Journal of Econometrics 5 (1977) 265–293. © North-Holland Publishing Company

CAUSALITY IN TEMPORAL SYSTEMS

Characterizations and a Survey*

David A. PIERCE

Federal Reserve System, Washington, DC 20551, USA

Larry D. HAUGH

University of Vermont, Burlington, VT 05401, USA

Received October 1975, final version received October 1976

A time series $\{X_t\}$ 'causes' another time series $\{Y_t\}$, in the sense defined by C.W.J. Granger, if present Y can be predicted better by using past values of X than by not doing so, other relevant information (including the past of Y) being used in either case. In this paper we (1) classify the possible causality relationships between two series X and Y, using an analogy to events in a sample space; (2) review existing work and present some new results on alternative characterizations of the more important causality events; and (3) compare several recent procedures for the empirical detection of causality.

1. Introduction

The elucidation of causal relationships among a set of variables is one of the major goals of empirical research. It has long been recognized that the finding of high correlation among variates does not in any necessary sense establish that they are causally related. Variables may be functionally related yet be uncorrelated; and, perhaps more often, they may be correlated yet not causally related. The former effect arises because correlation is a measure of linear association only; the latter because of common association of each with additional factors.

If the system or 'universe' were entirely linear, and if one identified all the influences or variables within that system, then it could perhaps be more fairly said that correlation between two variables (partial correlation) implies causation. But there still remain issues such as, given that two variables are causally related, which is causing which.

For temporal systems, Granger (1969) notes that the movement through time provides a natural answer to these questions. Granger's definition[1] is in

*Views expressed are not necessarily related to those of the Federal Reserve System. Much of the work of the second author was done at the University of Florida.

[1]This use of the term causality also corresponds to that of Wiener (1956), and the concept is thus also referred to as 'Wiener–Granger causality', e.g., Zellner (1975).

terms of predictability: a variable X causes another variable Y, with respect to a given universe or information set that includes X and Y, if present Y can be better predicted by using past values of X than by not doing so, all other information contained in the past of the universe being used in either case.

One may still hesitate to use the word 'causality' to describe the nature of the association that would then exist, but we shall adopt this usage here, noting as does Granger (1973) that it appears difficult to present an alternative definition for causality which can be tested empirically.

Allowing for instantaneous causality depending on whether current X is used in 'predicting' current Y, and defining causality from Y to X in similar fashion, there are many statements which can be made concerning the causal relationship between the two variables. In the next section we present more explicitly Granger's definition of causality, and classify 256 possible causality 'events' that can occur relating X and Y, which are the subsets of the 8-element 'sample space' generated by the truth values of the three binary outcomes 'x causes y', 'y causes x', and 'instantaneous causality exists'. In section 3 we set forth the bivariate time series model on which the rest of the paper is based. Sections 4 and 5 then present necessary and sufficient conditions for the occurrence of several of the causality events which have important implications for practical modelling problems. Questions of identifiability in multiple time series models enter into this treatment.

Beginning with section 6 we review and compare several recent methodologies proposed (and in some cases utilized) for empirically testing for the occurrence of certain causality events. The 'ad hoc filtering procedures' are based on a regression of one series on past, present and future values of the other, each series being first transformed by a common pre-chosen filter. Insofar as this filter, which is not determined empirically, inadequately removes serial correlation, this procedure is vulnerable to the 'spurious regression' phenomenon of Granger and Newbold (1974). On the other hand, the procedures of sections 7 and 8 pay careful attention to the univariate structures of the series as well as their interrelationships, which are assessed by interrelating the univariate residuals. Thus, the 'intrastructure', being more effectively dealt with, does not contaminate our inference concerning the 'interstructure'. But these methods may in certain causality tests underestimate the statistical significance of empirical causal relationships, due to a phenomenon previously observed by Durbin (1970) and Box and Pierce (1970). In section 9 we summarize some more general time series regression and modelling procedures which readily lend themselves to causality detection.

Several of the studies reviewed have focused on the relationship between money and income, and we summarize and compare the results obtained, which often are in conflict with each other. Section 10 presents some further discussion and conclusions.

2. Patterns of causality

Following Granger (1969), let $\{A_t, t = 0, \pm 1, \ldots\}$ be the given information set, including at least $\{(X_t, Y_t)\}$. Let $\bar{A}_t = \{A|:s < t\}$, $\bar{\bar{A}}_t = \{A|:s \leqq t\}$, and similarly define $\bar{X}_t$, $\bar{Y}_t$, $\bar{\bar{X}}_t$, $\bar{\bar{Y}}_t$. Let $P_t(Y|B)$ denote the minimum MSE single-step predictor of Y_t given an information set B, and $\sigma^2(Y|B)$ the resulting MSE.[2] Granger's definitions are:

(1) *X causes Y*

$$\sigma^2(Y|\bar{A}) < \sigma^2(Y|\bar{A}-\bar{X}), \tag{2.1}$$

(2) *X causes Y instantaneously*

$$\sigma^2(Y|\bar{A}, \bar{\bar{X}}) < \sigma^2(Y|\bar{A}). \tag{2.2}$$

Note that these definitions are in terms of single-period predictions. However, it was shown by Pierce (1975) that if (2.1) and (2.2) are true for the MSE of any multiperiod prediction, they also hold for the single-period prediction, for the systems we shall later be considering.

Causality from Y to X is defined in the same manner. *Feedback* occurs if X causes Y and Y causes X. There are in fact a great many different patterns which can occur. Accepting the result (which we shall prove later) that X causes Y instantaneously if and only if Y causes X instantaneously, a complete classification logically has 3 dimensions, (a) whether (or not) X causes Y, (b) whether Y causes X, and (c) whether instantaneous causality exists. The set of 'basic outcomes' or 'simple events' of interest would then consist of the $2^3 = 8$ possible combinations of these 'generators'. This set is akin to the sample space of an experiment (although we are not concerned with a probability structure), and indeed we may use a coin-tossing analogy by likening (a) above, 'X causes Y', to 'heads on first toss', (b) to 'heads on second toss', and (c) to 'heads on third toss'. Binary notation can be used to describe the elementary outcomes; for example, 'only instantaneous causality exists' would be represented as (001). The eight possible causality interrelationships are displayed in table 1, using both this binary notation and an alternative notation reminiscent of the arrows used to describe interrelationships in path analysis.

Pursuing the sample space analogy, every possible subset of the set of 8

[2]The original definition of Granger (1969) used predictors of y_t which are optimal in the sense of being minimum variance unbiased least-squares predictors. There is apparently no danger of practical misapplications in generalizing slightly to minimum squared error loss, i.e., min MSE predictors. Note also that y is restricted sufficiently in order that $\sigma^2(y|B)$ be independent of t.

outcomes is a *causality event*. For example, the simple statement 'X causes Y' refers to the event

$$\{(100), (110), (101), (111)\},$$

which may alternatively be depicted as

$$(x \to y) \vee (x \Rightarrow y) \vee (x \leftrightarrow y) \vee (x \Leftrightarrow y).$$

It is easily seen that there are $2^8 = 256$ possible causality events. Some of the more important ones in practice are displayed in table 2.

Table 1

Space of causality interrelationships.

	Description	Notations	
(1)	X and Y are independent	$(x\ y)$	(000)
(2)	Instantaneous causality only	$(x - y)$	(001)
(3)	X causes Y only and not instantaneously	$(x \to y)$	(100)
(4)	X causes Y only and instantaneously	$(x \Rightarrow y)$	(101)
(5)	Y causes X only and not instantaneously	$(x \leftarrow y)$	(010)
(6)	Y causes X only and instantaneously	$(x \Leftarrow y)$	(011)
(7)	Feedback, not instantaneously	$(x \leftrightarrow y)$	(110)
(8)	Feedback and instantaneous causality	$(x \Leftrightarrow y)$	(111)

Table 2

Selected causality events.

	Condition	Equivalent to
(I)	X causes Y	$\{(100), (101), (110), (111)\}$
(II)	X causes Y instantaneously	$\{(101), (111), (001), (011)\}$
(III)	Y causes X	$\{(010), (011), (110), (111)\}$
(IV)	Y causes X instantaneously	Same as II (Theorem 4.1)
(V)	Feedback	$\mathrm{I} \cap \mathrm{III} = \{(110), (111)\}$
(VI)	X causes Y but not instantaneously	$\mathrm{I} \cap \mathrm{II}^c = \{(100), (101)\}$
(VII)	Y does not cause X	III^c
(VIII)	Y does not cause X 'at all'	$\mathrm{III}^c \cap \mathrm{IV}^c$
(IX)	There is unidirectional causality from X to Y	(a) $\mathrm{I} \cap \mathrm{III}^c = \{(100), (101)\}$ (b) $\mathrm{I} \cap \mathrm{III}^c \cap \mathrm{II}^c = \{(100)\}$
(X)	X and Y are related only instantaneously (if at all)	$\mathrm{I}^c \cap \mathrm{III}^c = \{(001), (000)\}$
(XI)	X and Y are related instantaneously but in no other way	$\mathrm{I}^c \cap \mathrm{III}^c \cap (\mathrm{II} \cup \mathrm{IV}) = \{(001)\}$
(XII)	X and Y are independent	$\mathrm{I}^c \cap \mathrm{II}^c \cap \mathrm{III}^c \cap \mathrm{IV}^c = \{(000)\}$

We shall adhere to this framework in the sequel, first in characterizing causality patterns when the variates are generated by a certain type of bivariate time series model, and then in testing empirically for various causal relationships.

3. Time series models for causality assessment

In this paper we shall assume that the information set (the system or universe) consists of two variables X and Y, and that there exist transformations

$$x_t = T_x X_t, \qquad y_t = T_y Y_t,$$

such that (x_t, y_t) is a non-singular linear, covariance stationary, purely non-deterministic time series, and such that x and y are related causally in the same manner as are X and Y. Very often T_x and T_y will consist of first-difference or seasonal-difference operations, as frequently this type of transformation is necessary (and sufficient) to render the observed series stationary. Since such transformations are linear, and since the optimal predictors in terms of which causality was defined in section 2 are now also linear, any causality event is true of (X, Y) if and only if it is true of (x, y). Also, if any 'detrending' is done, the resulting causation analysis is relative to the information set (X, Y, Z) where Z consists of the trend variables. Moreover, certain nonlinear transformations of individual variates, such as logarithms or those of Box and Cox (1964), are also causality-preserving in this sense.

Under the above restrictions on X and Y it is known [see Hannan (1970) for example] that the bivariate process $(x_t, y_t)'$ has the representation

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \sum_{j=0}^{\infty} \Psi_j \begin{pmatrix} a_{t-j} \\ b_{t-j} \end{pmatrix} = \Psi(B) \begin{pmatrix} a_t \\ b_t \end{pmatrix}, \tag{3.1}$$

where $\{\Psi_j\}$ is a sequence of 2×2 matrices; $(a_t, b_t)'$ is a vector white noise sequence satisfying

$$E\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \mathbf{0},$$

$$E\left[\begin{pmatrix} a_t \\ b_t \end{pmatrix}(a_s b_s)\right] = \begin{cases} \Sigma \text{ (positive definite)}, & t = s, \\ 0, & t \neq s; \end{cases}$$

and

$$\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j \tag{3.2}$$

is a matrix polynomial in the lag operator B [defined by $B^j w_t = w_{t-j}$ for

$w_t = (w_{1t}, w_{2t})']$, convergent for $|B| \leqq 1$. Restrictions on Ψ_0 and/or Σ are required for identifiability of this model, a point returned to later. Eq. (3.1) is the moving-average representation of $(x, y)'$, and we shall also suppose that there exists an autoregressive representation

$$\Pi(B)\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{bmatrix} A(B) & H(B) \\ C(B) & D(B) \end{bmatrix}\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a_t \\ b_t \end{pmatrix}, \tag{3.3}$$

where $A(B) = \sum_{j=0}^{\infty} A_j B^j$ and similarly for the other operators. Thus for $|z| \leqq 1$,

$$\Pi(z) = \Psi^{-1}(z) \quad \text{and} \quad |\Pi(z)| \neq 0.$$

Now our assumptions imply that x_t and y_t each have representations as univariate linear processes, which we write in autoregressive form as

$$\begin{bmatrix} F(B) & 0 \\ 0 & G(B) \end{bmatrix}\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix}. \tag{3.4}$$

Thus, following Haugh (1972), we can derive from (3.4) a joint model for the univariate residuals which is of the form

$$\pi(B)\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{bmatrix} \alpha(B) & \beta(B) \\ \gamma(B) & \delta(B) \end{bmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_t \\ b_t \end{pmatrix}, \tag{3.5}$$

where α, β, γ, δ are of the same form as A, B, C, D. The various operators in these equations are seen to be connected by the relation

$$\begin{bmatrix} A(B) & H(B) \\ C(B) & D(B) \end{bmatrix} = \begin{bmatrix} \alpha(B) & \beta(B) \\ \gamma(B) & \delta(B) \end{bmatrix}\begin{bmatrix} F(B) & 0 \\ 0 & G(B) \end{bmatrix}. \tag{3.6}$$

We would expect that an analysis of (3.5) would yield quite directly information on the causality patterns concerning x and y (or X and Y), as u_t and v_t are the components of x_t and y_t that cannot be predicted from their own pasts. Indeed, it follows from the definitions (2.1) and (2.2) [where $\bar{Y}_t$ appears throughout] that (3.4) is a causality-preserving transformation: the linear nature of the transformation (3.4) insures that u and v are causally related in the same way as x and y. Furthermore, we shall see that the *cross-correlations*,

$$\rho_{uv}(k) = \frac{E(u_{t-k}v_t)}{[E(u_t^2)E(v_t^2)]^{\frac{1}{2}}}, \tag{3.7}$$

between the *whitened* or *filtered* series, $u_t = F(B)x_t$ and $v_t = G(B)y_t$, can be used to characterize any of the 256 causality events.

The cross-correlations $\{\rho_{uv}(k)\}$ are also closely related to the coefficients in the regression (projection) of v_t on all u's, past, present and future. If

$$E(v_t|u_s\colon -\infty < s < \infty) = \sum_{j=-\infty}^{\infty} \nu_j u_{t-j} = \nu(B)u_t,$$

then the projection of v_t on past, present and future $\{u_s\}$ is a *two-sided relation*,

$$v_t = \nu(B)u_t + f_t. \tag{3.8}$$

Note that u_t and v_t are each separately white noise; thus, f_t is necessarily serially correlated if at least two ν_j are non-zero. However, f_t is uncorrelated with u_t at all lags; and the cross-correlations in (3.7) are expressable as

$$\rho_{uv}(k) = (\sigma_u/\sigma_v)\nu_k. \tag{3.9}$$

The analogous relation with u as dependent variable is of the form

$$u_t = \omega(B)v_t + g_t, \tag{3.10}$$

where $\omega(B)$ is in general doubly infinite, g_t is uncorrelated with past, present and future v_t, and

$$\rho_{uv}(k) = (\sigma_v/\sigma_u)\omega_k. \tag{3.11}$$

Such two-sided relations between u and v are analogous to similar relations that also hold for x and y. From (3.8) and (3.10), using (3.4), we have

$$y_t = V(B)x_t + h_t, \tag{3.12}$$

and

$$x_t = W(B)y_t + k_t, \tag{3.13}$$

where

$$\begin{aligned} V(B) &= \frac{F(B)}{G(B)}\,\nu(B), \\ W(B) &= \frac{G(B)}{F(B)}\,\omega(B)\,. \end{aligned} \tag{3.14}$$

In (3.12) and (3.13) the operators $V(B)$ and $W(B)$ can be doubly infinite, and h_t and k_t have zero-mean and are uncorrelated with x_t and y_t, respectively. Rarely however will h_t and k_t be serially uncorrelated (this is relevant to section 6), as they are connected to the generally autocorrelated series f_t and g_t through the relations

$$G(B)h_t = f_t, \qquad F(B)k_t = g_t. \tag{3.15}$$

In the following section certain causality events will be characterized in terms of restrictions on two-sided relations such as (3.8) and (3.12). This framework has been widely employed by Sims, e.g. (1972, 1974), and others for causality detection (sections 6, 8.3, 9.2).

4. Characterizations of causality

In this section and the next we present several results pertaining to conditions on the bivariate time series model above which are necessary and sufficient for the occurrence of certain of the more important causality events of section 2.

4.1. Instantaneous causality

The assessment of instantaneous causality is closely connected with the concept of identifiability in multiple time series models. We shall work with the model for $u_t = F(B)x_t$, $v_t = G(B)y_t$ given in (3.5), noting that since F and G are one-sided, nonzero (at least $F_0 = G_0 = 1$) and possess inverses, statements concerning $A(B), \ldots, D(B)$ in the bivariate autoregressive model (3.3) can be inferred from those concerning $\alpha(B), \ldots, \delta(B)$ in (3.5), through (3.6).

Recalling that $\Sigma = E[(a_t b_t)'(a_t b_t)]$, the matrix autocovariance function (or the spectral density matrix) of (u_t, v_t), and of (x_t, y_t), is unchanged if the model

$$\pi(B)\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_t \\ b_t \end{pmatrix}, \qquad \mathrm{Cov}\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \Sigma, \tag{4.1}$$

is replaced by

$$\pi^*(B)\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} a_t^* \\ b_t^* \end{pmatrix}, \qquad \mathrm{Cov}\begin{pmatrix} a_t^* \\ b_t^* \end{pmatrix} = \Sigma^*, \tag{4.2}$$

where

$$\pi^*(B) = P\cdot\pi(B), \qquad \Sigma^* = P\Sigma P', \tag{4.3}$$

and P is any non-singular matrix. Moreover [Granger (1969)], if $\Sigma = I$ (the identity matrix), then also $\Sigma^* = I$ for all orthogonal P, so that there exist infinitely many specifications of the bivariate models (3.3) and (3.5), even with $\Sigma = I$. On the other hand, the specification for which $\pi_0 = I$ (Σ unrestricted) is unique: $P\pi_0 = I$ implies $P = I$.[3]

From an alternative vantage point, supposing $\pi_0 = I$, there exists a unique lower triangular matrix P_1, and therefore a unique upper triangular matrix P_2, such that each has unit diagonal, and such that $P_i \Sigma P_i'$ is diagonal for any positive definite matrix Σ [Haugh (1972)]. With $\pi_0 = I$, $P_1\pi_0$ is lower triangular ($\beta_0 = 0$) and $P_2\pi_0$ is upper triangular ($\gamma_0 = 0$). Thus, if any one of the conditions $\gamma_0 \neq 0$, $\beta_0 \neq 0$, or $\mathrm{Cov}(a_t, b_t) \neq 0$ exists, there are alternative but equivalent parameterizations of the model such that either of the other two conditions holds.

To connect this result with the condition (2.2) for instantaneous causality, let us temporarily adopt the specification $\beta_0 = \mathrm{Cov}(a_t, b_t) = 0$. In this parameterization it can be shown [referring to (2.1) and (2.2) with $A = \{u, v\}$] that

$$\sigma^2(v|\bar{A}) = \gamma_0\sigma_a^2 + \sigma_b^2 = \gamma_0\sigma_a^2 + \sigma^2(v|A, \bar{\bar{u}}).$$

Thus, u causes v instantaneously (X causes Y instantaneously) if and only if $\gamma_0 \neq 0$. But $\gamma_0 \neq 0$ if and only if $\beta_0 \neq 0$ in the equivalent parameterization in which $\gamma_0 = \mathrm{Cov}(a_t, b_t) = 0$ is the model-identifying restriction, which by an entirely symmetric argument is true if and only if Y causes X instantaneously.

It follows that the only issue concerning instantaneous causality is whether it exists. Whether there is instantaneous causality from X to Y, from Y to X, or both (instantaneous feedback) cannot be ascertained from the data. Noting also from (3.5) that $E(u_t v_t)$, and thus $\rho_{uv}(0)$ in (3.7), are non-zero unless $\beta_0 = \gamma_0 = \mathrm{Cov}(a_t, b_t) = 0$, we have:

Theorem 4.1. Instantaneous causality exists (II or IV in table 2) if and only if the following equivalent conditions hold:[4]

(1) $\rho_{uv}(0) \neq 0$.
(2) $v_0 \neq 0$ *in* (3.8).
(3) $\omega_0 \neq 0$ *in* (3.10).
(4) *It is not the case that* Σ *and* π_0 *are both diagonal in* (3.5).
(5) *At least one of* $\mathrm{Cov}(a_t, b_t)$, H_0, *and* C_0 *is non-zero in* (3.3).

Conditions (2) and (3) follow from eqs. (3.9) and (3.11).

[3] A third parameterization, found useful by Haugh (1972) in specifying multiple time series models, results from taking Π_0 triangular with unit diagonal and Σ diagonal.

[4] Additional equivalent conditions may hold in special cases. For example, if y does not cause x (Theorem 4.2), then $v_0 \neq 0$ in (3.8) if and only if $V_0 \neq 0$ in (3.12).

4.2. Unidirectional causality

The results in Theorem 4.1 also have a bearing on the assessment of unidirectional causality. Sims (1972) and Granger (1969, 1973) both use the 'diagonal-Σ' characterization of the bivariate process (x_t, y_t). Granger (1969) notes that $H(B) = 0$ in (3.3) if y does not cause x. And working with the moving average representation (3.1), Sims proves that y does not cause x if and only if $\psi_{12}(B)$ can be chosen to be zero [actually $\psi_{12}(B)$ *or* $\psi_{11}(B)$, but for identifiability we take $\psi_{11}^{(0)} = 1$].[5] Both results are correct, Granger's because it is only a one-way implication and Sims' because of the words 'can be'. Each result may be expanded into two more specific results, depending on whether 'at all' is appended to the phrase 'y does not cause x'. In doing so we denote the moving average representation of (u, v) by

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \theta(B)\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \begin{bmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{bmatrix}\begin{pmatrix} a_t \\ b_t \end{pmatrix}, \tag{4.4}$$

where $\theta_{11}^{(0)} = \theta_{22}^{(0)} = 1$, so that from (3.1) and (3.4)

$$\Psi(B) = \begin{bmatrix} F^{-1}(B) & 0 \\ 0 & G^{-1}(B) \end{bmatrix}\theta(B)\,. \tag{4.5}$$

Permitting instantaneous causality, we have:

Theorem 4.2. y does not cause x (VIII, table 2) if and only if the following equivalent conditions hold:

(1) $\psi_{12}(B)$ [*equivalently*, $\theta_{12}(B)$] *can be chosen zero.*
(2) $\theta_{12}(B)$ *is either* 0 *or a constant.*
(3) $\psi_{12}(B)$ *is either* 0 *or proportional to* $\psi_{11}(B)$.
(4) $V_j = 0, j < 0$, *in* (3.8).
(5) $\beta(B)$ *is either* 0 *or a constant.*
(6) $H(B)$ *is either* 0 *or proportional to* $A(B)$.
(7) $\rho_{uv}(k) = 0,\ k < 0.$

Conditions (1) and (4) were proved by Sims (1972). Conditions (2) and (5), also noted by Haugh (1972), explicitly allow for different parameterizations of instantaneous causality. Condition (3), shown previously by Pierce (1975) using another method, follows from (2) using eq. (4.5) and the fact that, since u_t, a_t and b_t are each themselves white noise, $\theta_{11}(B) = 1$ whenever $\theta_{12}(B)$ is constant or zero. Similarly (6) follows from (5), eq. (3.6), and the fact that $\alpha(B) = 1$

[5] Caines and Chan (1975) have proved a similar result for vector x and y

whenever $\beta(B)$ is constant or zero. Finally, to establish (7), choosing $\beta(B) = 0$ in condition (5) requires that u_t be proportional to a_t. Thus, from the second equation in (3.1),

$$y_t = \psi_{21}(B)a_t + \psi_{22}(B)b_t,$$

we have that, for $-k = h > 0$,

$$E(u_{t+h}v_t) = E\{a_{t+h}[\psi_{21}(B)a_t + \psi_{22}(B)b_t]\}, \tag{4.6}$$

which is zero since $a_{t+h}(h \geqq 1)$ is correlated with neither a_s nor b_s, $s \leqq t$. [Condition (7) and eqs. (3.9) and (3.14) provide an alternative proof of condition (4).]

As a corollary to this result, suppose XI in table 2 obtains, i.e., y does not cause x, nor vice versa, yet instantaneous causality is present. Then (i) either x or y may be related to present and past values of the other, with possibly long lag distributions, yet (ii) neither series is of value in predicting the other. This possibility has been noted elsewhere [e.g., Sims (1972)], and our experience with macroeconomic time series [e.g., Haugh and Box (1974), Pierce (1976)] is that the largest estimated values of $\rho_{uv}(k)$ has often been at $k = 0$.

The analogous result with no instantaneous causality present is:

Theorem 4.3. y does not cause x 'at all' (VIII in table 2) if and only if the covariance matrix Σ of (a_t, b_t) is diagonal and additionally the following equivalent conditions hold:

(1) $\psi_{12}(B) = 0$, $\psi_{21}^{(0)} = 0$.
(2) $\theta_{12}(B) = 0$, $\theta_{21}^{(0)} = 0$.
(3) $\rho_{uv}(k) = 0$, $k \leqq 0$.
(4) $V_j = 0, j \leqq 0$, *in* (3.12).
(5) $\beta(B) = 0$, $\gamma_0 = 0$.
(6) $H(B) = 0$, $C_0 = 0$.
(7) *The conditions of Theorem 4.2 are met and those of Theorem 4.1 are not.*

5. The 'basic' causality events

The three 'observations' on whether (a) x causes y, (b) y causes x, (c) instantaneous causality exists, were seen in section 2 to generate the causality 'sample space'. As with discrete sample spaces in general, these may be regarded either as dimensions (generators) of the causality space or as events ('cylinder sets') within that space. Based on the results in section 4 we can now present several equivalent characterizations of these basic events, and hence of any causality event. Additionally, we reformulate a spectral characterization of Granger (1969).

5.1. The cross-correlations of the univariate innovations

Theorems 4.1–4.3 above can be used to enable $\{\rho_{uv}(k)\}$ in (3.7) to characterize whatever causality patterns occur. From Theorem 4.1, the zero-lag cross-correlation signals the existence or non-existence of instantaneous causality.

Table 3

Conditions on cross-correlations of whitened series for causality events in table 2.

	Relationship	Restrictions on $\rho_{uv}(k)$
(I)	X causes Y	$\rho_{uv}(k) \neq 0$ for some $k > 0$
(III)	Y causes X	$\rho_{uv}(k) \neq 0$ for some $k < 0$
(II, IV)	Instantaneous causality	$\rho_{uv}(0) \neq 0$
(V)	Feedback	$\rho_{uv}(k) \neq 0$ for some $k > 0$ *and* for some $k < 0$
(VI)	X causes Y but not instantaneously	$\rho_{uv}(k) \neq 0$ for some $k > 0$ *and* $\rho_{uv}(0) = 0$
(VII)	Y does not cause X	$\rho_{uv}(k) =$ for all $k < 0$
(VIII)	Y does not cause X at all	$\rho_{uv}(k) = 0$ for all $k \leqq 0$
(IX)	Unidirectional causality from X to Y	$\rho_{uv}(k) \neq 0$ for some $k > 0$ *and* $\rho_{uv}(k) = 0$ for either (a) all $k < 0$ or (b) all $k \leqq 0$
(X)	X and Y are related only instantaneously (if at all)	$\rho_{uv}(k) = 0$ for all $k \neq 0$
(XI)	X and Y are related instantaneously and in no other way	$\rho_{uv}(k) = 0$ for all $k \neq 0$ *and* $\rho_{uv}(0) \neq 0$
(XII)	X and Y are independent	$\rho_{uv}(k) = 0$ for all k

And stating Theorem 4.2 in contrapositive form with y and x reversed, we have that x causes y if and only if $\rho_{uv}(k) \neq 0$ for *some* $k > 0$; and y causes x if there is a non-zero negative lag coefficient. Thus, the above binary observations may be equivalently expressed as

(a) $\rho_{uv}(k) \neq 0, \quad \exists k > 0,$

(b) $\rho_{uv}(k) \neq 0, \quad \exists k < 0,$

(c) $\rho_{uv}(0) \neq 0.$

Any causality event can be stated in terms of these cross-correlations. Table 3 does this for the events of table 2.

5.2. The model parameters

We will consider first the model (4.1) or (4.4) for the univariate innovations (u_t, v_t). Assuming Σ is diagonal and that the upper right-hand element (H_0, β_0, $\psi_{12}^{(0)}$, or $\theta_{12}^{(0)}$) is specified to be zero, our generators (a), (b), and (c) above may be expressed as

(a) $\gamma_j \neq 0, \quad \exists j > 0,$
(b) $\beta_j \neq 0, \quad \exists j > 0,$
(c) $\gamma_0 \neq 0,$

or equivalently as

(a) $\theta_{21}^{(j)} \neq 0, \quad \exists j > 0,$
(b) $\theta_{12}^{(j)} \neq 0, \quad \exists j > 0,$
(c) $\theta_{21}^{(0)} \neq 0.$

Modifications of these for the autoregressive and the moving average representations of (x_t, y_t) themselves are straightforward applications of Theorems 4.1 and 4.2.

A characterization which is relevant to several of the projection or regression methods for detecting causality which we shall describe is the following based on the single-equation representations at the end of section 3:

(a) $W_j \neq 0, \quad \omega_j \neq 0, \quad \exists j < 0,$
(b) $V_j \neq 0, \quad v_j \neq 0, \quad \exists j < 0,$
(c) $v_0 \neq 0, \quad \omega_0 \neq 0.$

5.3. The cross-spectrum

Granger (1969) showed that the cross-spectrum between two variables x and y can be decomposed into three parts, the first two each relating to a single causal arm of a feedback situation, and the third representing the influence of instantaneous causality. The model form used by Granger is a finite-parameter version of the autoregression (3.3); however, his development is equally relevant for general linear processes with autoregressive representations.

Using the adjoint formula for a matrix inverse, (3.3) can be rewritten as

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = [\Delta(B)]^{-1} \begin{bmatrix} D(B) & -H(B) \\ -C(B) & A(B) \end{bmatrix} \begin{pmatrix} a_t \\ b_t \end{pmatrix}, \tag{5.1}$$

where $\Delta(B)$ is the determinant of $\pi(B)$, which is of course an alternative expression for the moving average representation (3.1). Taking Σ to be diagonal, the parameters H_0 and C_0 are used to represent instantaneous causality.

We shall give an alternative derivation of the cross-spectral decomposition of Granger, based on cross-correlation techniques like those in Haugh (1972). First, for the model calculation without instantaneous causality ($H_0 = C_0 = 0$), it is easily shown that the cross-covariance between x_{t-k} and y_t is

$$\gamma_{xy}(k) = -\Delta(F)^{-1}\Delta(B)^{-1}\{D(F)C(B)\gamma_a(k)+H(F)A(B)\gamma_b(k)\}, \tag{5.2}$$

where $\gamma_a(0) = \sigma_a^2$, $\gamma_b(0) = \sigma_b^2$, $\gamma_a(k) = \gamma_b(k) = 0$ for $k \neq 0$, and B and $F = B^{-1}$ operate on the lag k. If y is not causing x, then $H(z) \equiv 0$, so that the second term in braces above vanishes. Similarly, if x is not causing y, then $C(z) \equiv 0$ and the first term vanishes. Thus, the two components of the cross-covariance generating function $\Gamma(z) = \sum_{-\infty}^{\infty} \gamma_{xy}(k)z^k$ identified by Granger are

$$c_1(z) = -\sum_{k=-\infty}^{\infty} \{\Delta(F)^{-1}\Delta(B)^{-1}D(F)C(B)\gamma_a(k)\}z^k, \tag{5.3}$$

$$c_2(z) = -\sum_{k=-\infty}^{\infty} \{\Delta(F)^{-1}\Delta(B)^{-1}H(F)A(B)\gamma_b(k)\}z^k. \tag{5.4}$$

Transformation of these two expressions to the frequency domain (setting $z = e^{i\omega}$) gives what could be called 'causal cross-spectra', one for the 'x to y' direction and the other for the 'y to x' direction. In fact, Granger further considers the phase and coherence spectra corresponding to each of these cross spectra.

If instantaneous causality exists, let $H'(B) = H(B)-H_0$ and similarly define $C'(B)$. It can then be shown that

$$\Gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)z^k = c_1(z)+c_2(z)+c_3(z), \tag{5.5}$$

where $c_1(z)$ and $c_2(z)$ are as in (5.3) and (5.4) except that H' and C' replace H and C. Moreover,

$$c_3(z) = \sum_{k=-\infty}^{\infty} -[\Delta(F)\Delta(B)]^{-1}\{C_0D(F)\gamma_a(k)+H_0A(B)\gamma_b(k)\}z^k \tag{5.6}$$

represents the presence of instantaneous causality. Thus the three basic events may be represented as

(a) $c_1(z) \not\equiv 0$,
(b) $c_2(z) \not\equiv 0$,
(c) $c_3(z) \not\equiv 0$.

An analogous decomposition of the cross-spectrum of the innovations (u_t, v_t) can also be made, based on the model (3.5).

6. Causality detection: Ad hoc filtering methods

We now consider the inferential aspects of causality. Given data $\{(x_t, y_t), 1 \leqq t \leqq n\}$, which as noted may be the result of a transformation on the actual non-stationary data $\{X_t, Y_t\}$, how may causality events be tested for empirically? We shall review several procedures for doing this which have recently been advocated, in some cases explicitly as causality detection procedures and in others as parts of time series modelling procedures which only implicitly embody the concept of causality we are employing. We begin with a regression procedure, due to Sims (1972), in which the data are transformed by a common filter, the purpose of which is to remove serial correlation in the residuals of a two-sided regression of y on x. The filter is specified a priori rather than from an empirical investigation of the data at hand (hence the word 'ad hoc'), although subsequent serial correlation tests are generally conducted on the regression residuals. The paper by Sims (1972) has attracted wide attention in econometric research, and several subsequent studies have also applied this technique [e.g., Barth and Bennett (1974)].

The basis for this procedure is condition (4) of Theorem 4.2, namely, that y does not cause x if and only if, in the unconstrained projection of y_t on x_s, $-\infty < s < \infty$, only present and past x enter the relation. The tests for causality may be briefly summarized as follows. The operator $V(B)$ in (3.12) is truncated at values $-N$ and M, sufficiently large to include any expected non-negligibly non-zero coefficients, and ordinary least-squares estimates of V_j are computed, after filtering the data. A prechosen filter $P(B)$ is applied both to x and to y, producing from (3.12) a relation of the form

$$y_t^* = V(B)x_t^* + h_t^*, \tag{6.1}$$

where asterisks denote transformed (filtered) variables. Sims utilizes in most cases the filter $P(B) = (1-0.75B)^2$, in the expectation that serial correlation in the disturbance h_t in (3.12) will thereby be reduced. However, if serial correlation remains in the filtered residuals h_t^* problems can occur since, while the least-squares estimates $\hat{V}_j$ will be consistent, bias occurs in the estimates of their variances. Very often this bias is downward, producing inflated t- and F-statistics and R^2 values. [See Granger and Newbold (1974) for an extensive study of this phenomenon.]

One thus suspects that in the application of this procedure, in those cases where the prechosen filter leaves substantial serial correlation in the filtered series, it is possible that causality may be believed to have been found where it does not exist. That this may have occurred in the case of money and income, the principal focus of the study by Sims (1972), is suggested by the findings of Feige and Pearce (1974). The filters they found appropriate to whiten the various series differed considerably from the filter $(1-0.75B)^2$, the difference being such that substantial positive autocorrelation was found by these authors to remain in the series x_t^* and y_t^*. [This does not always mean that serial correlation remains in h_t^*; we take up this point in the following paragraph.] Thus, while Sims found unidirectional causality from money [quarterly data on (i) currency plus demand deposits and (ii) the monetary base] to income [quarterly nominal GNP], case IX in table 2, Feige and Pearce were unable to establish causality in either direction (case XII) using an approach discussed in section 7.

We noted above that, while the filter for $(1-0.75B)^2$ may leave substantial autocorrelation in the series x_t^* and y_t^*, the issue for the current procedure is whether serial correlation is removed in the residuals h_t^*. That is, while Sims (1972, p. 545) makes the claim that this filter 'flattens the spectral density of most economic time series', the procedure would be appropriate (as Sims also states) if it only flattened the spectral density of h_t. To check on this, Sims applies a modification of the periodogram test of Durbin (1969), the results of which are (in the case of money and income) marginal at the 5% level of significance. Yet from the result of Feige and Pearce that money and income are at best weakly related it follows that y^* in (6.1) is primarily explained by h^*, so that the substantial (and very highly significant) autocorrelation remaining in y^* will *per force* show up in h^* [see also eq. (3.15) in this regard]. Possibly the periodogram test used by Sims is not always a reliable indicator of the type of serial correlation patterns likely to be of importance.

The need for a more adequate treatment of autocorrelation suggests two approaches. One is to augment an approach such as described in this section with a treatment of the regression residuals through a variant of generalized least squares. This is discussed in section 9.2. Another approach is to relate the prefiltering to the characteristics of the particular series being studied, an idea which is discussed in the following two sections.

7. Causality detection via cross-correlating univariate residual series

Haugh (1972) developed an approach to identify the degree and direction of association between two covariance stationary time series which immediately yields a causality detection procedure in the present framework. It is distinguished from the procedure of Sims by (i) the use of cross-correlation analysis rather than regression analysis on the filtered data; (ii) the use of separate filters on x_t and y_t to ensure that each is very nearly pre-whitened; and (iii) the empirical determination of these filters from the particular time series realizations under study. We shall consider each of these aspects in turn. Once the second and third of these points are incorporated into the procedure, the first will be seen to be largely a matter of individual preference.

7.1. Regression and correlation

Suppose, in the context of the procedure discussed in section 6, the prechosen filter $P(B)$, whether $(1-0.75B)^2$ or anything else, happened to correspond exactly to the x_t-process, so that

$$u_t = P(B)x_t \tag{7.1}$$

is white noise. If $z_t = P(B)y_t$, then, as noted by Box and Jenkins (1970), the lag-k cross-correlation

$$\rho_{uz}(k) = \frac{E(u_{t-k}z_t)}{\sigma_u\sigma_z}, \qquad -\infty < k < \infty, \tag{7.2}$$

is proportional to the distributed lag weight V_k in (3.12). In particular,

$$V_k = \frac{\sigma_z}{\sigma_u}\rho_{uz}(k), \tag{7.3}$$

and $V_k = 0$, $k < 0$, if and only if $\rho_{uz}(k) = 0$, $k < 0$, a condition which, as noted above, is equivalent to y not causing x.

Given data on u_t and z_t, $1 \leqq t \leqq n$, an estimate of $\rho_{uz}(k)$ is

$$r_{uz}(k) = \sum u_{t-k}z_t \Big/ \left[\sum u_t^2 \sum z_t^2\right]^{\frac{1}{2}},$$

so that, from (7.3),

$$\hat{V}_k = \frac{\hat{\sigma}_z}{\hat{\sigma}_u} r_{uz}(k) = \sum u_{t-k}z_t \Big/ \sum u_t^2 \tag{7.4}$$

is the resulting estimate of V_k.

Now if the $\{u_{t-k}\}$ were exactly orthogonal, then (7.4) would also be the formula for the ordinary least-squares estimate of V_k, $-N \leqq k \leqq M$, in the regression

$$z_t = \sum_{-N}^{M} V_j u_{t-j} + h_t', \tag{7.5}$$

whatever the values of M and N. Since $\{u_t, 1 \leqq t \leqq n\}$ is a finite sample series, the required exact orthogonality will not hold; however, to the extent that other cross-products do enter into the least-squares estimates, their effect should not be substantial. In particular, for any test of the hypothesis $V_j = 0$, $j < 0$, in

B

(7.5) based on regression analysis, one can construct an asymptotically equivalent test based on cross-correlation analysis.

If one wishes to examine causality in the other direction, the analogue of both the regression and correlation procedures above is to use the filter $G(B)$ in (3.4) such that $v_t = G(B)y_t$ is white noise, and, with $w_t = G(B)x_t$, to examine the relation obtained from (3.13) by applying the filter $G(B)$ throughout, which is

$$w_t = \sum_{-\infty}^{\infty} W_j v_{t-j} + k'_t = W(B)v_t + k'_t. \tag{7.6}$$

For example, x does not cause y (at all) if and only if $W_j = 0, j < 0$ ($j \leqq 0$), i.e., if and only if $\rho_{vw}(j) = 0, j < 0$ ($j \leqq 0$).

For further discussion, with examples, of the modelling procedure based on 'input whitening', see Box and Jenkins (1970). It is also of considerable interest that this analysis can be implemented in the frequency domain; Hannan's (1963) 'inefficient' method is based on the fact that, in (3.12),

$$V(e^{i\omega}) = \gamma_{uz}(e^{i\omega}) = \gamma_{xy}(e^{i\omega})/\gamma_{xx}(e^{i\omega}), \tag{7.7}$$

where in general $\gamma_{hk}(e^{i\omega})$ is the cross spectrum of two time series h_t and k_t. After γ_{xy} and γ_{xx} are estimated, the inverse Fourier transform of their ratio supplies an estimate of $V(B)$.

7.2. *Whitening both series*

Haugh (1972), in a sense, combines the two procedures based on (7.5) and (7.6) by examining the cross-correlations $\{\rho_{uv}(k)\}$ in (3.7) between the pair of whitened series $u_t = F(B)x_t$ and $v_t = G(B)y_t$. From (3.14) and the fact that F_j and G_j are necessarily 0 for $j < 0$, lag distributions in (3.12) or (3.13) are one-sided if and only if those in (3.8) or (3.10) are. And, from (3.9), (3.11) and the discussion in section 7.1, these questions are equivalently assessed either by regressing v_t on u_t (or vice versa) or by cross-correlating u_t and v_t. Furthermore, since now *both* series are white noise, the cross-correlation procedure is symmetric, as is evident from the characterization in section 5.1.

7.3. *Estimating the time series filters*

The preceding discussion presupposes that one knows the univariate model filters $F(B)$ and $G(B)$ which whiten x_t and y_t. Since this is seldom the case in practice, it is necessary to estimate them from the sample series, for example [as Haugh does] by employing the methodology of Box and Jenkins (1970).[6] One

[6]One feature of the approach of Box and Jenkins (1970) is that the estimation of these filters can be combined with the selection of an appropriate differencing transformation (first para graph of section 3) to render the original data (X_t, Y_t) stationary.

thereby obtains *estimates* $\hat{F}(B)$ and $\hat{G}(B)$ of the true whitening filters, and *residuals* $\hat{u}_t$ and $\hat{v}_t$, given by

$$\hat{u}_t = \hat{F}(B)x_t, \qquad \hat{v}_t = \hat{G}(B)y_t. \tag{7.8}$$

The causality analysis is carried out using the sample *residual cross-correlations*

$$\hat{r}_k = r_{\hat{u}\hat{v}}(k) = \sum \hat{u}_{t-k}\hat{v}_t \Big/ \left[\sum \hat{u}_t^2 \sum \hat{v}_t^2\right]^{\frac{1}{2}}, \tag{7.9}$$

in place of the sample innovation cross-correlations,

$$r_{uv}(k) = \sum u_{t-k}v_t \Big/ \left[\sum u_t^2 \sum v_t^2\right]^{\frac{1}{2}}, \tag{7.10}$$

calculable only if the true filters F and G were known.

To use the statistics (7.9) to test hypotheses about the population quantities $\{\rho_{uv}(k)\}$, i.e., about causality events, it is necessary to know at least approximately their sampling distribution under the null hypothesis that the type of causality being tested for is absent. Consider first the hypothesis of series independence (XII in table 2). Under this assumption, Haugh (1976) showed that the residual cross-correlations (7.9) and the white noise cross-correlations (7.10) have the same asymptotic distribution. The common distribution is particularly simple: for any m-dimensional vector $\hat{\boldsymbol{r}} = (\hat{r}_{k_1}, \hat{r}_{k_2}, \ldots, \hat{r}_{k_m})'$, where $\{k_i\}$ are integers,

$$\sqrt{n}\,\hat{\boldsymbol{r}} \sim N(\boldsymbol{0}, \boldsymbol{I}). \tag{7.11}$$

In particular, therefore,

$$n\hat{\boldsymbol{r}}'\hat{\boldsymbol{r}} = n \sum_{i=1}^{m} \hat{r}_{k_i}^2 \tag{7.12}$$

is chi-square with m degrees of freedom.[7]

Thus, to detect whether there is in fact some sort of causality present, N and M may be chosen as with Sims' method and the statistic

$$U = n \sum_{k=-N}^{M} \hat{r}_k^2 \tag{7.13}$$

[7]An asymptotically equivalent test which was established to be more accurate in a Monte Carlo study is also given in Haugh (1976).

referred to a χ^2-distribution with $(N+M+1)$ degrees of freedom. This is essentially the test employed by Feige and Pearce (1974) who, as described above, were usually unable to reject the hypothesis of independence between money and income.[8]

However, the distributional problem is as yet unsolved in the general case. For example, to assert unidirectional causality from x to y (IX in table 2), one needs at a minimum to reject the hypothesis that all $\rho_{uv}(k)$, $k > 0$, are zero and to 'accept' the hypothesis that all $\rho_{uv}(k)$, $k > 0$, are zero. Even apart from such problems as the joint significance level of such composite test procedures, the distribution of the $\{\hat{r}_k\}$ depends on the non-zero $\{\rho_{uv}(k)\}$, even when non-zero ρ's occur only at lags different from the $\hat{r}$'s used in a test statistic. This dependence is twofold: the $\{\hat{r}_k\}$ are no longer asymptotically distributed as the $\{r_{uv}(k)\}$ [Durbin (1970)], which in turn are no longer the cross-correlations of *independent* white noise series. The former effect can be shown to result in an underestimation of significance, i.e., a tendency not to detect causality when it exists. On the other hand the existence of cross-correlation in the population may inflate the standard errors of the sample cross-correlations [Bartlett (1935)].

It should be noted, however, that there do exist distributional results for auto- and cross-correlations of the residuals of fitted (one-sided) dynamic regression models of the form

$$y_t = \tau(B)x_t + \xi(B)e_t = \sum_{j=0}^{\infty} \tau_j x_{t-j} + \sum_{j=0}^{\infty} \xi_j e_{t-j}, \tag{7.14}$$

where $\{e_t\}$ is white noise independent of $\{x_t\}$. Tests of fit of the dynamic model form $\tau(B)$ can be based on the cross-correlations

$$\tilde{r}_k = r_{\hat{e}\hat{u}}(k) \tag{7.15}$$

between the residuals $\hat{e}_t$ and the filtered series $\hat{u}_t = \hat{F}(B)x_t$. The asymptotic distribution of these residual cross-correlations is known [Haugh (1972), Pierce (1972)], and is such that only a degree-of-freedom correction need be made in the χ^2-statistic [analogous to (7.13)] calculated from the $\{\tilde{r}_k\}$.

Consequently, an asymptotically valid test of the null hypothesis of unidirectional causality versus the alternative of feedback (V in table 2) [or more precisely, a test of the null hypothesis that y does not cause x whatever the instantaneous and $x \to y$ patterns] can be based on the χ^2-statistic formed from the cross-correlations (7.15) for negative lags.[9]

[8]Feige and Pearce present results for two different money series (and include in each case results with and without seasonal adjustment), and perform the X^2-test separately for lags $k \geqq 0$ and lags $k \leqq 0$. If the latter pairs are combined, one of the four resulting 2-sided tests is significant at the 10% level.

[9]A further problem with significance levels can result from the effects of *pretesting*, as in practice the dynamic regression models [as well as the univariate models (3.4)] are often

8. Other approaches relating univariate residual series

The method of Haugh just discussed differs most importantly from the ad hoc filtering method in its use of filters geared to the series under study. The procedures we discuss in this section combine a regression approach with the prior use of univariate time series analysis on the individual series.

8.1. *Granger (1973)*

Granger's (1973) framework is the bivariate autoregressive model (3.3), in which $\gamma_{ab}(0)$ is taken to be zero, so that, as we saw in section 3, H_0 or C_0 is allowed to be non-zero. It is assumed $A_0 = D_0 = 1$. As does Haugh, Granger begins by fitting univariate models to each series; thus the subsequent bivariate model for the filtered series (u_t, v_t) is (3.5), which can be written as two equations,

$$u_t = -\sum_{j>0} \alpha_j u_{t-j} - \sum_{j\geqq 0} \beta_j v_{t-j} + a_t, \tag{8.1}$$

and

$$v_t = -\sum_{j>0} \delta_j v_{t-j} - \sum_{j\geqq 0} \gamma_j u_{t-j} + b_t. \tag{8.2}$$

To test for causality from y to x [analogous procedures can be described for $(x \rightarrow y)$], Granger first expresses (8.1) in the form

$$\begin{aligned} u_t &= -\alpha^{-1}(B)\beta(B)v_t + \alpha^{-1}(B)a_t \\ &= \tilde{\omega}(B)v_t + \tilde{g}_t. \end{aligned} \tag{8.3}$$

We use tildas to distinguish (8.3), in which $\tilde{\omega}(B)$ is one-sided, from (3.10), wherein $\omega(B)$ may be two-sided. Then, using the residuals $\{\hat{u}_t\}$ and $\{\hat{v}_t\}$ calculated from the estimates $\hat{F}(B)$ and $\hat{G}(B)$ of the univariate models for x and y, Granger proposes a regression to estimate the coefficients $\{\tilde{\omega}_j\}$. As he notes, under the null hypothesis that x does not cause y, $\tilde{g}_t$ in (8.3) is white noise. Now while the regression data are fitted residuals, the estimates $\hat{\tilde{\omega}}_j$ will, as noted in section 7.1, be asymptotically equivalent to the sample residual cross-correlations $\hat{r}_k$ (and to $\hat{\nu}_j$). Thus the same features and limitations as concern the $\hat{r}_k$ in Haugh's procedure apply to the $\hat{\tilde{\omega}}_j$ estimated from (8.3) [and to the $\hat{\tilde{\nu}}_j$ analogously defined].

specified from examination of the same data (e.g., its autocorrelations) used later to fit the models. This is a problem with any iterative model-building procedure, including explicitly those of Box and Jenkins (1970) and those discussed in sections 7 and 8 – but including as well any econometric modelling wherein the researcher's specifications are influenced by previous persons' experience with the same or related macroeconomic data (and thus including, probably, virtually all econometric modelling).

Granger's procedure actually involves a variant of this, namely, rather than fixing M in advance, to employ a stepwise regression procedure and then apply tests such as those of Pope and Webster (1972).

8.2. Priestley (1971)

Although the methodology discussed by Priestley (1971) was not developed for the purpose of testing causality events, it bears so many resemblances to the work of Haugh (1972) and Granger (1973) that a discussion of its characteristics is relevant in a review of those methods that interrelate univariate residual series in detecting relationships between two time series. The underlying model considered by Priestley is the two-sided transfer function or lag distribution of (3.12).

As a diagnostic check of feedback in the presence of one-way causality, Priestley indicates that the residual (or unexplained) variance obtained from the fitted one-sided lag distribution model could be compared with the theoretically minimum variance obtainable by the most general model of the form (3.12). That is, there exist expressions for the minimum residual variance (σ_f^2) to be expected assuming that (3.12) is correct. Priestley gives such an expression in terms of the coherency between x and y,

$$\begin{matrix}\text{Lower bound on}\\ \text{residual variance}\end{matrix} = \int_{-\pi}^{\pi} h_{yy}(\omega)\{1-|C_{xy}(\omega)|^2\}\mathrm{d}\omega\,, \tag{8.4}$$

where $C_{xy}(\omega)$ is the coherency spectrum between x and y, and $h_{yy}(\omega)$ is the autospectrum of y. In practice the above bound would be calculated by replacing $h_{yy}(\omega)$ and $C_{xy}(\omega)$ by estimated spectral functions.

Once a proposed one-sided distributed lag model has been fit, its residual variance can be compared to the estimated lowest bound attainable with a two-sided distributed lag model. If the one-sided residual variance is too much greater than the bound above, it would appear that feedback or reverse causality may be present. We note, though, that an incorrectly specified one-sided model could also have a similar effect, so that other diagnostic checks of the fitted model are necessary [Box and Jenkins (1970)]. Also, a rule is needed for judging at what point the fitted residual variance is too large.

The particular method used to assess the nature of the relationship between x and y begins, as do Haugh's and Granger's, by finding univariate models (3.4) for each series separately. Again, the cross-correlations of the two univariate residual series are examined. However, what Priestley proposes is to fit by ordinary least squares the following regression model [compare (3.5) or (8.2)], after identification of the forms of $\delta(B)$ and $\gamma(B)$,

$$\delta(B)v_t = -\gamma(B)u_t+b_t, \tag{8.5}$$

using the *residual* series $\hat{u}_t$ and $\hat{v}_t$. A difficulty here is that, as $\hat{v}_t$'s are regressed on lagged values of themselves, estimates of δ_j involve the residual autocovariances from the model $\hat{G}(B)y_t = \hat{v}_t$ for y, their asymptotic distribution being different from the usual one where $G(B)$ is known [Box and Pierce (1970)].

The identification method outlined in Priestley (1971) would seem to have a further disadvantage in that the correlation structure of the noise is ignored at various steps, as in the parameterization of (8.5) above. Although it is claimed that the loss of efficiency in estimating the distributed lag parameters will usually be small, the fact that the noise may be autocorrelated will possibly lead to spuriously high t-values used in testing the significance of the regression coefficients, an effect analogous to that outlined in section 6.

Another feature of the approach of Priestley (1971) is the use of the method of covariance contraction to preliminarily identify the forms of $\delta(B)$ and $\gamma(B)$. By 'covariance contraction' is meant the initial identification of $\delta(B)$ in (8.5) by guessing (perhaps iteratively) by trial and error the form of $\delta(B)$ necessary to 'contract' the residual cross correlation function $\hat{r}_k$ in (7.9) to a position in which only a finite number of such cross correlation coefficients are significantly non-zero.[10] Rather than identifying $\delta(B)$ in its general form as a polynomial in B, Priestley suggests that repeated application of specific operators will serve to contract the cross-correlation function. For example, if this function 'decays slowly and smoothly', the use of the difference operator $(1-B)$ is recommended. However, if an operator $\delta(B)$ involving $(1-B)$ is required to contract the cross-correlation function, then in general the univariate innovation series v_t is not stationary. This contradicts the results of the first stage of identification wherein $G(B)$ is identified and fitted in order that v_t be approximately white noise.

It is recommended that the final identification of $\delta(B)$ and $\gamma(B)$ be made by fitting (8.5) as above. It seems though that one would prefer $\delta(B)$ and $\gamma(B)$, to be determined from the data in a more direct manner.

8.3. *Two-sided regressions with empirically determined filters*

Several recent studies, aiming primarily at assessing money–income relationships, have employed a two-sided regression analysis (following Sims' approach) in which the original series are prefiltered by empirically determined filters. These include Auerbach and Rutner (1975), Bisignano (1974), Ciccolo (1975), Dy Reyes (1974), and undoubtedly others. Aside from variations in the generality of the classes of univariate filters considered, the two-sided regression models

[10]Note that if $\delta(B)v_t = -\gamma(B)u_t+b_t$, then $\delta(B)\rho_{uv}(k)$ is proportional to $\gamma(B)\rho_{uu}(k)$, where $\rho_{uu}(k) = 1$ if $k = 0$ and 0 otherwise, and where $\delta(B)$ and $\gamma(B)$ now operate on the lag k. Hence, the pattern in the cross-correlation function $\rho_{uv}(k)$ is determined by $\delta(B)$ and $\gamma(B)$, with $\delta(B)$ influencing the large-lag pattern. By a 'contraction' of $\rho_{uv}(k)$ is meant the filtering of $\rho_{uv}(k)$ by $\delta(B)$ to produce $\delta(B)\rho_{uv}(k)$.

are essentially of the form (3.10) or (7.6) depending on whether, respectively, each series was whitened separately or the regressor-whitening filter was also applied to the regressand.

One of the non-trivial variations in the univariate model specifications that have been employed is whether the data were differenced as part of the univariate modelling procedure. The dissimilarities in the causality test results, due to whether differencing was or was not done, provide an interesting corroboration of the Sims/Feige–Pearce comparison for money and income. Those test procedures that allow differencing tended to find weaker interrelationships than those that require the assumption of stationarity of the series levels (section 10 picks up this point again).

In other respects these studies serve to highlight the importance of auto-correlation, as variations in the prefilter produced, for the case of money and income, 'findings' of independence, unidirectional causality in each direction, and feedback. Again the remarks concerning the distribution of $\{\hat{r}_k\}$ in section 7 or of $\{\hat{\omega}_j\}$ in section 8.1 are relevant here if the null hypothesis is something other than series independence. An exception is the work by Ciccolo, in which the filter for x is obtained from a model for x which also includes past y. But such a model, if one is confident that it is adequately specified, can be used directly to assess whether y causes x (see section 9.1); no further analysis is required.

9. Other methods

We conclude our survey by summarizing some existing methods which either have not been widely used in econometric applications or do not fall uniquely into one of the preceding categories.

9.1. A 'direct' method

A procedure not actually suggested, but implicitly contained in Granger's (1969) paper, is simply to regress x on its own past and that of y, giving directly an estimate of the first (in this case) equation in (3.3) [taking $H_0 = C_0 = 0$ and Σ possibly non-diagonal]. The possibility that y causes x can be examined by testing whether the coefficients of lagged y are zero [(6), Theorem 4.2], or equivalently, by comparing $\hat{\sigma}_a^2$ and $\hat{\sigma}_u^2$. This procedure has been implemented by Sargent (1976) in testing for causal relationships implied by Keynesian and classical economic models.

The main caution with this approach is to ensure that the specification is adequate so that the residual is indeed the white noise a_t in (3.3). In particular, omitting relevant lagged x values could again produce the spurious regression phenomenon. Perhaps a reasonable specification is for $A(B)$ to contain coefficients at all lags which are non-negligibly non-zero in the estimated autoregression $\hat{F}(B)$, as $A \equiv F$ under the null hypothesis that y does not cause x, though a proper specification would in general involve solving the bivariate modelling problem (section 9.3).

9.2. Generalized least-squares procedures

Again within the regression framework, asymptotically optimal causality-assessment procedures can be based on the two-sided distributed lag equations, (3.12) and (3.13), using generalized least squares (GLS).[11] There are several ways to implement this, which generally involve a first-stage consistent estimation of $V(B)$ and $W(B)$ in (3.12) and (3.13) [as before, truncated at suitably distant points], followed by an appropriate residual analysis. In the time domain this may involve an autoregressive and/or moving average modelling of the first-stage residuals followed by either (i) a transformation (filtering) of x and y by this estimated model or (ii) an estimation, from this model, of the covariance matrix of the residuals and use of the appropriate GLS formulas. Alternatively [Hannan (1963) 'efficient' method], use may be made of the fact that the covariance matrix of the Fourier transform of the residuals is approximately diagonal and thus easily inverted.

A disadvantage of these methods is that the regression procedure may become unwieldly and perhaps numerically unstable with a large number of possibly highly autocorrelated regressors. Of course, various procedures might be combined; for example, a filtering aimed at *reducing* [with no claim to eliminating, as is required for the method of section 6] autocorrelation or collinearity, which could include simply differencing the series, may be followed by one of the procedures in this section; e.g., see Sims (1974) and Williams et al. (1976).

9.3. Explicitly multivariate approaches

In a sense most of the foregoing are procedures for using single-equation methods to gain insight concerning a multivariate problem. If one can estimate A, B, C, D in (3.3) and their sampling distribution, one can assess causality relationships using such results as (5) of Theorem 4.1 and (6) of Theorem 4.2. Extensions to wider 'universes' or information sets are also more straightforward within this framework. Parzen (1969) outlines the basic multivariate modelling problem and references other related work.[12] If a parsimonious multivariate autoregressive moving average model can be identified [which is in general difficult when feedback exists; see Wilson (1971), Haugh (1972)],

[11]Granger (1973) for example mentions the possibility of using GLS with respect now to a regression involving the two univariate residual series derived from (3.5).

[12]This reference is also the earliest we have come across which explicitly advocates fitting univariate models as a part of the multiple time series modelling procedure.

Hannan (1970), Wilson (1973) and Newton (1975) have given estimation procedures and their asymptotic properties. Alternatively, if an appropriate order for a long autoregressive approximation can be determined [e.g., Parzen (1975)], procedures based on multivariate regression can be employed. Caines and Chan (1975a, b) have developed and employed likelihood ratio procedures to test for 'feedback' from y to x, which can themselves be vectors.[13]

10. Final remarks

We have reviewed and analyzed several procedures for the empirical detection of causal relationships between variables. Many of the differences we have noted in these procedures are largely a matter of a researcher's individual preference: indeed this is not surprising in view of the many equivalent characterizations of the same causality events found in sections 4 and 5. The real difference would seem to be in how autocorrelation problems (the univariate series structures) are handled. If the procedures are not geared to the particular series under study, 'spurious regressions' can result; if they are, 'spurious independence' may on occasion occur. We conclude with some further brief observations on the procedures we have discussed.

First of all, it seems possible to argue that the downward bias in significance resulting from using estimated residuals from filtered univariate models is probably of a small order of magnitude in many if not most applications. Recall that there is no (asymptotic) bias at all in Haugh's and Granger's procedures if all the population coefficients $\rho_{uv}(k)$ are zero. Thus a small degree of cross-correlation should only result in a small depression of the lowest significance level at which the null hypothesis (that a particular causality event does not occur) could be rejected; and often only a small degree of cross-correlation remains in many economic systems after the autocorrelation patterns are properly taken into account [see Granger and Newbold (1974), Haugh and Box (1977), Pierce (1977); the last of these describes and documents this 'independence phenomenon'].

We would expect also that the 'spurious regression' problem might be reduced by a better choice of the prior filter. There is ample evidence for choosing '$(1-B)$' as a factor of this filter for much economic data.[14]

[13]Quotes are used since their definition of feedback is different than in table 2, implying an ordering and moreover not requiring that x cause y. Similarly, the term 'causal ordering' also appears to have two definitions. Generally, the three statements 'there exists a causal ordering from x to y', 'the ordered process (x, y) is feedback free' (à la Caines and Chan), and 'y does not cause x' are equivalent, though the additional stipulation that x cause y is sometimes associated with the first of these.

[14]A test of the null hypothesis that an autoregression contains a unit root (is non-stationary) has been developed by Dickey (1975). Williams et al. (1976) studied both possibilities and concluded that differencing was more appropriate for U.K. money, income, and price series.

Again it should be emphasized that although 'causality' has been the topic of investigation, empirical association (correlation and regression) has been the mode of investigation, and hence the usual provisos discussed briefly in section 1 apply. The fact that (possible) lead-lag relationships in time can aid in the determination of causality in temporal systems is of course an important advantage distinguishing these efforts from earlier non-dynamic correlation–causation studies. But several important qualifications remain. First, the universe of information available must be recalled as an important component in Granger's definition of 'temporal causation'. In particular for many of the practical applications made to date, only bivariate or trivariate systems have been investigated; and the bivariate methods discussed in this paper have essentially investigated causality relative to the information set consisting only of x and y.

It should also be noted that the causality definition assumes the nature of the series given. We have not therefore addressed questions of measurement error or added noise (including seasonality when it is so viewed). Thus if $x = x_1+x_2$ and $y = y_1+y_2$, all components being stationary, it may well be that different causality events will hold for x_1 and y_1 than for x and y. Quite often x_1 and y_1 will represent 'seasonally adjusted' x and y. Notions of 'statistical exogeneity' in structural models [e.g., Sims (1974)] are sometimes related to this distinction: x is (statistically) exogenous relative to y (either may be vectors) if, for a suitable decomposition, y_1 does not cause x_1 in Granger's sense.[15]

Another interpretational problem occurs whenever 'instantaneous causality' exists. As Granger (1969) notes if only 'simple' unidirectional causality is present (e.g., IX(b) in table 2), then one has much more of a basis for expecting philosophical or structural notions of the term to correspond to his definition than otherwise, as we have seen that 'true source' of instantaneous causality is empirically undecidable. A similar uncertainty holds concerning Wold (1954) causal orderings [see Basmann (1965), Sims (1975) and Zellner (1975)].

Finally, that spurious Granger causal orderings can occur even with optimal estimation and testing procedures has been noted by Box and MacGregor (1954), Caines and Chan (1975a, b), Sims (1975), Zellner (1975) and others. Under some fairly general conditions in which an 'independent' variable can be controlled or influenced by someone who does so on the basis of movements in a dependent variable y, this feedback control rule or 'reaction function' can dominate the empirical relationship. This would result in reporting that y causes x (which could still be the basis for a valid forecasting relationship) in cases where the structural relationship is one of causation only from x to y.

[15]A separate issue concerns the *estimation* of x_1 and y_1. For example, passing either series through a two-sided filter (unless both are filtered identically), as in seasonal adjustment or more generally in signal extraction for a large class of stochastic unobserved components models, can fundamentally alter causality relationships.

References

Auerbach, R.D. and J.L. Rutner, 1974, Money, income and causality in the time and frequency domains, manuscript (Federal Reserve Bank, Kansas City, KS).

Barth, J.R. and J.T. Bennett, 1974, The role of money in the Canadian economy: An empirical test, Canadian Journal of Economics 7, 306–311.

Bartlett, M.S., 1935, Stochastic processes (Cambridge University Press, London).

Basmann, R.L., 1965, A note on the statistical testability of 'explicit causal chains' against the class of 'interdependent' models, Journal of the American Statistical Association 60, 1080–1093.

Bisignano, J., 1974, Money, income and causality: Another look, manuscript (Federal Reserve Bank, San Francisco, CA).

Box, G.E.P. and D.R. Cox, 1964, An analysis of transformations, Journal of the Royal Statistical Society (B) 26, 211–252.

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis forecasting and control (Holden-Day, San Francisco, CA).

Box, G.E.P. and J.F. MacGregor, 1974, The analysis of closed-loop dynamic-stochastic systems, Technometrics 16, 381–398.

Box, G.E.P. and D.A. Pierce, 1970, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, Journal of the American Statistical Association 65, 1509–1526.

Caines, P.E. and C.W. Chan, 1975a, Estimation, identification and feedback, in: E. Lainiotis and R.K. Mehra, eds., System identification: Advances and case studies (Marcel Dekker, New York).

Caines, P.E. and C.W. Chan, 1975b, Feedback between stationary stochastic processes, IEEE Transactions on Automatic Control 20, 498–508.

Ciccolo, J.H., 1975, Four essays on monetary policy, Ph.D. dissertation (Yale University, New Haven, CT).

Dickey, D.A., 1975, Hypothesis testing for nonstationary time series, unpublished manuscript (Iowa State University, Ames, IA).

DyReyes, F., 1974, A test of the direction of causality between money and income in Canada, Japan and the United States, Ph.D. dissertation (Department of Economics, Iowa State University, Ames, IA).

Durbin, J., 1970, Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, Econometrica 38, 410–421.

Feige, E. and D.K. Pearce, 1974, The causality relationship between money and income: A time series approach, presented at the Midwest Economic Conference (Chicago, IL).

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross spectral methods, Econometrica 37, 424–438.

Granger, C.W.J., 1973, Causality, model building, and control: Some comments, presented at the IFAC/IFORS International Conference on Dynamic Modelling and Control, July 9–12 (University of Warwick, Coventry).

Granger, C.W.J. and P. Newbold, 1974, Spurious regressions in econometrics, Journal of Econometrics 2, 111–120.

Hannan, E.J., 1963, Regression for time series, in: M. Rosenblatt, ed., Time series analysis (Wiley, New York).

Hannan, E.J., 1970, Multiple time series (Wiley, New York).

Haugh, L.D., 1972, The identification of time series interrelationships with special reference to dynamic regression, Ph.D. dissertation (Department of Statistics, University of Wisconsin, Madison, WI).

Haugh, L.D., 1976, Checking the independence of two covariance-stationary time series: A univariate residual cross correlation approach, Journal of the American Statistical Association 71, 378–385.

Haugh, L.D. and G.E.P. Box, 1977, Identification of dynamic regression (distributed lag) models connecting two time series, Journal of the American Statistical Association 72, forthcoming.

Newton, H.J., 1975, Fitting mixed models to multiple time series, presented at the IMS/ASA Eastern Regional Meeting, May 21–23 (Rochester, NY).

Parzen, E., 1969, Multiple time series modeling, in: P.R. Krishnaiah, ed., Multivariate analysis – II (Academic Press, NY) 389–409.

Parzen, E., 1975, Multiple time series: Determining the order of approximating autoregressive schemes, presented at the IMS/ASA Eastern Regional Meeting, May 21–23 (Rochester, NY).

Pierce, D.A., 1972, Residual correlations and diagnostic checking in dynamic-disturbance time series models, Journal of the American Statistical Association 67, 636–640.

Pierce, D.A., 1975, Forecasting in dynamic models with stochastic regressors, Journal of Econometrics 3, 349–374.

Pierce, D.A., 1977, Relationships – and the lack thereof – between economic time series, with special reference to money, reserves, and interest rates (with discussion), Journal of the American Statistical Association 72, forthcoming.

Pope, P.T. and J.T. Webster, 1972, The use of an F-statistic in stepwise regression, Technometrics 14, 327–340.

Priestley, M.B., 1971, Fitting relationships between time series, Bulletin of the International Statistical Institute 34, 1–27.

Priestley, M.B., W.W. Brelsford and R.L. Chadda, 1971, Fitting relationships between two time series, unpublished manuscript (Bell Telephone Laboratories).

Sargent, T.J., 1976, A classical macroeconometric model for the United States, Journal of Political Economy 84, 207–237.

Sims, C.A., 1972, Money, income and causality, American Economic Review 62, 540–552.

Sims, C.A., 1974, Output and labor input in manufacturing, Brookings Papers on Economic Activity, 695–728.

Sims, C.A., 1975, Exogeneity and causal ordering in macroeconomic models, presented at the Seminar on New Methods in Business Cycle Research, Nov. 13–14 (Minneapolis, MN).

Wiener, N., 1956, The theory of prediction, in: E.F. Beckenbach, ed., Modern mathematics for engineers, series 1 (McGraw-Hill, New York) ch. 8.

Whittle, P., 1963, Prediction and regulation by linear least squares methods (English Universities Press, London).

Williams, D., C.A.E. Goodhart and D.H. Gowland, 1976, Money, income and causality: The U.K. experience, American Economic Review 66, 417–423.

Wilson, G., 1971, Recent developments in statistical process control, presented at the 38th Session of the International Statistical Institute, Aug. 10–20 (Washington, DC).

Wilson, G.T., 1973, Estimation of parameters in multivariate time series models, Journal of the Royal Statistical Society Series B 35, 76–85.

Wold, H.O.A., 1954, Causality and econometrics, Econometrica 22, 162–177.

Zellner, A., 1975, Comments on time series and causal concepts in business cycle research, presented at the Seminar on New Methods in Business Cycle Research, Nov. 13–14 (Minneapolis, MN).