Economist's View
Mark Thoma

Monday, October 10, 2011

The Nobel Prize in Economics: A Note on Chris Sims' Contributions

Let me talk a bit about Sims contributions to economics, and if I have time I'll try to cover Sargent later.

Prior to Sims work, in particular his paper “Macroeconomics and Reality,” the state of the art in macroeconometrics was to use large-scale structural models. These models often involved scores or even hundreds of equations, essentially a S=D equation for every important market, identities to make sure things add up correctly, etc. But in order to estimate the parameters of these models, the structural parameters as they are known, you had to overcome the identification problem.

Without getting into the details, the identification problem essentially asks if its possible to estimate the structural parameters at all. The answer, in general, is no. For example, if every variable in the model appears in every equation, then it won't be possible to estimate the structural model. Let me give an example to illustrate. Suppose that X and and Y are the endogenous variables, e.g. price and quantity for some market, and that the structural model is:

$$Y_t = a_0 + a_1X_t + a_2Y_{t-1} + a_3X_{t-1} + u_t$$

$$X_t = b_0 + b_1Y_t + b_2Y_{t-1} + b_3X_{t-1} + v_t$$

The a's and the b's are the parameters that economists are generally interested in, but in this form it is not possible to estimate them. There must be what are known as exclusion restrictions before estimation is possible. In this case, for example, identification can be achieved by making either a_1 or b_1 equal to zero (more on this below), i.e. excluding one of the variables from one of the equations. If there is a reason for this, then excluding the variable is okay, but a variable can't be left out simply to achieve identification -- there must be good reason for excluding X_t from the first equation, or Y_t from the second (or both). Omitting a variable that ought to be in a model in order to satisfy the identification restrictions results in a misspecified model and biased estimates.

In large models, these exclusions are numerous, and many researchers simply assumed whatever exclusion restrictions were needed to achieve identification, and then went on to estimate the model. In “Macroeconomics and Reality,” Sims pointed out the problem with this approach. The assumptions that researchers were imposing to achieve identification had no theoretical basis. They were ad hoc and difficult to defend (especially when expectations are in the model -- expectations tend to depend upon all the variables in a model making it difficult to exclude anything from an equation involving expectations).

What Sims suggested as an alternative was to drop structural modeling altogether, and to use generalized reduced forms as the basis for estimation. There would be no hope of recovering structural parameters in most cases, but there was still much that could be learned by using reduced forms instead of structural models.

For example, the reduced form for the model above is (you can find the reduced form by expressing the endogenous variables X_t and Y_t in terms of exogenous and predetermined variables):

$$X_t = [1/(1-a_1b_1)]\{(a_0 + a_1b_0) + (a_1b_2 + a2)Y_{t-1} + (a_1b_3 + a_3)X_{t-1} + a_1v_t + u_t\}$$

$$Y_t = [1/(1-a_1b_1)]\{(b_0 + b_1a_0) + (b_1a_2 + b_2)Y_{t-1} + (b_1a_3 + b_3)X_{t-1} + v_t + b_1u_t\}$$

To estimate this, write it as:

$$X_t = c_0 + c_1Y_{t-1} + c_2X_{t-1} + a_1v_t + u_t$$

$$Y_t = d_0 + d_1Y_{t-1} + d_2X_{t-1} + v_t + b_1u_t$$

This is a VAR model. At first, Sims thought we could draw important conclusions from this model, e.g. suppose that X is money and Y is output. Then this model could tell us how a shock to money would change output over time (these are called impulse response functions -- you hit the system with a shock, and then use the estimated model to trace out the path of the endogenous variables over time). We could use this model to answer important questions such as whether money causes output (Sims' technique for testing causality was essentially the same as Granger causality, but Sims' made an important contribution in extending the causality techniques to systems with three or more variables when he introduced impulse response functions and variance decompositions).

But, as Cooley and LeRoy pointed out in an important paper, these models don't avoid structural assumptions after all, at least not if you want to say anything about how variables in the model respond to structural shocks. To see this, note first that the shock we are interested in is the shock to money, v_t. Now look at the errors in the two reduced form equations. We can estimate each equation by OLS, and when we do the error terms will be estimates of $a_1v_t + u_t$ for the first equation and $v_t + b_1u_t$ for the second. Thus, we get estimates of linear combinations of the v_t and u_t shocks we are interested in, but we don't get the shocks in isolation like we need. And there's no way to isolate the shocks, i.e. to determine their individual values. That's a problem because we need to find the money shock alone if we want to estimate its effect on output.

How can we do this? One way is to make either a_1 or b_1 equal to zero. Let's set $b_1=0$ because that's the easiest to discuss. In this case, when we estimate the second equation by OLS (the equation with the d parameters), the error will now be an estimate of v_t, which is just what we need. However, notice that this is nothing more than an exclusion restriction -- by assuming that $b_1=0$, we are excluding Y_t from the second equation (see the structural model). Thus, we have come full circle.

This is where Sims Structural VARS come into play. The reduced form above is known as a VAR model (in its estimable form, i.e. the second set of reduced for equations above involving the c and d parameters). It turns out that if we can often defend particular restrictions theoretically, e.g. if money can only respond to output with a lag, perhaps due to information problems, then there is no reason to have the contemporaneous value of output on the right-hand side of the structural equation for money, i.e. this implies that $b_1=0$.

Thus, while this still amounts to an exclusion restriction, the restriction is no longer ad hoc -- simply imposed as necessary to achieve identification as back in the old, large-scale structural model days -- it is grounded in theory. And the fact that we insist these restrictions be grounded in theory marks an important difference from the work that came before Sims.

And even better, this technique also allows the model to be identified without using exclusion restrictions at all. For example, if we think that some variables in the model have short-run but not long-run effects, e.g. that money can affect output in the short-run, but only produces price effects in the long-run -- a standard assumption in most macro models -- then the zero impact in the long-run can be imposed as an identifying restriction. Exclusion restrictions won't be needed (this is the Blanchard-Quah and Shapiro-Watson techniques).

This just scratches the surface of Sims' work -- I wish I had time to do more -- but *hopefully* this provides a window into one part of Sims' contributions.

Tuesday, October 11, 2011

Christopher Sims and Tests for Causality

To tell the full story of Christopher Sims' contributions to causality, we need to go back to the state of the art in policy evaluation in the 1960s, in particular, to something known as the St. Louis equation:

$$Y_t = c + a_0M_t + a_1M_{t-1} + a3M_{t-2} + b_0G_t + b_1G_{t-1} + b_2\, G_{t-2} + e_t$$

In this equation, output (Y) is regressed on current and lagged values of money (M) and government spending (G). The idea was to see how output responded historically to changes in money and government spending, and then use these estimates to guide policy. If we know how Y responds to M, then we can use that knowledge to set monetary policy optimally.

Now, there is a fundamental problem with this approach highlighted by the Lucas critique (the negative reaction to the other common approach, using large-scale structural models to evaluate policy, was discussed yesterday). If you change monetary policy you also change the values of the a and b coefficients so that the estimates are no longer reliable, and hence no longer a guide, but that criticism came later. At the time there was another worry.

The worry was something known as simultaneity bias. Consider the M_t term in the equation above. If M_t is "econometrically exogenous," i.e. if it doesn't depend upon Y_t, then the estimated value of a_0 will be unbiased. But if M_t depends upon Y_t , perhaps through and equation such as $M_t = h_0 + h_1Y_t + u_t$, then the estimate of will be biased and hence a poor guide to policy decisions.

The first use of causality tests was to test to see if h_1 in the "policy equation" was equal to zero, and Sims was a key player in the development of these tests. Thus, Sims starts his 1972 AER paper with:

> This study has two purposes. One is to examine the substantive question: Is there statistical

> evidence that money is "exogenous" in some sense in the money-income relationship? The other is to display in a simple example some time-series methodology not now in wide use. The main methodological novelty is the use of a direct test for the existence of unidirectional causality.

If there was unidirectional causality from M to Y, then the estimate would be unbiased. But if there was two-way causality, i.e. if Y causes M (h_1 is not zero), then the estimate would be problematic.

Sims contributed greatly to this literature, and once this work was largely complete, it quickly became clear that these tests could be used to assess causality more generally, the method was not limited to checking for econometric exogeneity.

But there was also a problem. The basic technique (an F-test on a set of coefficients) to test for causality worked well on 2-variable systems, but it didn't work reliably for systems with three or more equations (the problem was that X can cause Y, and Y can then cause Z so that there is a causal path from X to Z, but the F-test approach will miss this).

Sims Second major paper on causality addresses this problem by providing two new tools to assess causality, impulse response functions and variance decompositions (along the way it was also shown that Sims and Granger causality are equivalent). Impulse response functions, which have since become a key analytical device in macroeconomics, trace out the response of the variables in the model to a shock to another variable in the system (identification restrictions are needed to ensure that the shock is actually a policy shock, see here). If the variable, say output, responds robustly to a shock to, say, the federal funds rate, then we say that the federal funds rate causes output. But if we shock the federal funds rate and output essentially flat-lines in response, then causality is absent.

However, even when there is causality according to the impulse responses, impulse response functions do not tell us how important one variable is in explaining the variation in another variable (the impulse response function could look impressive, but it may be that we are only explaining 1% of the total variation in the other variable so that the response we are seeing is not very important in explaining why the other variable fluctuates over time). Variance decompositions solve this problem. They don't tell you the sign/pattern of the response like impulse response functions do, but the do give an indication of how important one variable is in explaining the variation in another variable (e.g. if M explains 75% of the variance in output, that's impressive and notable, but if it's only 1% then money isn't very important in explaining why output changes over time).

Sims second paper also made another important point. In his first paper, he found that money causes output (so it could not be treated as econometrically exogenous as in the St. Louis equation). But that was in a two-variable system including only M and Y. In his second paper he adds interest rates (i) and prices (P) to get a four variable system, and he finds that this overturns the results in his first paper. Once i is added to the model, M no longer causes Y. Thus, the lesson is that if you leave important variables out of a VAR system, it can produce misleading results.

But Sims' main contributions were, initially, the F-tests for testing causality in bivariate systems, and the addition of IRFs and VDCs to assess causality in higher order systems. In addition, he also provided many of the common "pitfalls of causality testing," -- causality testing can be misleading in a number of ways. One is above, leaving a variable out of a system. If A causes B to change tomorrow, and C to change the next day, a system containing only B and C will look as though B causes C when in fact there is no causality at all, a third variable causes both. Other pitfalls can occur, for example, when there is optimal control or when expectations of future variables are in the model. Identifying the pitfalls of the methods he (and others) developed was also an important contribution to the literature.

Sims' work on causality was highlighted in the Nobel announcement, and I hope this provided some background on this topic. But there's a lot more to be said about Sims' work over and above his work on causality testing discussed above and his work on structural VARs I discussed yesterday, e.g. his recent papers on rational inattention, and I hope to write more about both Sims and Sargent when I can find the time.