Vector Autoregressions[1]

We now turn our attention to estimating an entire system of variables through the use of vector autoregressions. Ultimately, we wish to estimate and forecast how shocks to one variable in the system may affect others over time. This estimation strategy is one of two covered in this class (the other being panel estimators) that is very widely used in senior theses at Bates.

The vector autoregression is a method that deals with endogeneity and the temporal nature of the data (including delayed effects) by estimating an entire system of variables. There is thus no longer a single endogenous (or dependent) variable. Rather all of the variables in the system are endogenous.

Seemingly Unrelated Regression Models

Consider the following example of two regressions:

$$y_t = \alpha x_t + \beta z_t + u_t \tag{1}$$

$$h_t = \gamma x_t + \omega g_t + v_t \tag{2}$$

Supposing that these regressions are properly specified (no omitted variables, stationarity, etc.) They are seemingly unrelated because, at first glance, it seems appropriate to simply estimate them separately. Notice, however, that each specification includes a vector (x_t) of common independent variables. As a result, it is possible that the error terms (u_t and v_t) may be correlated between equations. OLS ignores this information. It is consistent (the coefficient estimates are unbiased), but it is not efficient (the standard errors are incorrect).

A seemingly unrelated regressions (SUR) model exploits potential correlation across error terms by explicitly estimating a covariance matrix for the error terms. This leads to more efficient estimation.

In Stata, estimating a SUR model consists of the following command:

$$surreg \ (y \ \ x)(z \ \ x) \tag{3}$$

where each expression inside parenthesis is a seemingly unrelated regression.

---

[1]These are undergraduate lecture notes. They do not represent academic work. Expect typos, sloppy formatting, and occasional (possibly stupefying) errors.

VARs

Estimating an entire system has two complications. First, it typically yields a SUR model that should be efficiently estimated as discussed above.[2] Second, the system typically suffers from endogeneity. The solution to the second problem is thus to instrument using lags.[3] A VAR specification thus includes lags of all the variables on the right hand side. Because we typically suspect that each variable also depends on its own lags, we include these as well. As a result, the right hand side variables are the same for each row. Consider the example of output, inflation, and interest rates. A VAR(1) (one lag) takes the following form:

$$y_t = \alpha_1 + \beta_{1,1} y_{t-1} + \beta_{1,2} i_{t-1} + \beta_{1,3} \pi_{t-1} + u_{1,t} \tag{4}$$

$$i_t = \alpha_2 + \beta_{2,1} y_{t-1} + \beta_{2,2} i_{t-1} + \beta_{2,3} \pi_{t-1} + u_{2,t} \tag{5}$$

$$\pi_t = \alpha_3 + \beta_{3,1} y_{t-1} + \beta_{3,2} i_{t-1} + \beta_{3,3} \pi_{t-1} + u_{3,t} \tag{6}$$

or using matrix notation:

$$x_t = \alpha + \beta x_{t-1} + u_t \tag{7}$$

The key assumption here is that lagged independent variables are uncorrelated with current error terms. If so, then the system may be estimated as a SUR model. The regression coefficients are thus the same as running OLS for each row, but the estimated covariance matrix, $E[u_t u_t'] = \Sigma$, is different in that it does not have zeros on all of the off-diagonals.

This is known as a VAR(1) because it includes just one lag on the right hand side. More generally, however, a VAR(p) includes p lags. We can choose the correct lag length using an information criteria.

Impulse Response Functions

Even though our estimation does not employ a specific theoretical model, we should keep in mind that our data are generated by some model (most likely one of mine as they are clearly the best ones out there). Such a model is likely to look like:

---

[2]Although for a SUR model that has the exact same set of regressors for each equation, OLS is both consistent and efficient. Because we focus on this case with our VARs, we do not need to worry about this issue. In generalk, however, VARs need not have the same set of regressors and this issue will thus be important.

[3]A VAR is not, however, literally an IV (instrumental variables) estimator.

$$\theta x_t = \kappa + \Phi x_{t-1} + e_t \tag{8}$$

or:

$$x_t = \theta^{-1}\kappa + \theta^{-1}\Phi x_{t-1} + \theta^{-1}e_t \tag{9}$$

To truly understand the effect of a shock to monetary policy, as an example which may be one element of e_t, we must know θ^{-1}.

We have the following estimates. α includes 3 elements. β includes 9. Although Σ is $3x3$, it is also symmetric. It thus only has 6 unique elements. We have thus estimated 18 unique elements.

The model of (9) includes θ and Φ which each include 9 unique elements. κ includes 3 elements. The theoretical model's error matrix, $E[e_t e_t^{'}]$, also has 6 unique elements. The theoretical model that generates the data thus has 27 unique elements.

We are thus 9 estimated parameters short. We solve this problem by imposing *identifying restrictions*. We do this in two stages:

1. We assume that $E[e_t e_t^{'}]$ is the identity matrix. This isn't costly. Almost all theoretical models are represented with independent shocks. It imposes zeroes on all the off diagonal elements of $E[e_t e_t^{'}]$. There are 6 of these, but because the matrices is symmetric, only 3 of these are unique. That leaves six more identifying restrictions to go.

2. We then normalize one element of each row to one. This is costless. It is the equivalent of re-writing:

$$3x_t = 2y_t \tag{10}$$

as:

$$1.5x_t = y_t \tag{11}$$

which, of course, changes nothing.

We choose to normalize the diagonal elements of $E[e_t e_t^{'}]$ to one. This leaves three more identifying restrictions.

3. This is the costly one. We now assume that the upper-right diagonal elements of θ^{-1} are equal to zero.

$$x_t = \theta^{-1}\kappa + \theta^{-1}\Phi x_{t-1} + \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} e_t \tag{12}$$

We are then left with 18 estimated parameters to identify the 18 remaining elements of (8).

Suppose, that there is a shock to the third element of e_t so that $e_{3,t} \neq 0$. Examining (12) shows that this shock affects only the third element of x in period t. The first two elements can be affected only with a lag. Likewise, a shock to the second element of e_t affects only the second two elements of x in period t. A shock to the first element of e_t, however, may affect all of the variables in the system without delay. It is therefore important that we order the variables from the most exogenous to the least exogenous.

This is an unusual result. Order matters in VARs. There are two ways of dealing with this issue. The first is to make a convincing case for a given order. The second is to run the VAR with different orders and show that the results do not depend on the order.

Having obtained θ^{-1}, we can now obtain an impulse response function. Suppose, for example, that we wish to obtain the reaction of output to a change in interest rates. We take the following steps:

1. We start the system at the steady state which is easily obtained by setting $e_t = 0$.

2. We then hit the economy with a one time shock to the element of e_t that corresponds to i_t. All other shocks always equal zero.

3. We then use our estimate of $\theta^{-1}\Phi$ to calculate y_t in future periods.

IRFs are easily created in Stata using the irf command.

The identification approach that we are focusing on is known as the *Cholesky Decomposition*, after the matrix decomposition that we are using. There are, however, many other ways to identify a VAR and the literature on doing so is voluminous. Time permitting, we may discuss some of these later in the semester.

VAR regression coefficients are usually not reported, theya re both too numerous and not very informative. IRFs are instead the main tool for reporting VAR results. Here is an example from a VAR that includes the Federal Funds rate, output, prices, and teh money supply[4] .

Figure 1: Estimated Impulse Response Functions
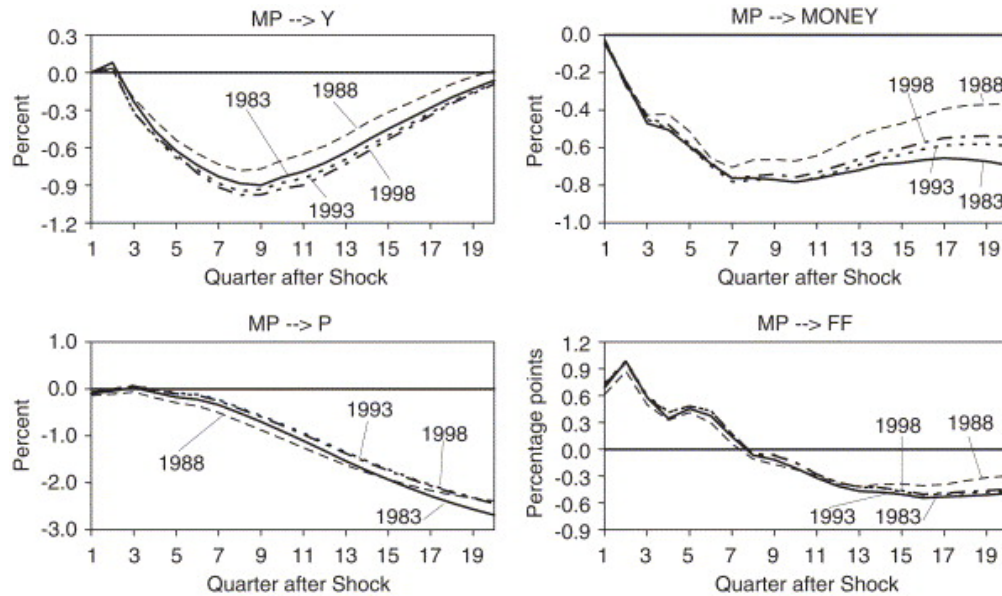
If done right, these IRFs shows how an increase (impulse) to the Federal Funds Rate causes changes to the other variables over time. If the identification approach is sound, then we will have identified an exogenous shock to the variable of interest and we will have overcome the issue of endogeneity. Because we have estimated a dynamic model, we are also able to pick up on delayed or complex (over time) dynamic responses. In this example, it takes several quarters for most of the more interesting results to manifest themselves.

An Example

Suppose that we estimate the following VAR:

$$
\begin{bmatrix} a_t \\ b_t \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ b_{t-1} \end{bmatrix} + u_t \tag{13}
$$

where $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.75 \end{bmatrix}$

and we wish to fit this to the following theoretical model:

---

[4]Taken from Gail, J. 1992. "How Well Does the IS-LM Model Fit Postwar U.S. data?" *Quarterly Journal of Economics*, vol. 107: 709-738.

$$\begin{bmatrix} a_t \\ b_t \end{bmatrix} = \theta^{-1} \Phi \begin{bmatrix} a_{t-1} \\ b_{t-1} \end{bmatrix} + \theta^{-1} e_t \tag{14}$$

We have estimated 7 unique elements (too keep things simple, I have omitted a vector of constants). Our theoretical model has 11 unique elements (four each from θ and Φ, and three from the model's error matrix).

The first step is to impose that the theoretical model's error matrix is the identity matrix. That leaves one identifying restriction to go. We achieve this be imposing that:

$$\theta^{-1} = \begin{bmatrix} c & 0 \\ d & e \end{bmatrix} \tag{15}$$

which implies that a_t is the more exogenous variable.

The theoretical model's equivalent to Σ is:

$$E[u_t u_t'] = \Sigma = E[\theta^{-1} e_t e_t' (\theta^{-1})'] = \begin{bmatrix} c & 0 \\ d & e \end{bmatrix} \begin{bmatrix} c & d \\ 0 & e \end{bmatrix} \tag{16}$$

Equating (23) to Σ:

$$c^2 = 1 \tag{17}$$

$$cd = 0.5 \tag{18}$$

$$d^2 + e^2 = 0.75 \tag{19}$$

which consists of three equations and three unknowns and can be easily solved to obtain θ^{-1}.

Forecasting

We can also forecast in Stata using the $fcast$ command after we run our VAR. Conceptually, forecasting works in the same way as with an autoregression. Re-date (7) one period forward to obtain the one period ahead point estimate:

$$E_t[x_{t+1}] = \alpha + \beta x_t \tag{20}$$

Iterating one-period further ahead then yields the two period ahead forecast:

$$E_t[x_{t+2}] = \alpha + \beta E_t[x_{t+1}] \tag{21}$$

$$E_t[x_{t+2}] = \alpha + \beta(\alpha + \beta x_t) \tag{22}$$

Another iteration yields the three period ahead forecast:

$$E_t[x_{t+3}] = \alpha(1 + \beta + \beta^2) + \beta^3 x_t \tag{23}$$

More generally

$$E_t[x_{t+k}] = \alpha(1 + \beta + \beta^2 .... + \beta^{k-1}) + \beta^k x_t \tag{24}$$

Note that we can analyze the stationarity of the simulations. If β is a scalar, then the forecast will be stationary if and only if $|\beta| < 1$. If, however, β is a matrix then we need all of the eigenvalues of β to be inside the unit circle. Suppose for example that an eigenvalue is of the form $a + bi$, where i is the imaginary part: The modulus of this eigenvalue is then $\sqrt{a^2 + b^2}$. This eigenvalue is inside the unit circle if and only if the modulus is less than one in absolute value.

It is possible that β will yield a non-stationary forecast. This is typically a sign of misspecification, for example including a non-stationary time series in your specification.

Confidence bands for VAR forecasts are generally obtained either analytically or through bootstrapping.

Exogenous Variables

We can also include exogenous variables as regressors in a VAR. Mechanically, this doesn't change the analysis much:

$$x_t = \alpha + \beta x_{t-1} + \gamma z_t + u_t \tag{25}$$

I have denoted the exogenous variables as z_t. If we are concerned about endogeneity with them, however, it would not be uncommon to use lags instead of time t variables. We would include exogenous variables if we fear that their omission will bias the results and if we are confident that they are not affected by the endogenous variables in the system. In macroeconomics, for example, this latter condition is not usually met. It would be more common to include some microeconomic variables as endogenous in a VAR, with macro variables then being added as exogenous.

Including exogenous variables does make forecasting with a VAR a little more complicated. Suppose, for example, that z_t is inflation. We then wish to form a one period ahead forecast using (25):

$$E_t[x_t] = \alpha + \beta x_t + E_t[\gamma z_{t+1}] \tag{26}$$

Forecasting now requires that we form $E_t[\gamma z_{t+1}]$. It is not obvious how we should do this. Unlike our endogenous variables, this forecast is not endogenous. Consider three, non-exhaustive possibilities:

1. We could assume that inflation (z_t is white noise). This makes forecasting easy, we just insert the sample average into $E_t[\gamma z_{t+1}]$. If z_t is not white noise, however, then our forecast will be biased.

2. We could fit z_t to a separate stochastic process (e.g. AR(1)) and use this to form an expectation of z_t.

3. We could go outside the model to find a forecast. For inflation, we could collect the Survey of Professional Forecaster's estimate and insert this into (26).

Cointegration

An important source of misspecification that may occur in VAR settings is *cointegration*. Consider the following example:

$$x_t = 1.1 x_{t-1} + e_{1,t} \tag{27}$$

$$x_t - y_t = e_{2,t} \tag{28}$$

where $e_{1,t}$ and $e_{2,t}$ are white noise error terms. Combining (27) and (28), we get:

$$y_t = 1.1 y_{t-1} + e_{1,t} - e_{2,t} + e_{2,t-1} \tag{29}$$

Both variables have one unit root. It may seem that difference them and including them in a VAR would be appropriate.

The problem is that although neither variable is stationary, their difference is as shown in (28). The two variables are thus said to be *cointegrated*. The cointegrating vector, (28), contains information and failing to incorporate this information into our specification is a form of omitted variable bias.

The solution to cointegration is a *vector error correction model* or VEC. The concept is simple. First, we must test for cointegrating vectors. There can be more than one, if we have M variables there

can be up to $M - 1$ cointegrating vectors. The most common approach is to use Johansens test, the math of which is beyond the scope of this class. Second, we then just include all of the cointegrating vectors in what is otherwise a VAR with the non-differenced variables. For a VECM(1):

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \alpha + \beta \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \gamma(x_t - y_t) + u_t \tag{30}$$

Note that we are not fixing the non-stationary through differencing. We are including the non-differenced, non-stationary variables directly in our specification. This works because including the cointegrating vector is a a different way to fix non-stationarity.

An Example

Recall the velocity of money: the number of times per period that each unit of money is spent. Suppose that the velocity of money is a stationary, stochastic process. Now recall the equation of exchange:

$$MV \equiv PY \tag{31}$$

This is an identity that must always be true. Taking logs:

$$p_t + y_t - m_t = v_t \tag{32}$$

The three variables on the left hand side are likely I(1), they must be differenced in order to render them stationary. Equation (32) is, however, a stationary stochastic process. To avoid misspecification, we should include this cointegrating vector in a VECM.

For illustration, the following chart plots (using the same data from the Money example), the logs of GDP, the CPI, and the M2 money supply.

9
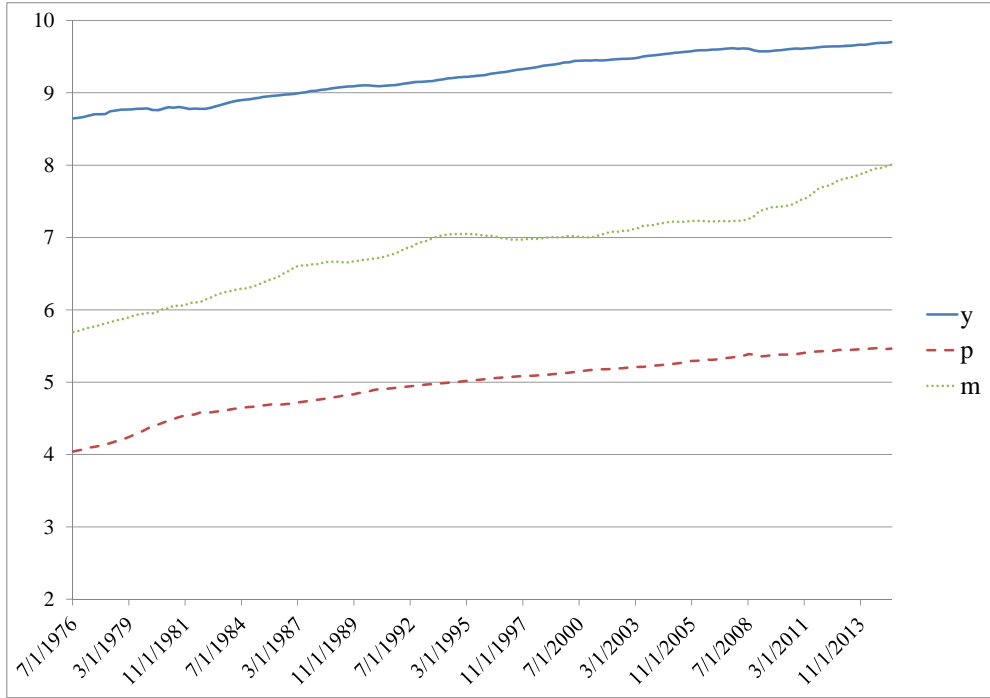
Because I am working with logs, the slope is one plus the growth rate. Each variable appears non-stationary and was confirmed to be so using Dickey Fuller tests. The following chart now plots logged velocity using (32):
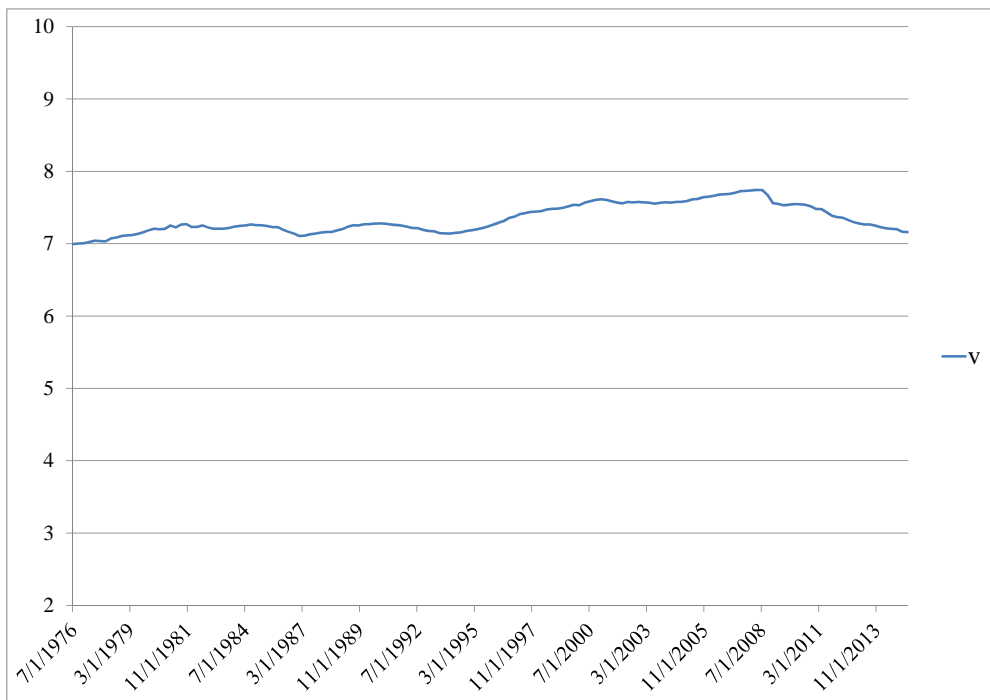
Note that this variable does appear to be stationary.

This is a hard topic. To help solidify it, we will next estimate a VAR using the monetary economics example in class.