Exogeneity and Causal Ordering in Macroeconomic Models

by Christopher A. Sims

This paper aims to clarify the interpretation of exogeneity tests of the type Sargent [129]† and I [126] among others have applied, as well as the closely related tests proposed or used by Caines and Chan [12], Granger [40], Pierce [119], Haugh [54], and others.‡

I have found that disquisition on this topic is capable of generating fierce or long-winded responses from economists. A good part, though not all, of this sort of dispute is purely semantic. Words like "causal," "exogenous," "structural," and "behavioral" inevitably come into use in the discussion, and economists appear to have strong and divergent opinions on the "natural" or "standard" meanings of those terms. Because these words do crop up, this paper sets out in section 2 to define them. The definitions in section 2 attempt to parallel the usage of earlier writers who have given these terms precise meaning—for example, Simon [132], Koopmans and Bausch [77], Hurwicz [63], Granger [40], and, in engineering literature, Zemanian [169]. However, many readers are likely to find the terminology of section 2 artificial or strange. While I am happy to discuss§ any possible alternative equally precise semantic scheme, the reader who finds section 2 irritating or opaque should recognize that it is purely instrumental. We need precise definitions of terms in order to discuss, in section 3, a variety of possible types of economic models which imply the existence of one-way Granger causal orderings. Interpreting the results of exogeneity tests when the null hypothesis of exogeneity is

†Numbers in [ ] correspond to reference list, p. 219.

‡This paper has changed substantially since the draft which was presented at the November 1975 Conference. Section 1 expands a few paragraphs which appeared previously in the middle of the paper without special emphasis. In section 2 the notion of a "behavioral" relation is new. In section 3 the "errors in variables" and "perfect market" mechanisms for generating causal orderings were not discussed in the earlier draft, and the discussion of particular empirical examples is new. Thus, the discussants' remarks could not have dealt with these aspects of the paper.

§For about 15 minutes.

accepted requires consideration of which type of economic behavioral mechanism capable of generating exogeneity is likely to have done so in the instances at hand.

Section 1 makes a point without any preliminary expedition through the semantic swamps. When, as econometricians estimating models ordinarily do, someone asserts that a particular variable or group of variables is strictly exogenous in a certain regression, that assertion is, in time series models, testable. "Exogeneity" here is given its standard econometrics textbook definition. Exogeneity tests are thus an easily applied test for specification error, powerful against the alternative that simultaneous-equations bias is present. The usefulness of these specification error tests ought not to be controversial and in particular is not connected to the controversial and complicated analysis of section 3. Section 3 discusses how we are to interpret the result that exogeneity is an acceptable null hypothesis. Section 1 discusses the simpler question of how we interpret rejection of that null hypothesis.

Probably most readers will prefer to look over sections 1 and 3 first, returning to section 2 only if section 3 looks interesting enough to justify the effort.

I. Exogeneity Tests as Specification Error Tests

It is a maintained hypothesis underlying the Gauss-Markov theorem and the distribution theory ordinarily applied to generalized least squares estimators that the right-hand-side variables in a regression equation are strictly exogenous, meaning that the expected value of the vector of residuals, conditional on the whole array of right-hand-side variables, is zero. To restate this mathematically, if our model is

$$y = Xb + u \tag{1}$$

with y $T\mathrm{x}1$, X $T\mathrm{x}k$, u $T\mathrm{x}1$, the hypothesis that X is strictly exogenous is the hypothesis that $E[u|X] = 0$. Sometimes this "first order independence" assumption is strengthened to complete independence of X and u, but the first-order version will be enough for us. If exogeneity holds for this model with sample size up to $T+s$, then we can add to the right-hand side of (1) the variable Z, whose t'th component is the $t+s$'th component of X, to get

$$y = Xb + Zc + u. \tag{2}$$

On the null hypothesis that (1) satisfies the assumptions of the Gauss-Markov theorem, (2) does also, with $c = 0$. Testing $c = 0$ by standard methods thus tests the null hypothesis of strict exogeneity of X in (1).

The argument so far has made no use of the notion that (1) should be a time series model. However, the test has good power against reasonable alternatives to exogeneity only if those alternatives make it likely that

estimates of c in (2) will be significantly non-zero. In some cross-section applications, the order of observations has no behavioral significance. Hence, we have good reason to believe that the t'th vector of observations on y and X, $(y(t), X(t))$, is independent of the s'th such vector for any $t \neq s$. Thus if, in the behavioral relation we purport to be estimating, $E[u|X] \neq 0$, the only possible route for dependence between u and X is dependence between $X(t)$ and $u(t)$. Dependence between $X(t)$ and $u(t)$, when $u(t)$ is serially uncorrelated, will not generate a non-zero c in (2). In time series or in cross-section models where the ordering of observations is non-random, however, failure of exogeneity is likely to produce non-zero c and the test will be effective. Of course, it is still important to understand what special circumstances might make c zero in (2) even though X is not exogenous in the behavioral relation under investigation. This issue is taken up in section 3.

As Shiller points out in his discussion of this paper, least-squares regression estimates and the usual associated asymptotic distribution theory can be justified without a strict exogeneity assumption. However, these alternative stochastic specifications (making the right-hand side "predetermined" rather than exogenous) ordinarily rest on a different sort of *a priori* argument than exogeneity, and they also generally imply a different battery of specification error tests. Thus if a model fails an exogeneity test, it is not natural immediately to claim that the model is instead justified by an assumption that the right-hand-side variables are predetermined. A new behavioral argument would be required to justify the predeterminedness assumption, and a new specification error test suggests itself.†

For example, suppose we are estimating the supply function for wheat in the United States. Quantity is regressed on price, under the argument that the wheat market is international, so that price fluctuations depend on world supply and demand, which are not strongly related to U.S. supply. An exogeneity test rejects the null hypothesis that price is exogenous. The natural conclusion is that world supply and demand do depend importantly on U.S. supply, or that U.S. supply is at least related to some of the same omitted variables as world supply and demand. But why not now claim that supply depends on last period's price and that last period's price is predetermined? This requires a new behavioral argument: first, that there is no possibility of within-year supply response to price; then, that omitted variables influencing supply and demand are unrelated; and finally, that omitted variables which influence supply (that is, the residuals in the supply equation) are serially uncorrelated. The lack of serial correlation in supply equation residuals would appear to be a dubious *a priori* assumption, but it is a testable assumption; and testing for serial correlation in this specification plays the same central role as testing for

†Phillips [116] pointed out a long time ago the connections between assumptions on serial correlation properties of error terms and on orders of lag in behavioral equations which are required to justify a claim that right-hand-side variables are predetermined.

exogeneity in the first specification.

II. Semantics

A. Causal Orderings as Recursions

Using the central idea of Simon [132] and generalizing it slightly, consider a space S of "outcomes" and two sets of restrictions on it characterized by the subsets A and B of S. A and B together produce the "result" $A \cap B$. Now consider two spaces X and Y together with associated functions P_X, P_Y mapping S into X and Y respectively. The following definition covers many existing uses of "causal ordering" as special cases:

> *Definition:* The ordered pair (A,B) of restrictions on S determines a causal ordering from X to Y (equivalently, makes X causally prior to Y) if and only if $P_X(A \cap B) = P_X(A)$ and $P_Y(A) = Y$.

Paraphrasing, (A,B) makes X causally prior to Y if and only if A restricts X (if at all) without restricting Y, while the addition of B restricts Y (if at all) without further restricting X.

It should be clear that in this definition the causal ordering is a characteristic of the system (A,B) and the output space Y, not of the result $A \cap B$. Given a system (A,B) which does not make X causally prior to Y, we can always define $B' = A \cap B$, $A' = P_X^{-1}(P_X(A \cap B))$, and $Y' = P_Y(A)$, and (A',B') will by construction make X causally prior to Y', yet (A',B') has the same result as (A,B).

An example of causal ordering is a pair of linear equations in two unknowns, one of which involves only one unknown. The space S is Euclidean 2-space, X and Y are two copies of the real line, P_X projects a point in S into its first co-ordinate, and P_Y projects a point in S into its second co-ordinate. The set A is the line determined by the first equation, which involves only X (a vertical line, if X is the horizontal axis), and B is the line determined by the second equation.

Another example is a two-simultaneous-equation econometric model in Wold causal chain form. Here S is the space of joint distributions of the endogenous variables conditional on the predetermined variables; X is the space of marginal distributions (again conditional on predetermined variables) for the first endogenous variable; Y is the space of marginal distributions for the second endogenous variable; P_X, P_Y project joint distributions in S into corresponding marginals; A is the equation involving only one endogenous variable and hence determining its marginal distribution; and B is the other equation, which specifies the conditional distribution of the second endogenous variable given the first.

A third example is the triangular autoregressive representation of a bivariate covariance-stationary process $(x(t),y(t))$, in which x is causally prior to y in Granger's [40] sense. Here the space S is joint autocovariance functions of the processes $x(t)$ and $y(t)$, Y is the space of autocovariance functions for $y(t)$ alone, and X is the space of autocovariance functions for $x(t)$. The restrictions A are the first equation of the joint autoregres-

sive representation, which is a univariate autoregressive representation for X. The set B is determined by the second equation of the joint autoregressive representation.

Though here and in the remainder of the paper we will deal only with pairs (X, Y) of inputs and outputs, it should be clear that two-element orderings like those considered here can be extended to n-element orderings in a natural way.†

B. Characterizing a Relation as "Causal" or "Structural"

In this scheme it is natural to think of A as a particular input and of B as specifying the way inputs generate output. In the form of exogeneity test I used in my 1972 *American Economic Review* article [136], it is natural to think of "causal priority of x" as a characteristic of the distributed lag regression of y on x, that is, as a characteristic of B, not of B and A jointly. There is a nearly precise corresponding usage in engineering and physics, where operators mapping "input" functions into "output" functions are characterized as "causal" or "non-causal" (or as "realizable" or "non-realizable") again without reference to any particular input. These characterizations of B, the input-output connection, as causal or non-causal arise from intuitive notions of what causal systems "in nature" must be like. The reason it seems plausible to define a causal ordering as we have is that in a system (A,B) with a causal ordering from X to Y it is natural to contemplate varying the input A, holding B fixed, and obtaining outputs $P_Y(A \cap B)$ determined by A. Of course we can always undertake this experiment as a mathematical exercise, but the practical significance of the experiment depends critically on whether there is in nature a mechanism corresponding to the set of constraints B which would remain fixed while we varied A. Thus, even though there will be many systems (A,B) which generate the same results $A \cap B$ and imply different causal orderings, not all of them are of equal practical interest. The most interesting systems are those in which the input-output relation B is one which would in fact remain fixed if we varied A.‡ We will call such a B "structural," and a precise definition of this term follows.

If variation of A is to be even formally possible, B must have a form which "accepts" variation in A. To be precise, we will say

> *Definition:* The set $B \subset S$ *accepts* X *as input to* Y if for any $A \subset S$ which constrains only X (that is, any A such that $P_X^{-1} P_X(A) = A$), (A,B) makes X causally prior to Y.

Though similar notions are sometimes given the name "realizable,"

†For example, (A,B,C) orders X, Y, Z in that order if (A,B) makes X causally prior to Y and (A,B,C) makes $X \times Y$ causally prior to Z when mapping P_{XY}: $P_{XY}(s) = (P_X(s), P_Y(s))$ is used to map S into $X \times Y$.

‡The idea that causal orderings are at least implicitly linked to the possibility of varying the causally prior input has appeared before, in Simon [132] and Koopmans and Bausch [77], for example. Here, as throughout, I am providing references only to papers I have encountered in unsystematic reading in this area.

or "causal" in physical science literature, the formally explicit notion closest to that defined below seems to be Hurwicz's [63] definition of "structural," so I use that word.

> *Definition:* The set B is *structural* for inputs X if B accepts X as input; and when any set $C \subset X$ is "true" (or is "implemented") then $P_Y(P_X^{-1}(C) \cap B)$ is true.

In Hurwicz's [63] formulation, the "inputs" considered in defining structural relations are mappings from one space of equation systems into another. These mappings correspond to interventions of the form, for example, "fix the coefficient on x_i in the j'th equation at α." The role of B is played by some set of equation systems to which the transform is applied. The input is interpreted as implemented when, for example, a certain excise tax rate is fixed at α, in which case a "structural" equation system must be one in which the coefficient on x_i in the j'th equation system does in fact vary with the excise tax rate.

Since the property of being "structural" is a property of the way we interpret the system as applying to the real world, not of the system's form, there is no way of proving that a system is structural by examining the system's form. We can test whether a system is structural by using it to predict the effects of an intervention, making the intervention, and observing the result. In this way we may prove the system is *not* structural, but there can be no guarantee that other interventions, or even the same intervention repeated, will be predicted well by the system just because one or several test interventions are predicted well.

Nonetheless, it may be possible to specify enough properties which we know a structural system ought to have to allow us to distinguish potentially structural from surely non-structural systems. The use of such restrictions to distinguish non-structural systems is called "identification" in economics, "realizability theory" in some physical science applications.

In any application where inputs and outputs are dated, it is sensible to assume that a structural relation for inputs must not be one which determines past outputs from future inputs. This notion has not received much prominence in writings on econometric methodology because the hypothetical inputs with respect to which simultaneous equation models are identified are paradigmatically one-time transformations of a system which is assumed not to have been subject to intervention during the period of observation. We consider possible alterations of the supply or demand curve, for example, in a system which has had stable supply and demand curves in the sample period. In this context there is no time-stream of inputs and outputs. Furthermore, econometricians might resist the idea that dating of variables can be used to formulate general restrictions on structural relations. Indeed it is possible, for example, that a structural relation between two endogenous variables in an econometric model could involve a two-sided distributed lag. The claim that the conditional expectation of y_t given the past and future of x_t has non-zero

partial derivative with respect to future x's precludes that conditional expectation from being a structural relation only if variations in x are the inputs with respect to which the structure is claimed to be identified. More often than not, in economics there are no "variables" in the system whose time paths are the identifying interventions.

The condition that future inputs should not determine past outputs is called "causality" in some physical applications. For example, an operator mapping input functions of time into output functions of time is termed "causal" if it determines y_t (the value of the output function at t) from x_s (values of the input function) at $s \leq t$. In the formal framework of this paper, we can define this causality property as follows.

Consider a family of spaces $X_t^+, \bar{X}_t^-, \bar{Y}_t^-, Y_t^+$, t ranging over the real line, where X_t^+ is to be thought of as future inputs and Y_t^- as past outputs. The corresponding functions P_{Xt+} and P_{Yt-} map S into X_t^+ and Y_t^-, respectively.

> *Definition:* The subset B of S is *causal* if and only if:
> – For any t, B accepts X_t^- and X_t^+ as input; and
> – $t > r$, and $P_{xt+}^{-1}(P_{xt+}(A)) = A$, imply $P_{Yr-}(A \cap B) = Y_r^-$.

Paraphrasing, if we attempt to feed into B an input which specifies characteristics of future inputs only, the result will contain no information about past outputs.

Note that being causal in this sense is only a necessary condition for an input-output mechanism to be structural. The mistake of treating this causality condition as sufficient for a relation to be structural is a version of the old *post hoc ergo propter hoc* fallacy. But as we have already seen, causality of a relation is in this respect no different from any other identifying restriction. No characteristic of a relation's internal structure can guarantee that the relation is structural because being structural is a characteristic of the way we connect the relation to reality, not a property of the relation by itself.

C. Behavioral Relations

A reduced form of a standard econometric model in which government policy variables appear as exogenous variables is both causal and structural relative to interventions which take the form of variations in the time path of the policy variable. (Of course, this is the usual interpretation of such models. Whether the reduced form is actually structural relative to such variations in policy may be a matter of dispute.) Nonetheless, econometricians often talk and write as if reduced forms are not "structural." When structural systems are treated as different from reduced forms, the equations of the structural system are sometimes referred to as "behavioral."†

†"Behavioral" is also used, perhaps more frequently, to refer to any equation in a model which is not an accounting identity.

In this pattern of word usage, a "structure" is thought of as a system in which the hypothethical interventions defining what is structural are changes in the form of relations in the system. It is supposed that interventions which affect the form of only a subset of the full system of relations are possible, and "structural" equations are those which will remain fixed in form while interventions affecting other equations in the system are implemented.

Leading examples of such "structural" systems are supply-demand models and Keynesian multiplier models. Each equation is taken to apply to the behavior of a particular homogeneous group of people or institutions. The "demand" equation is what remains fixed when we somehow alter the behavior of "suppliers." The "consumption" function is what remains fixed when we change the behavior of "investors." Reduced forms are not "structural" because they reflect the combined effects of behavior of distinct types of individuals or institutions.

Lucas [89] and others have recently pointed out that a relation which is behavioral in this sense of applying to the behavior of a well-defined group need not be structural relative to alterations in the behavior of other groups (or sectors). In particular, a behavioral relation for a given group may reflect that group's methods of projecting the future behavior of other groups, and those projection methods are likely to change when the behavior of other groups changes.†

Thus, being behavioral is neither necessary nor sufficient to make a relation structural relative to an interesting class of possible interventions. Nonetheless, econometricians will go on trying to estimate behavioral systems because when a model has a behavioral interpretation it is usually much easier to guess how it will change when some definite disturbance to actual economic behavior of some sector occurs. Also, frequently models are estimated to investigate behavior without any direct policy application of the model in view.

Causality, in the sense we are giving that term, is an important identifying restriction on dynamic behavioral relations. Such relations ordinarily are meant to correspond to the decision rule of some class of economic agents. If we can distinguish variables taken as input by the agents from those determined as outputs by the agents, we expect that the decision rule will be causal from inputs to outputs. Thus in a model of a competitive market, where both suppliers and demanders take current and past prices as given in deciding on quantity, both supply and demand relations should be causal from price to quantity.‡

†In fact, this "rational expectations" critique of use of standard systems of behavioral equations to project policy effects is not limited to situations where expectations are involved. The basic idea is only that certain variables (for example, lagged price) are proxies for unobservable underlying quantities (for example, expected future prices) and that the nature of the proxy relation depends on the behavior of other groups. But expectations are far from the only type of unobservable concept for which proxies are commonly used in econometric models.

‡This does not mean, of course, that we can estimate either relation by least squares. The fact that a relation is causal does not mean that its input is statistically exogenous. No one

We have discussed four terms: "causal ordering," "causal," "structural," and "behavioral." The first two refer to properties of the logical structure of a model which may be plausible requirements if we are to contemplate treating the model as behavioral or as structural relative to variation in x as the identifying interventions. The controversy surrounding "causality" in economics tends to arise in situations where a model with a causal order from x to y or a model containing a relation which is causal for x as input fits some historical data; and then it is asserted that the model has a certain behavioral interpretation or can be used accurately to project the effects of varying the path of x. That is, the fit of the causal model to the data is used to buttress a claim that the model is behavioral or structural relative to variations in the path of x as identifying interventions.

In the remainder of the paper we discuss the justification for and dangers in interpreting fitted models displaying Wold or Granger causal orderings as behavioral or structural.

III. Interpreting Wold and Granger Causal Orderings

A. Orderings on Linear Dynamic Systems

Consider now a dynamic, stochastic, linear, econometric model

$$a_{11}*y_1 + a_{12}*y_2 = u_1 \tag{3}$$

$$a_{21}*y_1 + a_{22}*y_2 = u_2 .$$

The "*" indicates convolution, being read $a_{11}*y_1(t) = \sum_{s=-\infty}^{\infty} a_{11}(s)y_1(t-s)$. We will take the system to have been normalized with $a_{11}(0) = 1$ and $a_{22}(0) = 1$, and we will assume $a_{ij}(s) = 0$, all $s < 0$, $i=1,2$, $j=1,2$. This latter condition is natural because we would like the system to accept as input arbitrary initial conditions — values for $y_i(t)$, $t \leq 0$ and $u_i(t)$ $t \leq 0$. It is also natural in many applications to require that (3) be causal when u_i, $i=1,2$ are jointly covariance stationary, when (3) takes realizations of covariance-stationary processes u as input and produces realizations of covariance-stationary processes y as output, and when past and future input and output are defined by, for example, setting U_t^- as the values of $u_i(s)$, $i=1,2$, $s \leq t$. This amounts to the standard condition that the coefficients $a_{ij}(s)$ form a stable operator.†

presumes that because farmers take price as given, the supply equation for wheat can be estimated by a least-squares distributed lag regression of quantity of wheat on price of wheat. This point might seem self-evident, were it not that macroeconomists do sometimes seem to assume, for example, that if firms take wages as given in setting prices, a regression of prices on wages will capture firms' price-setting decision rule.

†It bears repeating that the reason for imposing stability of the operator applied to y in (3) is not simply that we know that the real-world y is not explosive. Systems of the form (3) with the $a_{ij}(s)$ operator "unstable" may fit covariance stationary pairs of u, y processes. The "instability" of the $a_{ij}(s)$ operator implies non-stationarity of y only if we impose the additional requirement that the system be causal with u as input and y as output. (For example, if the system is causal from stationary u's to stationary y's when we *reverse* the sign of the time index, then the left-hand-side coefficients will generally form an "unstable" operator, if instability is defined in the conventional way in terms of the absolute values of the roots of the characteristic polynomial.)

The system (3) displays a Wold causal ordering† if $a_{21}(0)=0$, u_1 and u_2 are serially uncorrelated and u_1 and u_2 are mutually uncorrelated. The system displays a Granger causal ordering‡ if $a_{21}(s) = 0$, all s, and $u_1(t)$ and $u_2(s)$ are mutually uncorrelated for all t, s. (Note that the possibility that $u_i(t)$, $u_i(s)$ are correlated is left open.)

Each ordering implies a convenient statistical property for the first equation of (3). The Wold ordering implies that $y_2(t)$ is uncorrelated with $u_1(s)$ for $s \geq t$, that is, that y_2 is *predetermined* in the first equation. The Granger ordering implies that $y_2(s)$ and $u_1(t)$ are uncorrelated for all t, s, that is, that y_2 is *exogenous*§ in the first equation. The conditions that y_2 be predetermined or exogenous in the first equation are not equivalent to Wold and Granger causal orderings because the second equation of (3) need not exist or take the form given in (3) in order for y_2 to be predetermined or exogenous in the first equation.‖ In fact, by appropriate definitions of input and output spaces, it is possible to make the conditions that y_2 be predetermined or exogenous in the first equation of (3) equivalent to the condition that the first equation be causal in the sense defined earlier in this paper. Nonetheless, in practice it can be helpful in organizing thought on the subject to think of exogeneity or predeterminedness as restrictions on a two-equation system. If there is a second equation of the form given in (3) and if u_1, u_2 are covariance-stationary, then Wold and Granger orderings are equivalent to y_2's being predetermined and exogenous, respectively, in the first equation.

The problem of whether and how to interpret Wold or Granger orderings as structural can arise in two guises. One of these is relatively familiar. Standard simultaneous equation models in most applications are implemented subject to numerous maintained hypotheses of exogeneity and predeterminedness. These maintained hypotheses are, at least in principle, generated by considering what we know about causal orderings in the real world phenomena being modeled and by translating those real world orderings into Wold or Granger orderings. This familiar problem of the criteria for making assumptions that variables are exogenous or predetermined was treated by Koopmans [75] some twenty years ago, and published work on the subject has advanced little since then. The discussion below, though aimed at interpreting exogeneity tests, is relevant also to purely *a priori* analysis of exogeneity assumptions.

The other guise of the problem arises when a model displaying a causal

†See Wold [162] for a forceful presentation of the argument that structural models are likely to take a form with a Wold ordering.

‡See Granger [40] for a presentation of this notion of causal ordering. Granger's original definition is not confined to linear covariance-stationary systems.

§ Some writers use the term "strictly exogenous" where we use "exogenous" to sharpen the distinction from "predetermined."

‖The potential advantages and disadvantages of testing exogeneity without estimating the second equation of (3) are discussed in Sims [141].

ordering has been shown to fit some sample of data and we must decide whether to interpret this historically observed Wold or Granger ordering as structural or behavioral. One reason this form of the issue has not received much attention until recently is that with the Wold ordering, which entered the econometric literature earlier, the issue never arises in pure form. If y_1 and y_2 are jointly covariance-stationary and have an autoregressive representation, then there is *always* a system of the form (1) displaying a Wold ordering.† Thus if we have no other identifying restrictions on (1), the demonstration that a system like (1) with a Wold ordering will fit the data is no evidence at all that the ordering is structural or behavioral. The structural or behavioral system could be practically any system of the form (1) and still imply that a Wold-ordered system would fit the historical data. This does not mean that a Wold ordering is untestable but only that a Wold ordering can be tested only in conjunction with other identifying restrictions on the system. Thus, debate over whether the estimated causal ordering is structural or behavioral is likely to be diverted into dispute over whether the other identifying restrictions are valid in the particular application under consideration.

The Granger ordering, on the other hand, is by itself a restriction on the class of jointly covariance-stationary processes y_1, y_2 which could satisfy (1). That this creates the possibility of testing the null hypothesis that y_2 is exogenous in the first equation – without any other identifying restrictions as maintained hypothesis – was pointed out as early as 1963 by Hannan [52]; but the first applications of the idea in economics, of which I am aware, were those by myself [136] and Sargent [129]. When a Granger ordering can be shown to fit some historical data and some plausible behavioral or structural interpretation implies a Granger ordering, the problem arises of determining what other possible behavioral or structural models, if any, might have produced the observed good fit of a model with a Granger ordering.

B. Generating Granger Orderings

The leading type of behavioral system which generates a causal ordering‡ occurs when there are two behavioral relations in the system, with one of the relations isolated from the other and with one of the two dependent variables in the system occurring in only one of the two relations. Thus, we may imagine weather and U.S. wheat production as determined by two behavioral relations, one describing farm behavior and one describing atmospheric behavior. Because the agents involved in the two relations are quite distinct and because the atmosphere presumably pays no attention to wheat prices or production in the United States, it is natural

†See Wold and Jureen [163].

‡One is tempted to label orderings arising in other ways "spurious," as I did in an earlier draft of this paper. But the other types of behavioral systems generating orderings might themselves be the center of interest, in which case this "leading" type of system generates a spurious ordering.

to suppose that weather will be exogenous in the wheat supply equation. And of course, we presume that the wheat supply relation is causal in our sense, so that only current and past weather helps determine current production; this is what guarantees equivalence of exogeneity with the Granger ordering.

While this sort of reasoning is entirely standard, it is subject to some pitfalls which are not always recognized. In particular it rests on assumptions about the residuals in the two relations which are sometimes ignored. While we may agree that in a properly specified "atmospheric behavior" model, U.S. wheat production would not play a significant part in determining rainfall, for rainfall to be exogenous requires also that residuals in the two relations be unrelated and that residuals in the supply relation be unrelated to any of the determinants of atmospheric behavior. If, for example, temperature affects wheat production but is omitted from our explicit list of right-hand-side variables in the supply relation, rainfall will not be exogenous. The supply equation residual will be related to temperature, which is in turn related to rainfall through the atmospheric behavior system.

So long as the need to be explicit about assumptions on residuals is kept in mind, the foregoing method is a legitimate way to justify an exogeneity assumption and is certainly what explicitly or implicitly underlies the exogeneity assumptions in most econometric models.

Econometricians are used to thinking in terms of modeling a p-variate system with p behavioral equations. Recently, however, considerable work has been done on models in which the error terms are given more complex structure – not every "random" component is taken to be generated as the residual in a behavioral equation. Ordinarily, if we consider models with both "behavioral" random terms and, say, "measurement error" random terms, a complicated simultaneous equation structure is implied. It is possible, however, for some components of the "behavioral" error vector and of the "measurement error" to be negligibly small, in which case a p-equation system with a causal ordering may fit the data with some or all of the equations being non-behavioral.

For example, suppose in a supply and demand system the supply equation has a negligibly small error term while the price variable, and only the price variable, is subject to substantial pure measurement error. That is, we have

$$
\begin{aligned}
(4)\qquad & p = a + bq + e_D && \text{(demand)} \\
& p = c + dq && \text{(supply)} \\
& p^* = p + e_M && \text{(measurement error)}
\end{aligned}
$$

Eliminating the unobservable p and writing the model in a form with the presumably uncorrelated random terms e_D and e_M on the right of the two equations, gives us

$$(5) \qquad p^* = c + dq + e_M$$
$$q = (a-c)/(d-b) + e_D/(d-b).$$

Clearly in this system q is exogenous in the first equation, which has the parameters of the supply equation. An exogeneity test would thus correctly suggest to us that we can recover a behavioral relation from a regression of p^* on q.

Whenever there is a very large measurement error in one variable, not in the other, the other variable will tend to appear exogenous in an equation with the error-ridden variable on the left. Because the "behavioral" error is relatively small, the residual in such an equation is dominated by the measurement error. As the above example suggests, this situation need not cause any problems when we have some way of determining from *a priori* considerations which (if any) of the behavioral relations in the system is recovered from the regression with error-contaminated variable on the left. Note that in the example above if there were an e_S term in the supply equation, with e_S and e_D of similar sizes, both small relative to e_M, then q would still show very little correlation with the residual in a regression with p^* on the left, but the coefficients in that regression would no longer have a behavioral interpretation.†

An example of a causal ordering arising from measurement error occurred in my work [138] on hours/output relations in manufacturing. Typically the estimated distributed lag relations, with monthly data, showed a tight relation of hours and output, with most of the response of one to the other completed within a month or so. The implied dynamics did not depend much on which variable appeared on the left-hand side of the regression, and exogeneity null hypotheses were accepted for either direction of regression. However, when deflated sales replaced industrial production as the measure of output, exogeneity tests were passed only with output on the left-hand side, and the implied dynamics were substantially altered if output was instead put on the right-hand side. This pattern of results fits the hypothesis of substantial pure measurement error in deflated sales, and since the data on sales are, at the monthly level, based on a relatively small sample, there is independent reason for believing this hypothesis.

In a perfect market for a durable commodity, price fluctuations must be unpredictable. This is a partial equilibrium result and deserves a more careful statement. But the economic logic of it is fairly simple: if it were known that the price was going to rise by more or less than the interest

†It is perhaps unnecessary to point out that the foregoing paragraphs should not be summarized as, "The presence of measurement error makes spurious causal orderings likely." Only certain special structures in the measurement error generate a spurious causal ordering. If one economist claims to have found a causal ordering allowing the estimation of a behavioral relation by least squares, another economist wishing to refute the first ought to show why measurement error of the required special structure is present, not just that measurement error of unspecified form is present.

rate, profit could be made either by withholding stocks from the market now (in the case of a known rise by more than the interest rate) or by selling from stocks now (in the case where the rise is known to be less than the interest rate). For such a commodity, any function of price of the form, for example, $p(t+s)-e^{rs}p(t)$, should be unpredictable – not only from past (before t) prices but also from any information publicly available at t. Granger's [40] original definition of a causal ordering was framed entirely in terms of series' predictive properties relative to one another; from his original definition, it is easy to see that a price from such a perfect market must always be "causally prior" to *any* time series of publicly available information. Furthermore when a causal ordering arises from this mechanism, there is no presumption that regressions with price on the right-hand side will be behavioral, even though they will pass exogeneity tests.

Exogeneity tests are thus a device for testing the perfect market hypothesis, so long as one avoids being misled by "spurious exogeneity" arising from the other mechanisms which can generate exogeneity. On the other hand, one clearly should be skeptical of behavioral interpretation of single-equation regressions, justified by an exogeneity test, with durable-goods prices on the right-hand side. This is unfortunate because an area of current research interest in which some economists† have applied exogeneity tests is behavioral modeling of price dynamics. It is likely that most prices are set in far from perfect markets, or apply to commodities which are not durable, or otherwise fail to meet the strict hypotheses of the perfect market price behavior proposition. When the perfect market hypothesis is appropriate, price should not only be exogenous but also should have an autoregressive structure of a particular form: the projection of $p(t+s)$ on values of p at t and earlier should be $e^{rs}p(t)$. If this autoregressive structure can be rejected, though p acts like an exogenous variable in a certain distributed lag regression, then a behavioral interpretation of the regression becomes more plausible.

Suppose we had a behavioral model in which one equation took the form

$$a_{22}*y_2(t) = b(y_1(t)-\hat{y}_1(t)) + u_2(t) \tag{6}$$

where $\hat{y}_1(t)$ is the minimum variance linear forecast of $y_1(t)$ based on values of $y_1(s)$, $y_2(s)$ for $s<t$ and where u_2 is serially uncorrelated and uncorrelated with past values of y_2 or y_1. Then if y_1, y_2 are jointly covariance-stationary and have an autoregressive representation, $\hat{y}_1(t)$ will be a linear combination of past values of y_2 and y_1, allowing us to write

$$y_1(t) - \hat{y}_1(t) = a_{11}*y_1(t) + a_{12}*y_2(t) = u_1(t). \tag{7}$$

Clearly $u_1(t)$ is serially uncorrelated and uncorrelated with past values

†See Geweke [33], for example. Geweke's exogeneity results are not easily explained by the mechanism of this paragraph, however.

of y_1, y_2. Now we can rewrite the right-hand side of (6) as $b\,u_1(t) + u_2(t)$. By construction, u_1 and u_2 are correlated, if at all, only contemporaneously. Thus, there is some constant α such that if we set $v_2 = bu_1 + u_2$, $v_1(t) = u_1(t) - \alpha v_2(t)$ is serially uncorrelated and uncorrelated with $v_2(s)$ for all s. Thus, taking linear combinations of (6) and (7), we arrive at

$$\begin{aligned} a_{11}{*}y_1(t) + (a_{12} - \alpha a_{22}){*}y_2(t) &= v_1(t) \\ a_{22}{*}y_2(t) &= v_2(t) \end{aligned} \tag{8}$$

which has the form of (3) and implies a Granger causal ordering.

Neither equation of (8) is likely to be structural, though the first equation would be consistently estimated by least squares regression and would pass a test for exogeneity of y_2.

The reason this example is interesting is that there are behavioral theories which lead to equations of the form (6). For example, suppose policy makers know that a structural relation between y_1 and y_2 exists of the form

$$b_{21}{*}y_1 + b_{22}{*}y_2 = u_2. \tag{9}$$

Suppose further that policy makers can set y_1 at any value they wish each period, subject to a "disturbance" u_1 and to the fact that when they choose $y_1(t)$, they know $y_1(s)$, $y_2(s)$ only for $s \leq t-1$. In particular then, they can use any rule of the form

$$b_{11}{*}y_1 + b_{12}{*}y_2 = u_1 \tag{10}$$

in forming y_1, subject to the condition that $b_{12}(0) = 0$, with $b_{11}(0)$ normalized at 1.0.

Let us take the policy makers to be minimizing a quadratic objective function of the form $Var[(g{*}y_2(t))^2]$. Now if u_1, u_2 form a linearly regular covariance-stationary process, they will have a joint moving average representation of the form

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = H{*} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{11}$$

where $v_1 = u_1 - \hat{u}_1$ and $v_2 = u_2 - \hat{u}_2$ are one-step-ahead forecast errors. The vector process v_1, v_2 will be serially uncorrelated.† Assuming that the optimal choice of (10) implies that the b_{ij} coefficients in (9) and (10) form a stable operator,‡ (9), (10), and (11) will jointly imply that we

†This is a version of the Wold decomposition of the process. See, for example, Rozanov [126].

‡This rules out some, but not all, interesting cases. See Sims [139].

can write

$$(12) \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = B^{-1} * H * \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

which in turn implies

$$(13) \qquad g * y_2 = f_1 * v_1 + f_2 * v_2 .$$

From the normalization rules $g(0) = b_{11}(0) = b_{22}(0) = 1$, from the fact that by construction $H(0) = I$ in (11), and from the fact that $b_{12}(0) = 0$, we can be sure that in (13) $f_2(0) = 1$, $f_1(0) = -b_{21}(0)$. Except for the requirement that the coefficient of $v_1(t)$ be one and of $v_2(t)$ be zero, equation (10) allows $y_1(t)$ to be chosen to be an arbitrary linear combination of current and past $v_1(s)$, $v_2(s)$.† Then if b_{21} has a one-sided inverse under convolution, that is, is "stable," we can also make $b_{21} * y_1$ equal to an arbitrary linear combination of current and past v_1, v_2 except for the restriction that the coefficient on current v_1 be $b_{21}(0)$ and that on current v_2 be zero. This in turn, through (9), implies that y_2 can be taken to be an arbitrary linear combination of current and past v_1, v_2 except that the coefficients on contemporary v_1, v_2 must be $-b_{21}(0)$, 1, respectively. Finally if g is invertible, the coefficients on the right-hand side of (13) can be chosen arbitrarily except for the previously listed restrictions on the 0-order coefficients.

Now the variance of $g * y_2$ is $\sum_{s=0}^{\infty} [f_1^2(s)\sigma_{11} + 2f_1(s)f_2(s)\sigma_{12} + f_2^2(s)\sigma_{22}]$, where σ_{ij} is the covariance of v_i with v_j. Since the summand is non-negative for all s, the minimum clearly occurs with $f_1(s) = f_2(s) = 0$, all $s \neq 0$.

We have arrived at the conclusion that the following equation will hold

$$(14) \qquad g * y_2 = v_2 - b_{21}(0) v_1 .$$

But this is precisely the form of equation given in (6), which we have already shown to imply that y_1 and y_2 will fit a model of the form (3) displaying a Granger causal ordering. Neither of the equations of the model displaying the ordering will be (9) or (10), so the ordering is certainly not behavioral. (One of the equations of the ordered model will be (14), however, so if we recognized the situation we could at least identify g.) Furthermore, it is the *policy* variable which will appear second in the causal ordering.‡

†To be strictly true, this result would require that v_1 and v_2 be expressible as linear combinations of current and past u_1, u_2, that is, that u_1, u_2 have an autoregressive representation. However, even when the autoregressive representation does not exist, it will be possible to make y_1 approximate an arbitrary linear combination of past v_1, v_2 arbitrarily well by appropriate choice of coefficients in (10).

‡The likelihood that optimal control might generate causal orderings was first pointed out to me by Milton Friedman in private correspondence.

Though this discussion has been framed in terms of abstract "policy makers," it might apply to "representative decision makers" as well. For example, employers setting a wage one period in advance, attempting thereby to achieve a target level of employment, might generate exactly such a structure, with causal ordering from employment to wages.

Since the foregoing discussion has introduced assumptions here and there along the way, it may be worthwhile to summarize formally what has been demonstrated.

> *Theorem 1:* Suppose
> - That (9) holds with $b_{22}(0)=1$, b_{22} and b_{21} both possessing one-sided inverses under convolution.
> - That u_1, u_2 form a covariance-stationary process with an autoregressive representation.
> - That the coefficients of (10) are chosen so as to minimize $Var[g*y_2]$, where g has a one-sided inverse under convolution.
>
> Then the resulting autocovariance structure for y_1, y_2 admits a Granger ordering from y_2 to y_1, with neither equation of the ordered system in general represented by (9) or (10).

The assumptions of the theorem are in fact quite restrictive and should not be read as implying that "optimal control generates causal orderings." Perhaps most restrictive is the requirement that the objective function involve y_2 alone — the objective cannot be to keep y_1 close to y_2, for example. Also restrictive is the requirement that the "information delay" be one period. If the delay is more than one period, no ordering is generated.† Note the special nature of the disturbance in the policy rule (10): u_1 must be influences on the policy variable which the policy makers cannot eliminate but which they do anticipate and attempt to counteract; u_1 cannot represent the effect of other policy-objectives or of imperfections in the process of policy-optimization. If u_1 is identically zero, the system becomes singular, the policy equation (10) can be estimated without error, and no ordering arises. Finally, it is by no means usual for us to have any good *a priori* reason to assume b_{21} to be invertible. Quite often, in fact, there will be a very small contemporaneous effect of y_1 on y_2 (small $b_{21}(0)$), which is likely to lead to b_{21}'s being non-invertible.

Equation (6) can be paraphrased to say that y_2 depends on forecast errors in y_1 plus an error term with certain properties. Behavioral relations of roughly this form have played central roles in some recent rational expectations macroeconomic models.‡ Lucas asserts a "supply equation" of this form without an error term. Sargent exploits the exogeneity implications of (6) by assuming the required properties for the error term

†Though I have not studied the case carefully, I believe that if the optimal control problem is solved by policy makers at a smaller time unit than applies to the fitted data, an approximate spurious ordering is likely to arise.

‡See, for example, Sargent [130] and Lucas [87].

arbitrarily. If $\hat{y}_1(t)$ is interpreted as the optimal forecast of $y_1(t)$, given the past of y_1 and y_2, and if it is assumed that the underlying behavioral relation has the form of (6) but with $\hat{y}_1$ replaced by $\hat{\hat{y}}_1$, where $\hat{\hat{y}}_1$ is a forecast based on the past of y_1 and y_2 and additional information, then (6) will hold and will have an error term with the required properties.† However, the assumptions required to claim a causal ordering by this route seem to me as restrictive as those required to claim an ordering resulting from optimal control. In particular, the assumptions of perfect information after a one-"period" delay and the requirement that there be no residual in (6) except that which is due to imperfect measurement of the expectations variable $\hat{y}_1$ appear to me to be highly restrictive. In Sargent's application to the labor market, they seem to me implausible.

As I showed previously [136], the Granger ordering is equivalent in a covariance-stationary system to the requirement that the two-sided distributed lag regression of y_1 on y_2 puts zero-coefficients on future values of y_2. The condition thus appears related to analysis of "leads and lags," and it naturally occurs to people that leads and lags between two series can be generated by their common dependence, with different lags, on some third series, even where there is no behavioral causal ordering between the two original series.

Consider the system

$$
\begin{aligned}
y_1 &= c_1 * z + v_1 \\
y_2 &= c_2 * z + v_2 .
\end{aligned}
\tag{15}
$$

Even if v_1 and v_2 are independent of each other and of z (so that (15) becomes what Sargent and I elsewhere in this volume have called an "unobservable index" model), there are no conditions on c_1 and c_2 alone which guarantee that y_1 and y_2 will satisfy a system like (1) with a Granger ordering. However, certain joint conditions on the c_i and the covariance properties of z and the v_1 will imply a Granger ordering.‡

The two-sided distributed lag regression of y_1 on y_2 has coefficients given by $R_{12} * R_{22}^{-1}$, where $R_{12}(s) = Cov(y_1(t), y_2(t-s))$, $R_{22}(s) = Cov(y_2(t), y_2(t-s))$, and R_{22}^{-1} is the bounded inverse of R_{22} under convolution.§ Using (15) and a convenient assumption that v_1, v_2, and z are mutually orthogonal, we can write $g = R_{12} * R_{22}^{-1} = c_1 * R_z * c_2' * (c_2 * R_z * c_2' + R_{v2})^{-1}$, where $R_z(s) = Cov(z(t), z(t-s))$ is the autocovariance function of z and R_{v2} is the autocovariance function of v_2. The "′" notation is

†See Shiller [131].

‡ Private conversation with Gary Skoog and unpublished work by John Geweke have been helpful to me on this topic.

§ R_{22}^{-1} is the inverse Fourier transform of the inverse of the spectral density of y_2, where this latter "inverse" is taken frequency-by-frequency and is the inverse under ordinary multiplication.

defined by $f'(s) = f(-s)$. This yields fairly directly a sufficient condition for a Granger ordering:

Theorem 2:
– If in (15) z, v_1, and v_2 are mutually orthogonal.
– If c_2 is invertible under convolution.
– If $c_2 * R_z * c_2' = \lambda R_{v2}$, where λ is a constant.
Then y_1 and y_2 can be written in the form (1) with a Granger ordering from y_2 to y_1.

Proof: The result follows when we rewrite the expression for g in the preceding paragraph as

$$g = c_1 * c_2^{-1} * c_2 * R_z * c_2' * (c_2 * R_z * c_2' + R_{v2})$$

and substitute

$$\lambda^{-1} c_2 * R_z * c_2' \text{ for } R_{v2}.$$

Theorem 2 is more interesting for the unlikeliness of its assumptions than for its positive result. The only likely example of a real-world case where its assumptions are plausible is where one is not dealing with time series at all, so that c_2, R_z, and R_{v2} all vanish for values of their arguments other than zero – that is, all variables are serially uncorrelated. In this case Granger orderings hold in both directions, y_1 to y_2 and y_2 to y_1, so it is easy to avoid the error of treating one of those orderings as structural.

Theorem 2 does not give necessary conditions for generating spurious orderings from (15), however. Another kind of sufficient condition for an ordering is:

Theorem 3: Suppose
– That y_2 has a univariate representation as a finite-order autoregression of order p.
– That z is a finite-order moving average process of order q.
– That c_2 is zero for $s > r$.
– That z, v_1, and v_2 in (15) are mutually orthogonal.
Then if $c_1(s)$ vanishes for $s \leq p + q + r$, y_1, y_2 can be represented as displaying a Granger ordering from y_2 to y_1.

Proof: Evident from inspection of

$$g = c_1 * R_z * c_2' * R_{22}^{-1}.$$

This theorem's assumptions are even more artificial than Theorem 2's.

The conclusion from this latter pair of theorems is that common dependence of y_1 and y_2 on a third variable with different lags does not,

in any natural way, tend to generate causal orderings. In fact, one might expect instead that testing for a Granger ordering would be a useful way to distinguish lead-lag relations which do not imply a behavioral causal ordering from those which do.

The cases we have discussed to this point are, as far as I know, the leading examples of structural models which might generate spurious Granger orderings. They show clearly that it is a mistake to act as if a behavioral causal ordering is automatically implied when a Granger ordering fits the historical data. On the other hand, it seems clear that there will be a large class of applications where, if a Granger ordering fits the data, the most plausible structure consistent with that empirical result will be one in which the ordering is behavioral.

IV. Conclusion

We can summarize by applying the analysis of this paper to the subject matter of my earlier paper [136] on money and income. Some monetarist economists claim that distributed lag regressions of GNP on a monetary aggregate capture a structural relation. It seems clear that the claim is that such a relation is behavioral, in the sense of describing the behavior of the rest of the economy in a two-equation system with the other equation describing the behavior of the Federal Reserve. (Of course, by a rational expectations argument, even if the GNP on M regression is behavioral, it might not be structural relative to systematic variations in policy.) For money to be exogenous in the rest-of-the-economy equation, it would have to be true that GNP does not feed back in to Federal Reserve behavior — influences on GNP other than M must not be related to Federal Reserve behavior. While not self-evident *a priori*, the idea that GNP does not feed back in to Federal Reserve behavior is not *a priori* unreasonable either, when one considers the shifting and conflicting policy objectives and economic theories under which the monetary authority acted in the post-war period. This possibility is consistent with the results of the exogeneity tests in that earlier paper.

Do the examples of section 3 suggest a likely mechanism for a spurious M-to-GNP causal ordering: that is, is there an alternative explanation for an M-to-GNP ordering which would imply the GNP-on-M regression does not have the postulated behavioral interpretation? The answer, in my opinion, is no. Money is not the price of a durable good. It is hard to see why there should be much greater measurement error in GNP than in M. For an optimal control model to produce an M-to-GNP ordering would require that GNP be subject to control while M was the target variable, which seems implausible. There appears to be no good reason to believe money supply or demand should be an exact function of current and past one-step-ahead prediction errors in some other variable. While the list of mechanisms to generate causal orderings given in section 3 is certainly not exhaustive, none of the possibilities listed there applies naturally to the money and GNP case.

Thus, it appears that economists who do not believe that GNP on money distributed lag regressions are structural ought to be basing their argument either on rational expectations or on the straightforward, old-fashioned possibility of Type II error. As was pointed out in the original article, standard errors of estimate on the coefficients of future M in the GNP on M regressions are large. The null hypothesis, while acceptable at standard significance levels, could nonetheless be false. In other words, while the results of the GNP *vs.* M exogeneity tests may be characterized as fuzzy or ambiguous, there does not appear to be good reason to believe them likely to be spurious.