Econ 556b Spring 1999 C.Sims

Granger Causality

1. Definitions

Definition 1. X *does not Granger-cause* Y (not $X \,\mathrm{GC}\, Y$, or $X \sim\!\mathrm{GC}\, Y$) iff prediction of Y based on the universe U of predictors is no better than prediction based on $U - \{X\}$, i.e. on the universe with X omitted.

When applied to a VAR model with reduced form

$$\left(I - \begin{bmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{bmatrix} \right) \begin{bmatrix} y(t) \\ x(t) \\ z(t) \end{bmatrix} = \varepsilon(t) \,, \tag{1}$$

where the universe U is interpreted as consisting of past values of x, y and z, the definition specializes to stating that $x \sim\!\mathrm{GC}\, y$ iff $B_{12} = 0$.

This definition does not provide a statistical magic wand that allows us to discover true causal structures via data analysis, without substantive theory. It is best thought of as an attempt at specifying a necessary condition for a causal relation. Even as a necessary condition, it has its problems. The way we usually think about causality suggests that "x causes y" should be transitive. That is, if x causes y and y causes z, then x causes z. Granger's definition is not transitive. If $B_{13} = 0$ but the rest of the B matrix in (1) is non-zero, then z causes x, x causes y, but z does not cause y. However, Granger causality can be the basis for defining a transitive relation.

Definition 2. x is *Granger Causally Prior* (or GCP) to y in (1) iff it is possible to group all the variables in the system into two blocks, Y_1 and Y_2, such that y is in Y_2 and x is in Y_1 and Y_2 does not Granger-cause Y_1.

If x, y, and z in (1) are scalar variables, then z is GCP to y iff either $B_{31} = B_{32} = 0$ or $B_{31} = B_{21} = 0$. x is GCP to y iff either $B_{31} = B_{21} = 0$ or $B_{21} = B_{23} = 0$. In the latter case, the block triangular structure of B would not be apparent unless we re-ordered the variables to put x at the bottom of the vector.

It may be interesting to note that $x \,\mathrm{GCP}\, y$ implies $y \sim\!\mathrm{GC}\, x$, and furthermore among all transitive relations $\frown$ with the property that $x \frown y$ implies $y \sim\!\mathrm{GC}\, x$, GCP is the strongest (That is, it is, the collection of pairs for which it is true is largest.) This result was first obtained in unpublished work by Thomas Doan.

2. Causality and Exogeneity

Granger causal priority plays a big role in applied time series work in good part because it makes precise the sense in which putting variables on the right hand side in a regression is justified by claims that what is on the right is causally prior.

Definition 3. x is *strictly exogenous* in the regression equation

$$y(t) = C(L)x(t) + \nu(t) \,, \tag{2}$$

iff

$$E[\nu(t)\,|\,x(s),\ \text{all } s] = 0 \,. \tag{3}$$

Note that no claim is made here about serial dependence in ν.

Strict exogeneity is one of the standard assumptions made about regression equations in econometrics. It underlies the classical claims for efficiency of GLS estimation in single equation models. It turns out that if there is a regression equation of the form (2) with $C_s = 0$ for all $s < 0$ and x strictly exogenous, and if y and x have a joint representation as a VAR, then x is GCP to y in their VAR representation. Furthermore, if z and y are both part of a VAR system (possibly including other variables as well) and z is GCP to y in this system, then there is an equation of the form (2), with x strictly exogenous, $C_s = 0$ for $s < 0$, and z a subvector of the x vector.

3. Testing recursive restrictions

Suppose we have a system consisting of two blocks of regression equations that we can write

$$\left(I - \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix}\right) \begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} X(t) + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix} \tag{4}$$

$$\left.\begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}\right| \left\{\beta, \{Y_1(t-s), Y_2(t-s)\}_{s=1}^{\infty}, \{X(s)\}_{s=-\infty}^{\infty}\right\} \sim \text{ i.i.d. } N(0,\Sigma) \,, \tag{5}$$

where $B_{ij}(s) = 0$ for $s < 0$. Because this system has the same explanatory variables (X's and lagged Y's) in each block, its maximum likelihood estimates can be found by OLS and its likelihood function (using the conditional distribution of the data given initial conditions for Y and the entire X sequence) is proportional to $|S_T|^{-T/2}$, where S_T is the sample covariance matrix of residuals. The algebra underlying this result is not trivial, but it is explained in detail in standard econometrics textbooks under the heading "seemingly unrelated regressions".

It is useful to rewrite (4) as

$$\begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} Z(t) + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix} \,, \tag{6}$$

where $Z(t)$ collects the explanatory variables on the right-hand side of (4) and β_i collects the coefficients on explanatory variables of equation block i in (4).

Now suppose we want to consider a restriction of the form

$$R\beta_2 = \gamma \,. \tag{7}$$

Because this is a linear restriction, estimation of the Y_2 equations in (4) subject to the restriction can be formulated as estimation of the Y_2 equations with a shorter list of right-hand-side variables than appears in the Y_1 equations. This of course undoes the condition that allowed us to claim that OLS equation by equation on (4) delivers ML estimates. However, it turns out nonetheless not to be necessary to abandon OLS estimation, if we appropriately modify the estimation equations.

Because $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are jointly normal, we can write

$$\varepsilon_1(t) = \gamma\varepsilon_2(t) + \nu(t) , \tag{8}$$

where $\nu(t)$ is i.i.d., normally distributed, and independent of $\varepsilon_2(t)$. We can use this relationship to replace the Y_1 equations with linear combinations of the Y_1 and Y_2 equations that have $\nu(t)$ as their disturbance vector rather than $\varepsilon_1(t)$. That is, we can rewrite the system as

$$\begin{bmatrix} I & -\gamma \\ 0 & I \end{bmatrix} \begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} = \begin{bmatrix} \beta_1 - \gamma\beta_2 \\ \beta_2 \end{bmatrix} Z + \begin{bmatrix} \nu(t) \\ \varepsilon_2(t) \end{bmatrix} . \tag{9}$$

If we let $\phi = \beta_1 - \gamma\beta_2$, the system can be written as

$$Y_1(t) = \gamma Y_2(t) + \phi Z(t) + \nu(t) \tag{10}$$
$$Y_2(t) = \beta_2 Z(t) + \varepsilon_2(t) . \tag{11}$$

The fact that ϕ can be written as $\beta_1 - \gamma\beta_2$ places no restrictions on ϕ, even if β_2 is restricted. Whatever the values of β_2 and γ, if β_1 can range over the whole of $\mathbb{R}^{m_1}$, so can ϕ.

The residual $\nu(t)$ in (10) is Gaussian with zero mean conditional on the right-hand-side variables. Thus the likelihood for the full sample of $Y(t)$, $t = 1, \dots, T$ factors into two components, one for the Y_1 block involving only ϕ and γ, and one for the Y_2 block involving only β_2. So ML reduces to ML separately within each block. Within the Y_1 block right-hand-side variables are the same in all equations, so the covariance matrices of the block residuals, even if they are restricted, do not affect the fact that ML estimates of ϕ and γ are obtained by equation-by-equation OLS. The same may be true of the Y_2 block, depending on the nature of the restrictions on it. If the restrictions are exclusion restrictions (setting elements of β_2 to zero), then for OLS to be MLE on the second block requires that any variable excluded from any equation in the block be excluded from all the equations in the block.

We have now reparameterized the system without changing the space of models being considered. That is, there is a one-one mapping between the parameters of (10-11) and those of (4). So it must be that the maximum of the likelihood of these two representations of the system match. That is, if we set $\Omega = \mathrm{Var}(\nu(t))$, it must be that

$$\left|\hat{\Omega}\right|^{-\frac{1}{2}} \left|\hat{\Sigma}_{22}\right|^{-\frac{1}{2}} = \left|\hat{\Sigma}\right|^{-\frac{1}{2}} , \tag{12}$$

where the hats denote MLE's.

But now we see that, because our restrictions affect only the lower block of equations, the ML estimates of the upper block (10), including $\hat{\Omega}$, are unaffected by restrictions that may be imposed on β_2. So the log likelihood ratio for the restricted vs. unrestricted model is just

$$-\frac{1}{2}(\log\left|\hat{\Sigma}_{22}^{R}\right| - \log\left|\hat{\Sigma}_{22}^{U}\right|) , \tag{13}$$

where the U and R superscripts refer to unrestricted and restricted ML estimates, respectively. We arrive at the conclusion, therefore, that the likelihood ratios needed to test restrictions that affect the coefficients of the Y_2 equations alone can be formed from restricted and unrestricted ML estimates of those equations by themselves. There is no need to make any use of estimates of the parameters or residual covariance matrices of the Y_1 equations.

To test Granger causal priority of Y_2, therefore, one can simply estimate the Y_2 equations by OLS, with and without lagged Y_1 variables included, and compare the resulting residual covariance matrices. Use of OLS is justified, because a Granger causal priority restriction does by construction exclude variables from all the equations of the Y_2 block at once.

Note that this does not mean that, with the GCP restrictions imposed, one should use OLS estimates of (4) for forecasting or for reporting results. The OLS estimates of β_1 are inefficient if β_2 is subject to restrictions. If for some reason one needs the system in the form (4) rather than (10)-(11), the best way to form estimates is to apply OLS to (10)-(11), then transform the estimated parameters back to those of (4).