Economics 504
Chris Georges

Handout on VARs

General Discussion of VARs:

Suppose that we are looking at the interrelationship between a number of macroeconomic variables. We are interested in knowing how exogenous shocks to any one of these variables affects the others, both contemporaneously (i.e., at the time of the shock) and over time.

Ideally, we would like to be able to model these variables (including their interactions), then estimate the model using historical data, and use this fitted model to answer our questions. However, if there is substantial disagreement about what the correct model is, then we might want to be a bit more agnostic on the theory side.

So an alternative approach would be to run a kitchen sink regression with a variety of explanatory variables (i.e., throw 'everything but the kitchen sink' into the regression model). One version of this approach is to run a *vector autoregression* (VAR). The regressors are the variables themselves (that's the 'vector' part of the name) and a set of their lags (that's the 'autoregression' part of the name).

Suppose for starters that there are only two variables X and Y that we are interested in. Then we could decide, for example, to use only one lag and write down the following regression model:

$$X_t = a_0 \cdot Y_t + a_1 \cdot Y_{t-1} + b_1 \cdot X_{t-1} + \varepsilon_{X_t}$$
$$Y_t = c_0 \cdot X_t + c_1 \cdot X_{t-1} + d_1 \cdot Y_{t-1} + \varepsilon_{Y_t}$$

We could also include constant terms in each equation, and additional lags (e.g., $a_2 \cdot Y_{t-2}$, etc.), but I'll leave them out here to simplify the notation. We will assume that the error terms are uncorrelated ($\mathbf{Cov}(\varepsilon_{X_t}, \varepsilon_{Y_t}) = 0$).

This is a (very simple) 'structural' VAR. Now, if both a_0 and c_0 were zero, then we could estimate each equation separately using OLS, since the values of X and Y at time t should not be affecting the values of the regressors which were determined in the past (specifically in period $t-1$). However, if both a_0 and c_0 are non-zero, then each of the two equations has an endogenous regressor in it: the equation for X_t has Y_t as a regressor, and the equation for Y_t has X_t as a regressor. This generates the same *identification problem* that we have when we try to estimate supply and demand equations. Essentially, we have a problem with reverse causality: Y_t causes X_t according to the first equation, but at the same time there is a separate process by which X_t causes Y_t according to the second equation. If we regress one on the other, we will pick up both sets of causation (like we would if we tried to fit a scatter of equilibrium points generated by shifts in both supply and demand).

However, if we can assume that *either* a_0 *or* c_0 is equal to zero, then we *can* estimate both equations by OLS. E.g., if we can assume that a_0 is zero, then this assumption ($a_0 = 0$) is sufficient to allow us to identify all of the remaining coefficients in the system. Note that this assumption says that Y affects X only with a lag (via a_1), and not contemporaneously, whereas X affects Y both contemporaneously (via c_0) and with a lag (via c_1).

Suppose now that we have three variables (X, Y, and Z) rather than two. Then each of the three regression equation has two endogenous regressors in it (e.g., the equation for X_t has both Y_t and Z_t on the right hand side). We can identify this system of equations if we can assume that one of the three variables (say X) is affected by neither of the others contemporaneously, one (say Y) is affected by only one other contemporaneously, and the third (Z) is affected by both of the others contemporaneously. This would establish a recursive structure like that in the two variable case that would allow us to identify all the remaining parameters by OLS.

This procedure, look for an ordering of the variables in terms of contemporaneous causality is one way of solving the identification problem (it was advanced by Sims (1980)). There are other identification techniques that have also been suggested and are used by various researchers, but all require putting restrictions on the coefficients - and so all end up being not completely agnostic about theory - i.e., we need to appeal to theory at least to some degree to determine if any recursive ordering (or more generally, any set of identifying restrictions) makes theoretical sense.

If we can come up with identifying restrictions that we are happy with, we can then estimate the (remaining) coefficients of the regression model and the errors (ε_{X_t}, ε_{Y_t},...), which we can interpret as shocks to each variable. We can also use the fitted equations to plot the predicted effect of a shock to any one variable on each of the variables over time. This is called an *impulse response* plot. For example, in the two variable case above, if $a_0 = 0$, then a positive value of ε_{Y_t} (say set to one standard deviation of the estimated shock series) today (time t) will cause Y to increase directly today, but leave X unchanged today. However, the increase in Y today will cause X to increase tomorrow (as Y_t enters the equation for X_{t+1}), and both the increase in X_{t+1} and Y_t will increase Y_{t+1}, so X and Y will feed back on each other over time.

Blanchard's Paper:

An alternative approach is not to try to identify the coefficients in the structural VAR (i.e., the equations above), but rather drop all contemporaneous variables out of the right hand sides of the equations (without assuming that they are truly zero). This amounts to estimating the coefficients of the reduced form of the system (i.e., substitute out X_t and Y_t from the RHSs of the two equations above). However, we are not then estimating the structural coefficients (rather we are estimating composites of the structural coefficients), and we don't recover estimates of the direct shocks either (the residuals in the reduced form regression are composites of all the shock terms).

Blanchard (1993) takes this approach and then attempts to clean up these composite shocks, to extract something closer to the direct shocks, rather than trying to impose identifying restrictions directly on the structural VAR.

References:

Sims, Christopher, "Macroeconomics and Reality," *Econometrica*, **48**, January 1980, 1–49.

Blanchard, Olivier, "Consumption and the Recession of 1990–91," *American Economic Review*, May 1993, 270–274.