c}_j = \begin{bmatrix} j_{11} & 0 \\ j_{21} & j_{22} \end{bmatrix}$$

Note that $\sum_{i=1}^{\infty} \tilde{c}_j = \left(\sum_{i=1}^{\infty} C_j\right) D = C(1)D = (I_2 - A(1))^{-1} D$, where $C(L) = \sum_j C_j L^j = (I - A(L))^{-1}$,

$$\begin{bmatrix} \Delta y_t \\ u_t \end{bmatrix} = A(L) \begin{bmatrix} \Delta y_{t-1} \\ u_{t-1} \end{bmatrix} + e_t$$

is VAR regression. So, the plan is as follows

Step 1 Run OLS regression $Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + e_t$, get OLS coefficients $A_1, ..., A_p$ and variance of the residuals Ω

Step 2 Find D such that $(I_2 - \sum_{j=1}^{p} A_j)^{-1} D = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$ and $\Omega = DD'$.

3.2.1 Implementing long-run restriction as IV

This identification problem however can also be resolved as an IV regression. Again the problem is

$$Y_t = A_1 Y_{t-1} + ... + A_p Y_{t-p} + e_t; \quad e_t = D v_t$$

with the restriction $(I - \sum_{j=1}^{p} A_j)^{-1} D = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$ and $\Omega = DD'$.

Let us re-write it as

$$D^{-1} Y_t = D^{-1} A_1 Y_{t-1} + ... + D^{-1} A_p Y_{t-p} + v_t$$

or

$$B_0 Y_t = B_1 Y_{t-1} + ... + B_p Y_{t-p} + v_t, \tag{2}$$

where $B_0 = D^{-1}$ and $B_i = D^{-1} A_i$. Let $B(L) = B_0 - B_1 L - ... - B_p L^p$. We have that our identifying condition is

$$B(1)^{-1} = (D^{-1}(I - \sum_{j=1}^{p} A_j))^{-1} = (I - \sum_{j=1}^{p} A_j)^{-1} D = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix},$$

which is equivalent to $B(1) = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$. Or if $B_0 = \begin{pmatrix} 1 & \alpha \\ * & * \end{pmatrix}$ and $B_1 L + ... + B_p L^p = \begin{pmatrix} \beta(L) & \gamma(L) \\ * & * \end{pmatrix}$, then the identifying condition is $\alpha - \gamma(1) = 0$.

In our case $Y_t = \begin{pmatrix} \Delta y_t \\ u_t \end{pmatrix}$. Let us write the first equation in (2):

$$\Delta y_t + \alpha u_t = \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \gamma_1 u_{t-1} + ... + \gamma_p u_{t-p} + v_t^s,$$

So, we would love to run a regression

$$\Delta y_t = -\alpha u_t + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \gamma_1 u_{t-1} + ... + \gamma_p u_{t-p} + v_t^s,$$

but we also need to impose the restriction $\alpha - \sum_{i=1}^{p} \gamma_i = 0$. This gives us

$$\Delta y_t = \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} - \gamma_1 (u_t - u_{t-1}) - ... - \gamma_p (u_t - u_{t-p}) + v_t^s.$$

Now notice that $u_t - u_{t-1} = \Delta u_t$, $u_t - u_{t-2} = \Delta u_t + \Delta u_{t-1}, ..., u_t - u_{t-p} = \Delta u_t + ... + \Delta u_{t-p+1}$. Then our regression with the imposed restriction is

$$\Delta y_t = \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} - \widetilde{\gamma}_1 \Delta u_t - ... - \widetilde{\gamma}_p \Delta u_{t-p+1} + v_t^s, \tag{3}$$

where $\widetilde{\gamma}_1 = \gamma_1 + ... + \gamma_p$, $\widetilde{\gamma}_2 = \gamma_2 + ... + \gamma_p$, etc. The problem of running the regression (3) is that the regressor Δu_t is endogenous, the idea is to use $u_{t-1}, ..., u_{t-p}$ as instruments. Once one runs (3) as an IV regression, he gets residuals v_t^s that can be used as instruments on the second equation: which is a regression of u_t on Δy_t and p lags of both u_t and Δy_t.

Examples

- King, Plosser, Rebelo, and Watson (1991) – long-run money neutrality
- Gali (1999) – technology shock is the only shock that has a permanent effect on productivity.

Critiques of long-run restrictions

- Faust and Leeper (1997): Long-run effect is badly estimated in finite samples. Long-run restrictions transfer this uncertainty to other parameters.

  What assumptions do we need to estimate a long-run effect from a finite sample?

  - If we know that the process is $AR(p)$, then we can consistently estimate $\sum_{t=1}^{\infty} C_t = (1-\sum_{j=1}^{p} A_p)$, where A_j are autoregressive coefficients.
  - What if we think that AR(p) is only an approximation, and the true process is AR(∞)? If the process is some general $MA(\infty)$, then we cannot estimate $\sum_{t=1}^{\infty} C_t$ consistently because we could always have C_{T+1} large (more formally Faust and Leeper get a "nearly observationally equivalence"). Another way to see this is that if we have some arbitrary $AR(\infty)$, $a(L)$ and we approximate it by $\tilde{a}(L)$ of order p, then we can consistently estimate $a(L)$ by $\tilde{a}(L)$ by letting $p \to \infty$ slowly, but the long run effect is $a(1)^{-1}$, is not a continuous function of $a(L)$ in $\mathcal{L}^2$, so $\tilde{a}(1)^{-1}$ need not be close to $a(1)^{-1}$.

  There are two solutions to the problem: either to stick to the belief that the true process is AR(p), or we re-define "the long run" as say ten years.

- Another critical paper is Cooley and Dwyer (1998). They simmulated data with very persistent AR parameter and did long-run structural VAR estimation on it. They showed that the result may be highly misleading. The reason for this is that unit root is discontinuity point for long-run restriction. For IV type of estimation it corresponds to weak IV.
- More recent critique is in Chari, Kehoe, and McGrattan (2005). they simulated data from a model with a $VAR(\infty)$ representation with slowly declining coefficients, and showed that estimating an SVAR with long-run restrictions from this model gives misleading inferences.
- Christiano, Eichenbaum, and Vigfusson (2006) replied to the paper above in two ways: 1) they showed that the model of Chari, Kehoe, and McGrattan is not relevant and is strongly rejected by the data. 2) Additionally, they show that if you use a Newey-West type estimator to estimate the long-run response, the VAR works much better, even for the model of CKM.