Causal Orderings and Exogeneity

April 19, 2018

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 This result was first obtained in unpublished work by Thomas Doan.

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

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

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.

- $x(t)$ exogenous is generally impossible when $x(t)$ contains lagged dependent variables, and thus is an assumption usually made when the x and y variables are completely distinct. Exogeneity of x is the main assumption required in proofs that GLS is the best estimator for C.

Theorem 1 (GCP and exogeneity equivalence)**.** *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.*

Predeterminedness

Definition 4. x *is* **predetermined** *in* (2) *iff* $E[\nu(t) \mid \{x(s), y(s-1) \mid s \leq t\}] = 0$.

- While exogeneity is an interpretable assumption even if $C(L)$ in (2) is two-sided (i.e. has $C_s \neq 0$ for some $s < 0$), predeterminedness only makes sense when $C(L)$ is one-sided.

- Predeterminedness implies absence of serial correlation in ν_t (because lagged ν_t's are functions of lagged x's and y's).

- Predeterminedness of x, plus standard regularity conditions, are enough to deliver the usual normal asymptotic theory for OLS estimates of C, with asymptotic covariance matrix of the usual $\sigma^2(X'X)^{-1}$ form.

- Once it is assumed that $\varepsilon(t)$ is serially independent, exogeneity is a stronger assumption than predeterminedness.

- In fact exogeneity by itself, without any claim about serial correlation in ε, under standard regularity conditions delivers consistency and asymptotic normality of OLS estimators and justifies use of GLS to obtain efficient estimators when $\mathrm{Var}(\varepsilon(t)$ is not of the $\sigma^2 I$ form.

- Assuming x is predetermined does not justify GLS or any other procedure (e.g. differencing the data) that takes linear combinations of observations and assumes that the transformed data still satisfy the original regression equation.

Weaker still

- There is a weaker assumption than either of these that implies consistency of OLS under standard regularity conditions: $E[\nu(t) \mid x(t)] = 0$.

- This assumption does not imply absence of serial correlation in ν and does not justify GLS.

- Older econometrics textbooks do not discuss this case; indeed they often define predetermined variables as just exogenous variables and lagged dependent variables.

- Rational expectations theory does sometimes deliver equations that are of this form, however. Suppose we have data on y_t and on forecasts of F_t of y_t, where the forecasts are made k periods in advance. If the forecasts are minimum-variance, then $E_{t-k}[y_t - F_t] = 0$. The regression equation $y_t = a + bF_t + \varepsilon_t$ will satisfy $E[\varepsilon_t \mid F_t] = 0$ at the true parameter values $a = 0$, $b = 1$. Test $a = 0$, $b = 1$, to test "rationality".

- The theory does not imply that the ε_t series is uncorrelated across time. Under the usual assumption that all relevant lagga of y_t are in the information set at t, the ε's will be uncorrelated at lags greater than k, but not at lags shorter than k.

- F is not exogenous. GLS is not justified. F is also not predetermined, so that the usual OLS formulas for the covariance matrix are not justified.

- Nonetheless OLS is consistent. Developing measures of uncertainty for the OLS estimates requires additional restrictive assumptions.