

**CAUSAL ORDERINGS; ACF AND IMPULSE RESPONSES AS
DATA SUMMARIES**

1. DEFINITIONS

Definition 1. *X does not Granger-cause Y (not X GCY, or $X \sim$ GCY) 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

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

where the universe U is interpreted as consisting of past values of x , y and z , the definition specializes to stating that $x \sim$ 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

Date: November 25, 2001.

©2000 by Christopher A. Sims. This material may be reproduced for educational and research purposes so long as the copies are not sold, even to recover costs, the document is not altered, and this copyright notice is included in the copies.

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 $x \text{ GCP } y$ implies $y \sim \text{GC } x$, and furthermore among all transitive relations \sim with the property that $x \sim y$ implies $y \sim \text{GC } x$, 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

$$(2) \quad y(t) = C(L)x(t) + \nu(t) ,$$

iff

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

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.

Theorem 2.1 (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.*

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

- Predeterminedness of x in (2) is a standard minimal assumption in proofs that OLS is consistent and asymptotically normally distributed. It is clearly weaker than exogeneity. 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.
- $x(t)$ predetermined implies the absence of serial dependence in $\nu(t)$. Predetermined $x(t)$ is the main assumption required in proofs that in some sense least squares is the best estimator for C .

- $x(t)$ 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 B .

3. ENGLE, HENDRY AND RICHARD EXOGENEITY

The terms “predetermined” and “exogenous”, were in wide use in econometrics before Engle, Hendry and Richard 1983 (henceforth EHR) extended the terminology. The EHR notion of **weak exogeneity** in its general form makes no reference to equations or residuals. It is concerned entirely with conditions under which certain parameters of the joint distribution of two variables y and x can be estimated without loss of information from the conditional distribution of $y|x$ alone. But specialized to our case of linear regression with $\nu(t) \sim N(0, \Sigma)$, weak exogeneity becomes the requirement that

Definition 5 (Weak exogeneity in regression).

- (i) $\nu(t)$ is independent of $\{x(s), y(s-1), \text{ all } s \leq t\}$, for all t , in (2),
- (ii) the marginal distribution of $\{x(s), y(s-1), s \leq 1\}$ does not depend on any unknown parameters included in C, Σ , and
- (iii) the pdf of $x(t) | \{x(s), y(s), s < t\}$ does not depend on C or Σ .

- Clearly condition (i) here is stronger than predeterminedness, as defined above, and equivalent to it under the assumption of joint normality of the x and y processes. Predeterminedness and exogeneity are used in different senses in different contexts, however. It is not uncommon to label the stronger requirement (i) “predeterminedness”, or instead to use the term for the weaker condition $\mathcal{E}[\nu(t) | \{x(s), y(s-1) s \leq t\}]$, where the \mathcal{E} operator stands for “best linear predictor”. However in the definition of weak exogeneity, only the form given in (i) suffices.
- Under conditions (i)-(iii), the likelihood for all the data on x and y over the whole sample factors into two pieces, one of which is the usual Gaussian regression likelihood, the other a function of x 's and of parameters other than C, Σ . If the prior also makes C, Σ independent of the other parameters, the posterior pdf factors in the same way as the likelihood, justifying inference based on the regression likelihood alone.

Specialized to our regression model, EHR's notion of **Strong Exogeneity** becomes

Definition 6 (Strong exogeneity in regression).

- (i) $x(t)$ is independent of $\nu(s)$, all t, s ; and
 - (ii) the pdf of the sample $\{x(s), s = 1, \dots, T\}$ does not depend on C, Σ .
- Under the strong exogeneity condition, the likelihood factors into a marginal pdf for the observed x 's and a conditional pdf for the observed y 's given the x 's, with only the latter involving the C, Σ parameters.
 - If in addition to the conditions above we have that the prior pdf over all the unknown parameters makes C, Σ independent of any other unknown parameters, then the posterior also factors into a piece relating only to the parameters of the x process and another depending only on C, Σ .
 - EHR argued correctly that these notions were important, because econometricians were used to justifying single-equation estimation methods based on claims about exogeneity and predeterminedness, without careful attention to model parameterization, which could lead to mistakes.
 - However they presented their arguments in such a way that it might appear that predeterminedness and exogeneity, in their old senses, were mistaken notions in themselves. This is not true. It is useful to be able to discuss whether, say, "price is predetermined in the demand equation" as a property of the actual demand equation, without regard to how we have parameterized it or how our beliefs about its parameters relate to beliefs about other parameters. EHR are correct in pointing out that after we have decided what we think about predeterminedness, we still have work to do in deciding whether single-equation estimation is justified. But these considerations can usefully separated from the question of whether the predeterminedness assumption itself is justified.

REFERENCES

ENGLE, R. F., D. F. HENDRY, AND J.-F. RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.