AUTOREGRESSIVE MODELLING AND CAUSAL ORDERING OF ECONOMIC VARIABLES*

Cheng Hsiao

Econometric Research Program
Research Memorandum No. 274

December 1980

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

AUTOREGRESSIVE MODELLING AND CAUSAL ORDERING OF ECONOMIC VARIABLES *

by

Cheng Hsiao

Department of Economics
Princeton University
Princeton, N.J. 08544

and

Institute for Policy Analysis
University of Toronto
Toronto, Canada M5S 1A1

FIRST DRAFT - May, 1979
REVISED - November, 1980

Abstract

The use of autoregressive modelling of economic variables is explored. A multivariate generalization of Wiener-Granger notion of causality is suggested. Propositions about population properties of various causal events are derived. These propositions may be used to interpret the results as well as to check the empirical implications of various variants of models and in ruling out a number of variants as being inconsistent with prior theorems. Canadian money, income, and interest rate are used as an example to illustrate the methodology.

AUTOREGRESSIVE MODELLING AND CAUSAL ORDERING

OF ECONOMIC VARIABLES*

by

Cheng Hsiao

1. Introduction

Recently, Sims (1980) has persuasively argued the advantages of constructing economic models without restrictions based on supposed a priori knowledge. He suggests that in a first stage model construction we treat all variables as jointly dependent and fit an unconstrained vector autoregression to avoid an infection of the model with spurious or false restrictions.

It is one thing to fit a vector AR model; it is another thing to interpret the empirical results. One of the central concepts in the discussion of economic laws or econometric models is that of "causality". Unfortunately, as Granger (1980) has remarked, "the definition of causality as a concept has more people knowing what they do not like than knowing what they do like". Zellner (1978) has suggested a definition of causality in terms of "predictability according to well thought out economic laws". While we agree with this definition, there is a practical difficulty in applying it to model construction with time series techniques. The definition essentially assumes that in order to establish a causal ordering we must have a priori knowledge. But an empirical model building strategy such as that suggested by Sims (1980) is recommended when economists disagree about the set of laws governing economic relationships. If a model is specified according to a set of incorrect laws, the estimation is biased, and the model may thereby become useless as a framework within which to do formal statistical tests. In this sense, a definition of causality not relying on economic theory may provide useful insights to many problems.

Granger (1969) has suggested a definition of causality that makes no mention of economic laws. It is based on the stochastic nature of the variables and its central feature is the direction of the flow of time. It is a purely statistical criterion relying entirely on the assumption that the future cannot cause the past. This definition is at variance with the philosophical definition in certain important aspects (Zellner (1978)). In certain cases it may even obscure conventional causal ordering (Sims (1977)). However, Sims has also demonstrated that these possibilities exist only under special, rather restrictive assumptions. Geweke (1978) has shown that in the complete dynamic simultaneous equation model exogenous variables cause endogenous variables in the sense of Granger. It does seem that there will be a large class of applications where the causal ordering arising from the most plausible behavioural structure will be consistent with a Granger ordering. Therefore we shall adopt the basic Granger formulation in this paper despite the disagreement about its appropriateness. We must caution readers that the range of applicability of Granger's concept may be limited. But we also feel that in the circumstances when theorists disagree about the underlying structural relationships an empirical investigation of statistical regularities among economic variables without relying on supposed a priori restrictions may help shed light on the problem.

In this paper we intend to provide a basic framework to explain the causal relationship of a multivariate time series model based on the Wiener-Granger notion of causality. Skoog (1976) has also worked on this problem before. But his focus is mainly on the incompatibility of characterizing a causal variable in terms of predictability in a multivariate set up. Our focus here is on providing a Granger causal ordering of the events and on the reconciliation of the disparity between the results obtained from bivariate and multivariate analysis. In Section 2 we generalize Granger's notion of causality to make some provision for spurious and indirect causality which may arise in multivariate analysis. Section 3 characterizes mathematical (or population) properties of causal events. Section 4 proves the statements made in Section 3. Section 5 uses post-war Canadian money, income and interest rate data to illustrate the disparity which might arise between the bivariate and trivariate analysis and the reconciliation one might attempt based on the generalized Granger notion of causality. Conclusions are stated in Section 6.

2. Patterns of Causality

To provide a basic framework for interpreting a multivariate AR model, we shall generalize Granger's notion of causality in this section. Our aim is to identify whether a variable causes another variable directly, indirectly, or spuriously. Of course, theoretically, the notions to be discussed below can be further generalized into (n-1) different levels of causal ordering in an n variable system in a manner similar to that of

McElroy (1978) in a different context. We shall not attempt this here in the interest of simplicity of exposition.

Let $\{y, x, z\}$ be full rank, zero mean, joint covariance stationary, purely linearly indeterministic processes.[1] The analysis will remain unchanged if we let x be an r component, y be an s component and z be a q component stationary processes. However, for simplicity of exposition we shall assume that x, y, and z are univariate.

It has been shown by Wold (e.g. see Rosanov (1961)) that a regular full rank stationary process $\{y, x, z\}$ possess a unique one-sided moving average (MA) representation of the form

$$(1) \qquad \begin{pmatrix} y_t \\ x_t \\ z_t \end{pmatrix} = \xi_t + \Gamma_1 \xi_{t-1} + \Gamma_2 \xi_{t-2} + \cdots ,$$

where ξ_t is a three component zero mean orthogonal process (i.e., $E\xi_t = 0$, $E\xi_t \xi_s' = \delta_{t,s}\Omega$, $\delta_{t,s} = 1$, $t = s$, $\delta_{t,s} = 0$, $t \neq s$).

Under fairly general conditions (Masani (1966)), (1) also admits an autoregressive representation

$$(2) \qquad (I - \Psi_1 L - \Psi_2 L^2 - \ldots) \begin{pmatrix} y_t \\ x_t \\ z_t \end{pmatrix} = \xi_t$$

where

$$I - \Psi_1 L - \ldots = \begin{bmatrix} 1-\psi_{11}(L) & -\psi_{12}(L) & -\psi_{13}(L) \\ -\psi_{21}(L) & 1-\psi_{22}(L) & -\psi_{23}(L) \\ -\psi_{31}(L) & -\psi_{32}(L) & 1-\psi_{33}(L) \end{bmatrix}$$

$$= (I + \Gamma_1 L + \Gamma_2 L^2 + \ldots)^{-1}$$

and L is the lag operator, $Ly_t = y_{t-1}$. Typical elements of $\psi_{ij}(L)$ are given by $\sum_{\ell=1}^{\infty} \psi_{ij\ell} L^{\ell}$.

We note that if we multiply ξ by the lower triangular matrix P (such that $P\Omega P' = I$), (1) can be transformed as

$$(3) \quad \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \Phi_o \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} + \Phi_1 \begin{bmatrix} u_{t-1} \\ v_{t-1} \\ w_{t-1} \end{bmatrix} + \ldots\ldots$$

$$= \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) & \phi_{13}(L) \\ \phi_{21}(L) & \phi_{22}(L) & \phi_{23}(L) \\ \phi_{31}(L) & \phi_{32}(L) & \phi_{33}(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix}$$

$$= \Phi(L) \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix}$$

where

$$E \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} (u_s \; v_s \; w_s) = \delta_{t,s} I \quad ,$$

and Φ_o is an invertible, lower triangular matrix with positive elements on the diagonal. Either (1) or (3) is an identified process (Hannan (1969)). For ease of exposition, we shall take (2) and (3) as the AR and MA representation of $\{y, x, z\}$ in this paper.

Let $\bar{X}_t = \{x_s: s < t\}$, $\bar{\bar{X}}_t = \{x_s: x \leq t\}$, and similarly define $\bar{Y}_t$, $\bar{\bar{Y}}_t$, $\bar{Z}_t$, $\bar{\bar{Z}}_t$. Let A_t be the relevant information set accumulated since time $t-1$. Thus, A_t can be considered as a stochastic process, including $\{y_t, x_t, z_t\}$. We define $A_t - x_t$ as the set of elements in A_t without the element x_t. The set $\bar{A}_t$, $\bar{\bar{A}}_t$, $\overline{A_t - X_t}$, $\overline{\overline{A_t - X_t}}$ are defined analogously to $\bar{X}_t$ and $\bar{\bar{X}}_t$. Denote by $\sigma^2(y|\bar{A})$ the mean square error of the minimum mean square linear prediction error of y_t given information set $\bar{A}_t$.

We follow Wiener-Granger in defining various patterns of causality in terms of predictability of a variable. We first strengthen Granger's (1969) definition of causality as:

Definition 1 (Direct Causality): If $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$, and $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$, then we say z causes y directly, denoted by $z \Rightarrow y$.

This definition says that z causes y directly only when present y can be better predicted, in the mean square prediction error sense, by using past values of z, no matter which information set is used. We emphasize this aspect because, as the following discussion will show, there are cases where past z may help in predicting present y when one information set is used, but may not help in predicting present y when another information set is used.

Definition 2 (Direct Feedback): If $z \Rightarrow y$, and $y \Rightarrow z$, then we say that direct feedback occurs between y and z, denoted by $z \Leftrightarrow y$.

Definition 3 (No Causality): z does not cause y when either (i) $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-X-Z})$ or (ii) $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ and $\sigma^2(x|\bar{A}) = \sigma^2(x|\overline{A-Z})$.

Condition (i) implies that the best linear predictor of y makes use of past values of y only. Condition (ii) implies that past x and y are jointly sufficient for predicting present y and x.[2] However, Definition 3 does not exhaust all the cases where one would normally consider as an indication of no causality. There are cases where we may find $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$ but y and z may have no relationship at all. For example, consider the following model

$$(4) \quad \begin{pmatrix} y_t \\ x_t \\ z_t \end{pmatrix} = \begin{pmatrix} 1 & .5L & 0 \\ 0 & 1 & .5L \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_t \\ v_t \\ w_t \end{pmatrix} .$$

In this system, the white noise innovations (u_t, v_t, w_t) are mutually orthogonal, so z does not cause y by virtue of their independence, hence $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$. Yet when past x is used to predict y, it is advantageous to also use past z. This can be seen by taking the inverse of (4),

$$(5)\quad \begin{bmatrix} 1 & -.5L & .25L^2 \\ 0 & 1 & -.5L \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} .$$

That is, $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$. An analogous case in regression analysis is where y depends on x^* , but not on z , yet x^* is unobservable with x serving as its proxy. If z is correlated with the noise in the proxy variable x , then we can use z to eliminate or reduce the noise in x so that a more accurate prediction for y may be achieved. However, z appears in the y equation not because z somehow causes y , but rather because it serves as a purifying variable for x . We call this situation spurious causality.

Definition 4 (Type I Spurious Causality): If $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ and $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$, then we say Type I spurious causality from z to y occurs.

Definition 4 says that relying on $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$ as an indication of causality may lead to spurious conclusions. On the other hand, to indicate that $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ as an indication of no causality from z to y may also lead to wrong inferences. There are cases that even though z does not help predict y when other variables are used as predictor, it is indeed the primary driving force for y . We call this situation indirect causality.

Definition 5 (Indirect Causality): If $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z}) < \sigma^2(y|\overline{A-X}) < \sigma^2(y|\overline{A-X-Z})$ and $\sigma^2(x|\bar{A}) < \sigma^2(x|\overline{A-Z})$, $\sigma^2(x|\bar{X},\bar{Z}) < \sigma^2(x|\bar{X})$, then we say that z causes y indirectly, denoted by $z \rightarrow y$.

There are also cases where z may be found to cause y in a bivariate analysis, while in fact there is a third series x which causes both y and z (e.g. Granger (1969)). Past z is used as a predictor for present y in the bivariate analysis only because some other statistical variable of importance has not been included. We call this case another type of spurious causality.

Definition 6 (Type II Spurious Causality): When condition (ii) of no causality holds, but $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z}) < \sigma^2(y|\overline{A-X}) < \sigma^2(y|\overline{A-X-Z})$ and $\sigma^2(z|\bar{A}) < \sigma^2(z|\overline{A-X})$, $\sigma^2(z|\bar{Z}, \bar{X}) < \sigma^2(z|\bar{Z})$, we say Type II spurious causality from z to y occurs.

Both definitions 5 and 6 say that past z will not help predict present y when past x are used but will help predict present y when past x are not used. However, in Definition 5 z drives x which in turn causes y . On the other hand, Definition 6 assumes that x is the primary driving force for both y and z . Past z only serves as a proxy for the missing x .

Granger (1969) also suggests a definition of instantaneous causality. Here we prefer to ignore this case as it seems more natural to treat it as contemporaneous correlation (e.g. see Caines (1976) and Hsiao (1979b)).

3. Characterization of Causal Events

In this section we state some basic relationships among variables for various causal events. The proofs are in the next section.

Theorem 1: z does not cause y if and only if the following equivalent conditions hold:

(i) The moving average operator $\Phi(L)$ is lower block triangular.

(ii) The autoregressive operator $\Psi(L)$ is lower block triangular.

This theorem is a straightforward application of theorems proved by Caines and Chan (1975), Geweke (1978), Granger (1969), Pierce and Haugh (1977), and Sims (1972), etc.

Corollary 1: If z does not cause y, then either z does not cause x, or neither z nor x causes y.

This corollary follows straightforwardly from the lower block triangularity of $\Psi(L)$ and $\Phi(L)$.

Corollary 2: If $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ and $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$, then z does not cause y.

Note that Corollary 2 is a sufficient condition for z not causing y. It is not a necessary condition. It is possible that z does not cause y yet $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$. This is referred to as Type I spurious causality.

Theorem 2: Type I spurious causality from z to y occurs if and only if the following equivalent conditions hold:

(i) in the MA representation (3) there exists a $C(L) = c_o + c_1 L + c_2 L^2 + \ldots$ such that

$$[\phi_{31}(L) \quad \phi_{32}(L)] = C(L)[\phi_{11}(L) \quad \phi_{12}(L)] \quad , \tag{6}$$

and $\phi_{13\ell} = 0$ for all ℓ, $\phi_{12\ell} \neq 0$, $\phi_{23\ell} \neq 0$ for some ℓ.

(ii) in the AR representation (2), $\psi_{32\ell} = 0$ for all ℓ, $\psi_{13\ell} \neq 0$ for some ℓ, and there exists a nonzero $C(L)$ such that

$$[\psi_{12}(L) \quad \psi_{13}(L)] = C(L)[\psi_{22}(L) \quad \psi_{23}(L)] \quad . \tag{7}$$

Corollary 3: A sufficient condition for the existence of Type I spurious causality from z to y is that $\phi_{31\ell} = \phi_{32\ell} = \phi_{13\ell} = 0$ for all ℓ and $\phi_{12\ell} \neq 0$, $\phi_{23\ell} \neq 0$ for some ℓ .

The situations where z is merely a proxy variable for the important left-out variables and where z is a primary driving force may be distinguished according to the following theorem:

Theorem 3: A necessary condition for the existence of Type II spurious causality from z to y is that:

(i) in the moving average representation (3), $\phi_{13\ell} = 0$, $\phi_{23\ell} = 0$ for all ℓ , $\phi_{12\ell} \neq 0$ for some ℓ , and there does not exist a C(L) such that (6) holds.

(ii) in the AR representation (2), $\psi_{13\ell} = 0$, $\psi_{23\ell} = 0$ for all ℓ , and $\psi_{12\ell} \neq 0$, $\psi_{32\ell} \neq 0$ for some ℓ .

Theorem 4: z does not cause y directly, but causes y indirectly if and only if the following equivalent conditions hold:

(i) there exists a nonzero C(L) such that

(8) $$[\phi_{12}(L) \quad \phi_{13}(L)] = C(L)[\phi_{22}(L) \quad \phi_{23}(L)] \qquad ,$$

and $\phi_{23\ell} \neq 0$ for some ℓ in (3).

(ii) $\psi_{13\ell} = 0$ for all ℓ and $\psi_{12\ell} \neq 0$, $\psi_{23\ell} \neq 0$ for some ℓ in (2).

Corollary 4: If $z \Rightarrow x$, $x \Rightarrow y$, but z does not cause y directly, then either z causes y indirectly or z causes y spuriously (in a Type I sense).

4. Proof of Various Characterizations of Causal Events[3]

Let the notation $H_y(t)$ stand for the completion with respect to the mean-square norm of the linear space of random variables spanned by y_s for $s \leq t$. Let u_t be the difference between y_t and the projection of y_t on $H_{y,x,z}(t-1)$. Let $\tilde{v}_t$ be the difference between x_t and the projection of x_t on $H_{y,x,z}(t-1)$. Let v_t be that part of $\tilde{v}_t$ which is orthogonal to u_t. Similarly, let w_t be that part of the difference between z_t and the projection of z_t on $H_{y,x,z}(t-1)$ which is orthogonal to u_t and v_t. By definition, u_t, v_t, and w_t are contemporaneously uncorrelated and are uncorrelated with past values of each other. Also $H_{u,v,w}(t)$ is identical to $H_{y,x,z}(t)$, and $\{y_t, x_t, z_t\}$ has a moving average representation of the form (3) (Rozanov (1967)).

We similarly define u_t^* as the difference between y_t and the projection of y_t on $H_{y,x}(t-1)$ and v_t^* as that part of the difference between x_t and the projection of x_t on $H_{y,x}(t-1)$ which is orthogonal to u_t^*; hence $H_{u^*,v^*}(t)$ is identical to $H_{y,x}(t)$. We also let u_t^{**} denote the difference between y_t and the projection of y_t on $H_{y,z}(t-1)$ and w_t^{**} denote that part of the difference between z_t and the projection of z_t on $H_{y,z}(t-1)$ which is orthogonal to u_t^{**}. Thus, $H_{y,z}(t)$ is identical to $H_{u^{**},w^{**}}(t)$.

Lemma 1: If $\phi_{12\ell} \neq 0$ and $\phi_{13\ell} \neq 0$ for some ℓ in (3), then the $\{y, x\}$ process has an MA represenation of the form

$$(9) \quad \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^*(L) & \phi_{12}^*(L) \\ \phi_{21}^*(L) & \phi_{22}^*(L) \end{pmatrix} \begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix}$$

with $\phi^*_{12\ell} \neq 0$ for some ℓ .

Proof: We denote by $D(t)$ the orthogonal complement of $H_{y,x}(t)$ in the subspace $H_{y,x,z}(t)$:

(10) $$D(t) = H_{y,x,z}(t) \ominus H_{y,x}(t) \quad .$$

Then, by construction,

(11) $$u^*_t \in D(t-1) \oplus u_t$$

where $\oplus$ denotes the direct sum. By construction we know that $H_{u,v}(t) \subseteq H_{u*,v*}(t)$ and $D(t) = H_{u,v,w}(t) \ominus H_{u*,v*}(t)$. Hence $D(t) \subseteq H_{u,v,w}(t) \ominus H_{u,v}(t) = H_w(t)$. Therefore[4]

(12) $$H^*_u(t) \subseteq H_u(t) \oplus H_w(t-1) \quad .$$

By Wold's decomposition theorem (Rozanov (1967)), $\{y, x\}$ are obtained from moving averages of uncorrelated processes as follows:

(13) $$\begin{aligned} y_t &= \phi_{11}(L)u_t + \phi_{12}(L)v_t + \phi_{13}(L)w_t \\ &= \phi^*_{11}(L)u^*_t + \phi^*_{12}(L)v^*_t \quad . \end{aligned}$$

(14) $$\begin{aligned} x_t &= \phi_{21}(L)u_t + \phi_{22}(L)v_t + \phi_{23}(L)w_t \\ &= \phi^*_{21}(L)u^*_t + \phi^*_{22}(L)v^*_t \quad . \end{aligned}$$

Given that $H_v(t) \perp H_{u,w}(t)$ and (12), we know that $\{y, x\}$ may not be represented as

(15) $$y_t = \tilde{\phi}_{11}(L)u^*_t + \tilde{\phi}_{12}(L)v_t$$

(16) $$x_t = \tilde{\phi}_{21}(L)u^*_t + \tilde{\phi}_{22}(L)v_t \quad .$$

Thus

(17) $$H_v(t) \subseteq H_v^*(t)$$

However, $\phi_{12\ell} \neq 0$ for some ℓ means that the projection of y_t on $H_v(t-1)$ is non-zero. Consequently, the projection of y_t on $H_v^*(t-1)$ is non-zero as well. Hence $\phi_{12\ell}^* \neq 0$ for some ℓ.

Proof of Corollary 2: Expressing $\{y_t, x_t\}$ in terms of the u_t^*, v_t^*, we have (9). Since $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$, we know that the projection of y_t on $H_{y,x,z}(t-1)$ lies in $H_{y,x}(t-1)$, and therefore $u_t^* = u_t$. Furthermore $u_t \perp H_{v,w}(t)$ and $u_t \perp H_{v^*}(t)$; hence, $\phi_{11}^*(L) = \phi_{11}(L)$ and $\phi_{21}^*(L) = \phi_{21}(L)$. If $\phi_{12\ell}^* = 0$ for all ℓ, then $\phi_{12\ell} = 0$ and $\phi_{13\ell} = 0$ for all ℓ and it is trivial that z does not cause y. If $\phi_{12\ell}^* \neq 0$ for some ℓ, by (3) and (9) we have (Ansley, Spivey and Wrobleski (1977)).

(18) $$\phi_{12}^*(L)v_t^* = \phi_{12}(L)v_t + \phi_{13}(L)\, w_t$$

(19) $$\phi_{22}^*(L)v_t^* = \phi_{22}(L)v_t + \phi_{23}(L)w_t$$

Letting $C(L) = \dfrac{\phi_{12}^*(L)}{\phi_{22}^*(L)}$, we have

(20) $$[\phi_{12}(L) \quad \phi_{13}(L)] = C(L)[\phi_{22}(L) \quad \phi_{23}(L)]$$

Theorem 1 says that $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ if and only if there exists a moving average representation for $\{y_t, z_t\}$ of the form

(21) $$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{**}(L) & 0 \\ \phi_{31}^{**}(L) & \phi_{33}^{**}(L) \end{pmatrix} \begin{pmatrix} u_t^{**} \\ w_t^{**} \end{pmatrix}$$

By lemma 1, a necessary condition to construct (21) from (3) is that $\phi_{13\ell} = 0$ for all ℓ. However, (20) says that $\phi_{13\ell} = 0$ for all ℓ, so are $\phi_{23\ell} = 0$. From theorem 1 we know that $\phi_{13\ell} = 0$ and $\phi_{23\ell} = 0$ for all ℓ if and only if z does not cause y.

Proof of Theorem 2: We note that $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ implies that the moving average representation for $\{y_t, z_t\}$ has the form (21) with $\phi^{**}_{13\ell} = 0$ for all ℓ. If both $\phi_{12\ell} = 0$ and $\phi_{13\ell} = 0$ for all ℓ, the projection of y_t on $H_{y,x,z,}(t-1)$ lies in $H_y(t-1)$ alone. That is, $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-X-Z})$, and neither z nor x causes y. On the other hand, if both $\phi_{12\ell} \neq 0$ and $\phi_{13\ell} \neq 0$ for some ℓ, then by Lemma 1 $\phi^{**}_{13\ell} \neq 0$ for some ℓ. Therefore we assume $\phi_{12\ell} \neq 0$ for some ℓ and $\phi_{13\ell} = 0$ for all ℓ.[5] With $\phi_{11\ell} \neq 0$, $\phi_{12\ell} \neq 0$ for some ℓ, the only way for (21) to hold is that

(22) $$\phi^{**}_{11}(L)u^{**}_t = \phi_{11}(L)u_t + \phi_{12}(L)v_t$$

(23) $$\phi^{**}_{31}(L)u^{**}_t = \phi_{31}(L)u_t + \phi_{32}(L)v_t .$$

Such a representation exists if and only if there exists a $C(L)$ such that

(24) $$[\phi_{31}(L) \quad \phi_{32}(L)] = C(L)[\phi_{11}(L) \quad \phi_{12}(L)]$$

If $\phi_{23\ell}$ also equal zero for all ℓ, then $\Phi(L)$ is lower block triangular, and hence $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$. This is a contradiction and therefore $\phi_{23\ell} \neq 0$ for some ℓ. Together with $\phi_{12\ell} \neq 0$ for some ℓ, this implies that $\psi_{13\ell} \neq 0$ for some ℓ, i.e., $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$.

Condition (ii) can be derived by taking the inverse of the $\Phi(L)$ matrix.

Proof of Corollary 3: Condition (i) of theorem 2 is automatically satisfied with $C(L) \equiv 0$.

Proof of Theorem 3: We first prove condition (i). Theorem 1 says that $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$ if and only if

(25) $$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{**}(L) & \phi_{13}^{**}(L) \\ \phi_{31}^{**}(L) & \phi_{33}^{**}(L) \end{pmatrix} \begin{pmatrix} u_t^{**} \\ w_t^{**} \end{pmatrix}$$

with $\phi_{13\ell}^{**} \neq 0$ for some ℓ. However, condition (ii) of no causality from z to y implies $\phi_{13\ell} = 0$ and $\phi_{23\ell} = 0$ for all ℓ in (3). Therefore, $\{y_t, z_t\}$ has the following representation

(26) $$y_t = \phi_{11}(L)u_t + \phi_{12}(L)v_t$$

(27) $$z_t = \phi_{31}(L)u_t + \phi_{32}(L)v_t + \phi_{33}(L)w_t \;.$$

From $\sigma^2(z|\bar{A}) < \sigma^2(z|\overline{A-X})$ and $\sigma^2(z|\bar{X}, \bar{Z}) < \sigma^2(z|\bar{Z})$ we know that $\phi_{32\ell} \neq 0$ for some ℓ. Given (26) and (27), as the proof of corollary 2 shows that a necessary condition for the existence of (25) is $\phi_{12\ell} \neq 0$ for some ℓ, and no existence of $C(L)$ such that

(28) $$[\phi_{31}(L) \quad \phi_{32}(L)] = C(L)[\phi_{11}(L) \quad \phi_{12}(L)]$$

Condition (i) implies condition (ii) follows straightforwardly from taking the inverse of $\Phi(L)$. To show that condition (ii) implies condition (i), we note that $\psi_{13\ell} = 0$, $\psi_{23\ell} = 0$ for all ℓ and $\psi_{12\ell} \neq 0$ for some ℓ jointly imply $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z}) < \sigma^2(y|\overline{A-X})$ and $\sigma^2(x|\bar{A}) = \sigma^2(x|\overline{A-Z})$. That is $\phi_{13\ell} = 0$, $\phi_{23\ell} = 0$ for all ℓ and $\phi_{12\ell} \neq 0$ for some ℓ. $\psi_{32\ell} \neq 0$ for some ℓ implies that there does not exist a $C(L)$ such that (28) holds.

Proof of theorem 4: We first prove condition (i). We note that $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ if and only if (20) holds. If $C(L) = 0$, then this implies that $\phi_{12\ell} = 0$, $\phi_{13\ell} = 0$ for all ℓ. That is, y can be viewed as generated by an innovation process that is uncorrelated with the innovations in the x and z processes. Therefore $C(L)$ must be nonzero for $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$ to hold. On the other hand if $\phi_{23\ell} = 0$ for all ℓ, by (20) $\phi_{13\ell} = 0$ for all ℓ. Then by theorem 1, $\sigma^2(x|\bar{A}) = \sigma^2(x|\overline{A-Z})$. (Neither can $\phi_{12\ell} = 0$ for all ℓ, since this will imply $\phi_{22\ell} = 0$ for all ℓ.) When $\phi_{12\ell} \neq 0$, $\phi_{13\ell} \neq 0$ for some ℓ, by Lemma 1 the $\{y_t, z_t\}$ process will have $\phi^{**}_{13\ell} = 0$ for some ℓ in (25), that is $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$.

Condition (ii) can be proved by taking the inverse of $\Phi(L)$.

Proof of Corollary 4: Definition 1 says that if z does not cause y directly either $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ or $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ or both. We first show that $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ and $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ cannot hold simultaneously under the assumption that $z \Rightarrow x$ and $x \Rightarrow y$. We note that $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ implies (20) holds. Therefore, $\phi_{12\ell} \neq 0$ for some ℓ unless $C(L)$ is identically equal to zero which is ruled out by the assumption that $x \Rightarrow y$. Thus by Lemma 1, $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$ implies that $\phi_{13\ell} = 0$ for all ℓ. By (20), $\phi_{23\ell} = 0$ for all ℓ. This contradicts the assumption that $z \Rightarrow x$. Therefore, either $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ and $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$ or $\sigma^2(y|\bar{A}) < \sigma^2(y|\overline{A-Z})$ and $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$. The former is indicated as indirect causality, the latter spurious causality of type I.

We can also prove these assertions directly. Suppose $\sigma^2(y|\bar{A}) = \sigma^2(y|\overline{A-Z})$ and $x \Rightarrow y$, then there exists a non-zero $C(L)$ such that (20) holds. $\phi_{23\ell} \neq 0$ for some ℓ are implied by $z \Rightarrow x$; therefore $\phi_{13\ell} \neq 0$ for some ℓ. $\phi_{12\ell} \neq 0$ for some ℓ follows from $\phi_{22\ell} \neq 0$ for some ℓ.

Thus, by Lemma 1, $\sigma^2(y|\bar{Y}, \bar{Z}) < \sigma^2(y|\bar{Y})$; that is, z causes y indirectly.

Suppose $\sigma^2(y|\bar{Y}, \bar{Z}) = \sigma^2(y|\bar{Y})$. A necessary condition for this to hold is that $\phi_{13\ell} = 0$ for all ℓ and (24) hold. $x \Rightarrow y$ and $z \Rightarrow x$ imply that $\phi_{12\ell} \neq 0$, $\phi_{23\ell} \neq 0$ for some ℓ . Thus condition (i) of Theorem 2 is satisfied. Therefore spurious causation of type I operates over z to y .

5. Autoregressive Modelling of Canadian Money, Income, and Interest Rate

Although time series approach to model construction is appealing in the sense that it gives an impartial view to competing theorems, many problems remain. For instance, the determination of appropriate time series model from finite observations; reliability of large sample estimation and testing procedures when sample information is limited; transforming variables to induce stationarity and the role of seasonality are all complex issues which the profession has not reached a concensus yet (e.g. see Zellner (1977)). A study of any one of these problems would be a challenging research program. Doing all of these things in one paper would be so challenging as to be impossible. In this section we use the quarterly Canadian money stock (M2), nominal GNP, and bank rate (BR) from 1955 I to 1977 IV as an example to check the reliability of the inferences one may draw from an autoregressive model.[6]

The relationship between money and income has been much debated in the economic literature (e.g. Friedman (1970), Brunner and Meltzer (1966), Tobin (1970)). On one side are the monetarists, who view money as an independent source of economic disturbance. On the other side are critics of this view, who say that money is a passive adapter to business conditions with little independent influence. Previous tests of causality on this problem (Barth and Bennett (1974), Hsiao (1979a), Sims (1972)) have concentrated on the money and income variables. We now extend the analysis to include the interest rate. However, it does not mean that in the money-income causality analysis these three variables offer a complete information set, nor do we intend to give an answer to this debate. The purpose of this exercise is meant to show how the conditional information set affects Granger causal ordering and to see whether the problems of collinearity, shortages of degrees of freedom, etc. will affect the qualitative conclusions drawn from an autoregressive model when different methods are used to identify a model.

The basic model we have in mind relates rates of change of money and income to the level of the interest rate (e.g. see Cagan (1966)). In a hypothetical long-run moving equilibrium it is theoretically appealing to regard "normal" as constant percentage rates of change in the money and income variables, which are related to a constant nominal interest rate. We then take the first difference of each of these variables (i.e. the second differences of the logarithm of the money and income variables and the first difference of the logarithm of the bank rate variable) to remove any trend. Such prefiltering regards changes in the percentage rate of change in money and percentage change in the interest rate as the main features of monetary behavior contributing to the generation of cycles (changes in the percentage change in income) (Friedman (1961)).[7]

We use two procedures to identify a vector AR model. One is the conventional. We first determine the order of the unrestricted AR model, then formulate and test hypotheses with economic content in the second stage.

The other is a procedure suggested by Hsiao (1979a,b) using the concept of final prediction error (Akaike (1969a, b)). This procedure combines the determination of the order of an autoregressive form with some second stage hypotheses testing. We first report results based on the standard procedure.

The choice of the appropriate order of an unconstrained vector AR model is a multiple decision problem. Suppose that the maximum order of dependence is Q. We want to decide which of the following exclusive sets the parameter point $(\Psi_1,\ldots,\Psi_Q)$ belongs:

$$
\begin{aligned}
&H_Q: \quad \Psi_Q \neq 0 \quad , \\
&H_{Q-1}: \quad \Psi_Q = 0 \quad , \qquad \Psi_{Q-1} \neq 0 \quad , \\
&\quad \vdots \\
&H_1: \quad \Psi_Q = \Psi_{Q-1} = \ldots = \Psi_2 = 0 \quad , \qquad \Psi_1 \neq 0 \quad , \\
&H_0: \quad \Psi_Q = \Psi_{Q-1} = \ldots = \Psi_1 = 0 \quad .
\end{aligned}
\tag{29}
$$

The set H_q implies that the dependence is of order q. As shown by Anderson (1971, ch. 6.3), an optimum procedure is to test $\Psi_Q = 0$, $\Psi_{Q-1} = 0 \ldots$ in turn until either one rejects such a hypothesis, say rejects $\Psi_q = 0$, and hence decides H_q.

Because the maximum allowable order for using FIML option in TSP 78 in Toronto is a ninth order one we let $Q = 9$. The likelihood ratio statistic for H_9 is 12.112, H_8 is 7.176, H_7 is 4.696, H_6 is 23.79. Each H_q has 9 degrees of freedom. Based on these observations we choose $q = 6$

as the order of the unconstrained vector autoregressive process.[8]

To proceed to test the causal ordering of money, income and interest rate, we note that except for the Type I spurious causality (Theorem 2). other patterns of causality can be expressed in terms of zero constraints on $\Psi(L)$ matrix in an AR specification. Since the general linear constraints of the form (7) is difficult to formulate, we proceed to separate the test of causal ordering into two stages. In the first stage we check the zero constraints of a vector AR model. If the first stage hypotheses testing leads to the acceptance of zero constraints which may imply the existence of Type I spurious causality, we then check whether other linear constraints exist in the second stage by comparing the best models under various information sets according to the definitions and corollaries stated in Sections 2 and 3 (e.g. the best model of y given $\overline{Y}$, $\overline{X}$, and the best model of y given $\overline{Y}$, $\overline{X}$, $\overline{Z}$).

According to various patterns of causality, we know that in order to identify whether z causes y directly, indirectly or spuriously we have to test (i) $\psi_{13}(L) = 0$; (ii) $\psi_{12}(L) = \psi_{13}(L) = 0$; (iii) $\psi_{13}(L) = \psi_{23}(L) = 0$; (iv) $\psi_{12}(L) = \psi_{13}(L) = \psi_{23}(L) = 0$; (v) $\psi_{13}(L) = \psi_{23}(L) = \psi_{32}(L) = 0$ in turn. Treating an unconstrained sixth order AR model as the maintained hypothesis, we may either test each of these hypotheses against the maintained hypothesis or we may perform the following sequential test

$$H_1: \ \psi_{13}(L) \neq 0 ,$$

$$H_2: \ \psi_{13}(L) = 0 , \qquad \psi_{12}(L) \neq 0$$

$$H_3: \ \psi_{13}(L) = 0 , \qquad \psi_{23}(L) \neq 0$$

$$(30) \qquad \begin{aligned} &H_4:\ \psi_{13}(L) = \psi_{23}(L) = 0 \quad , && \psi_{12}(L) \neq 0 \quad , \\ &H_5:\ \psi_{13}(L) = \psi_{23}(L) = 0 \quad , && \psi_{32}(L) \neq 0 \quad , \\ &H_6:\ \psi_{13}(L) = \psi_{23}(L) = \psi_{32}(L) = 0 \quad , && \psi_{12}(L) \neq 0 \quad , \\ &H_7:\ \psi_{13}(L) = \psi_{23}(L) = \psi_{32}(L) = 0 \quad , && \psi_{21}(L) \neq 0 \quad , \\ &H_8:\ \psi_{13}(L) = \psi_{23}(L) = \psi_{32}(L) = 0 \quad , && \psi_{31}(L) \neq 0 \quad , \end{aligned}$$

The advantage of testing each of these hypotheses against the maintained hypothesis is that the size of the test for each of these hypotheses is the same. The disadvantage is that an otherwise significant coefficients might be contaminated by other insignificant coefficients, hence making it difficult to reject the null hypothesis. The advantage of testing each hypothesis sequentially by treating the previously accepted null hypothesis as the maintained hypothesis is that the test is more sharply focused. The disadvantage is that a test of ψ_{ij} may either be accepted or rejected depending on the order it is tested. Using a fixed significance level α in a sequential test, we increase the Type I error from $p_1 = \alpha$ to $p_2 = \alpha + (1-\alpha)\alpha$, to $p_3 = \alpha + 2(1-\alpha)\alpha + (1-\alpha)^2\alpha$, ... when we change the order of testing $\psi_{ij} = 0$ from the first to the second and so on. To balance the desirability of the sensitivity to nonzero constraints with not favoring a particular $\psi_{ij} \neq 0$, we let $y=(1-L) \log BR$, $x = (1-L)^2 \log GNP$, $z = (1-L)^2 \log M2$ and use full information maximum likelihood method to estimate models with different zero constraints. The results are reported in Table 1.

As we can see from Table 1, for the test of one or two ψ_{ij}'s = 0, some have likelihood values very close to the unconstrained model (model 1).

Table 1

The Log-likelihood Value of Various 6th Order Autoregressive Models

Model	Log-likelihood Value
1. all $\psi_{ij} \neq 0$	556.146
2. $\psi_{13} = 0$	555.08
3. $\psi_{23} = 0$	554.15
4. $\psi_{12} = 0$	551.28*
5. $\psi_{21} = 0$	548.754****
6. $\psi_{31} = 0$	550.39**
7. $\psi_{32} = 0$	554.01
8. $\psi_{13} = \psi_{23} = 0$	553.025
9. $\psi_{12} = \psi_{32} = 0$	548.09
10. $\psi_{21} = \psi_{31} = 0$	542.928****
11. $\psi_{12} = \psi_{13} = 0$	550.269
12. $\psi_{21} = \psi_{23} = 0$	546.34**
13. $\psi_{31} = \psi_{32} = 0$	547.43*
14. $\psi_{12} = \psi_{13} = \psi_{23} = 0$	548.22
15. $\psi_{21} = \psi_{31} = \psi_{12} = 0$	540.51***
16. $\psi_{21} = \psi_{31} = \psi_{32} = 0$	539.97****
17. $\psi_{12} = \psi_{32} = \psi_{21} = 0$	541.51**
18. $\psi_{12} = \psi_{32} = \psi_{13} = 0$	547.08
19. $\psi_{13} = \psi_{23} = \psi_{21} = 0$	545.22
20. $\psi_{13} = \psi_{23} = \psi_{32} = 0$	550.84
21. $\psi_{13} = \psi_{23} = \psi_{31} = 0$	547.47
22. $\psi_{21} = \psi_{31} = \psi_{13} = 0$	542.11**
23. $\psi_{21} = \psi_{31} = \psi_{12} = 0$	538.14****
24. $\psi_{12} = \psi_{32} = \psi_{21} = 0$	540.78***
25. $\psi_{12} = \psi_{32} = \psi_{13} = 0$	546.10
26. $\psi_{12} = \psi_{13} = \psi_{23} = \psi_{32} = 0$	545.028
27. $\psi_{13} = \psi_{23} = \psi_{31} = \psi_{32} = 0$	544.383
28. $\psi_{12} = \psi_{13} = \psi_{23} = \psi_{31} = \psi_{32} = 0$	538.546
29. $\psi_{12} = \psi_{13} = \psi_{21} = \psi_{23} = \psi_{31} = \psi_{32} = 0$	530.772 **

* significant at 15% level
** significant at 10% level
*** significant at 5% level
**** significant at 1% level
} against Model 1

Among those with three ψ_{ij}'s = 0 and are insignificant at 15% level, the model with $\psi_{13} = \psi_{23} = \psi_{32} = 0$ (model 10) have the largest log-likelihood value, 550.84. The likelihood ratio test of this model against the unconstrained 6th order AR model has a chi square value of 10.612 with 18 degrees of freedom. Nor is the test of each of these ψ_{ij}'s = 0 against the unconstrained model significant. So we in turn take the model with $\psi_{13} = \psi_{23} = \psi_{32} = 0$ as the maintained hypothesis.

The further tests of $\psi_{12} = \psi_{13} = \psi_{23} = \psi_{32} = 0$ (model 26), $\psi_{13} = \psi_{23} = \psi_{31} = \psi_{32} = 0$ (model 27), $\psi_{12} = \psi_{13} = \psi_{23} = \psi_{31} = \psi_{32} = 0$ (model 28) against the new maintained hypothesis (model 20) have chi square values of 11.12, 11.91 with six degrees of freedom, and 24.6 with twelve degrees of freedom; indicating the rejection of these hypotheses at 10% level. On the other hand, if we test these hypotheses against unconstrained 6th order AR model (model 1), the likelihood ratio statistics are 22.236, 23.71 with 24 degrees of freedom, and 35.2 with 30 degrees of freedom, which are not significant at 15% level. Finally, the test of $\psi_{12} = \psi_{13} = \psi_{21} = \psi_{23} = \psi_{31} = \psi_{32} = 0$ (model 29) against the unconstrained model has chi square value of 50.75 with 36 degrees of freedom and against model 20 has value 40.17 with 18 degrees of freedom, which are significant at 15% and 1% level respectively.

Thus depending on whether model 20 or model 1 is chosen as the maintained hypothesis, we may either choose

$$(31)\quad \begin{bmatrix} (1-L)\log BR \\ (1-L)^2\log GNP \\ (1-L)^2\log M2 \end{bmatrix} = \begin{bmatrix} \psi_{11}^6(L) & \psi_{12}^6(L) & 0 \\ \psi_{21}^6(L) & \psi_{22}^6(L) & 0 \\ \psi_{31}^6(L) & 0 & \psi_{33}^{(6)}(L) \end{bmatrix} \begin{bmatrix} (1-L)\log BR \\ (1-L)^2\log GNP \\ (1-L)^2\log M2 \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

or

$$(32)\quad \begin{bmatrix} (1-L)\log BR \\ (1-L)^2\log GNP \\ (1-L)^2\log M2 \end{bmatrix} = \begin{bmatrix} \psi_{11}^{6}(L) & 0 & 0 \\ \psi_{21}^{6}(L) & \psi_{22}^{6}(L) & 0 \\ 0 & 0 & \psi_{33}^{6}(L) \end{bmatrix} \begin{bmatrix} (1-L)\log BR \\ (1-L)^2\log GNP \\ (1-L)^2\log M2 \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

as an identified time series model for interest rate, income and money where the superscript indicates the order of $\psi_{ij}(L)$. Their full information maximum likelihood estimates are reported in Tables 2 and 3.

If we choose (31) as the time series model, then according to theorems stated in section 3, there is a direct feedback between interest rate and income, a direct causality from interest rate to money and an indirect causality from income to money; but money neither causes income nor causes interest rate. If we choose (32) as the time series model, then income does not cause interest rate and neither interest rate, nor income cause money.

To check the validity of the above assertion and whether Type I spurious causality exists, we perform bivariate analysis on the income-interest rate, money-interest rate and money-income pairs. A priori we may expect that the order of lags would be increased when we move from a three dimensional analysis to a two dimensional analysis. Computationally, it is also possible to increase the value of Q . We therefore let $Q = 13$ and test $\Psi_Q^* = 0$, $\Psi_{Q-1}^* = 0, \ldots$ in turn. Each of these hypotheses has four degrees of freedom.

For the income-interest rate pair, the likelihood ratio statistic for $q = 13$ is 7.77, $q = 12$ is 1.91, $q = 11$ is 3.19, $q = 10$ is 7.88, $q = 9$ is 7.18, $q = 8$ is 3.67, $q = 7$ is 1.33, $q = 6$ is 15.95. For the money-

Table 2

Maximum Likelihood Estimates of (31)*

	(1-L)log BR	$(1-L)^2$log GNP	$(1-L)^2$log M2
(1-L)log BR (-1)	.1986 (1.942)	-.0008 (-.071)	-.013 (-1.385)
(-2)	-.051 (-.509)	.0014 (.133)	-.014 (-1.427)
(-3)	-.2080 (-2.216)	-.0096 (-.959)	.212 (2.247)
(-4)	-.069 (-.751)	-.023 (-2.358)	-.168 (-1.750)
(-5)	-.0177 (-.192)	.012 (1.226)	-.131 (-1.368)
(-6)	-.076 (-.857)	-.0326 (-3.433)	.0045 (.476)
$(1-L)^2$log GNP (-1)	1.518 (1.638)	-.746 (-7.176)	
(-2)	1.914 (1.652)	-.657 (-5.053)	
(-3)	2.814 (2.283)	-.363 (-2.625)	
(-4)	2.404 (1.917)	-.464 (-3.291)	
(-5)	2.266 (1.942)	-.205 (-1.562)	
(-6)	-.703 (-.735)	-.197 (-1.834)	
$(1-L)^2$log M2 (-1)			-.425 (-3.842)
(-2)			-.272 (-2.241)
(-3)			-.005 (-.048)
(-4)			-.491 (-4.487)
(-5)			-.248 (-2.115)
(-6)			-.206 (-1.79)
Standard error of the regression	.101	.011	.010

* The number in parenthesis is t-statistic.

Table 3

Full Information Maximum Likelihood Estimates of Model (32) *

	$(1-L)\log BR$	$(1-L)^2\log GNP$	$(1-L)^2\log M2$
$(1-L)\log BR$ (-1)	.1650 (1.692)	-.00003 (-.0029)	
(-2)	-.0516 (-.5167)	.0021 (-.2009)	
(-3)	-.1464 (-1.58)	-.0109 (-1.084)	
(-4)	-.1568 (-1.683)	-.0221 (-2.250)	
(-5)	-.0479 (-.5205)	.0128 (1.2955)	
(-6)	-.0707 (.7984)	-.0329 (-3.468)	
$(1-L)^2\log GNP)$ (-1)		-7.413 (-7.131)	
(-2)		-.6497 (-5.003)	
(-3)		-.3530 (-2.553)	
(-4)		-.4551 (-3.233)	
(-5)		-.1964 (-1.5005)	
(-6)		-.1986 (-1.8517)	
$(1-L)^2\log M2$ (-1)			-.3984 (-3.6566)
(-2)			-.2748 (-2.3625)
(-3)			-.0068 (-.0653)
(-4)			-.4570 (-4.467)
(-5)			-.2040 (-1.843)
(-6)			-.1763 (-1.594)
Standard error of the regression	.108	.0108	.0107

* The numbers in parenthesis are t-statistics.

interest rate, the likelihood ratio statistic for $q = 13$ is 7.29, $q = 12$ is 11.88. For the income-money pair, the likelihood ratio statistic for $q = 13$ is 12.58. From these statistics, it appears that a 6-th order model for the interest rate-income, a 12th order model for the interest rate-money, and a 13th order model for the income-money pair are reasonable choices of unconstrained bivariate AR models.

Given the maintained hypotheses, the bivariate tests for $\sigma^2(y|\overline{Y}, \overline{X}) < \sigma^2(y|\overline{Y})$ and $\sigma^2(x|\overline{Y}, \overline{X}) < \sigma^2(x|\overline{X})$ have chi square values 9.612 and 15.62 with 6 degrees of freedom respectively. The former is significant at 15% level, but not at 10% level and the latter at 1% level. The bivariate tests for $\sigma^2(y|\overline{Y}, \overline{Z}) < \sigma^2(y|\overline{Y})$ and $\sigma^2(z|\overline{Y}, \overline{Z}) < \sigma^2(z|\overline{Z})$ have chi square values 8.92 and 15.10, which are not significant at 15% level with 12 degrees of freedom. The bivariate tests for $\sigma^2(x|\overline{X}, \overline{Z}) < \sigma^2(x|\overline{X})$ and $\sigma^2(z|\overline{X}, \overline{Z}) < \sigma^2(z|\overline{Z})$ have values 11.66 and 18.17 respectively. With 13 degrees of freedom the former is not significant at 15% and the latter is not significant at 10% level.

If we choose 10% significance level, the bivariate results confirm (32) as the appropriate time series model. If we choose 15% significance level, the bivariate results directly contradict (31) and (32). Because according to (32), income should not cause money and according to (31), interest rate should cause money in the bivariate analysis. But the test for $\sigma^2(z|\overline{X}, \overline{Z}) < \sigma^2(z|\overline{Z})$ is significant and the test for $\sigma^2(z|\overline{Y}, \overline{Z}) < \sigma^2(z|\overline{Z})$ is insignificant.

One reason this may happen is that theorems derived in sections 3 and 4 refer to population properties. With finite observations there is no guarantee that the results would be consistent with what one would expect from the population properties of various patterns of causality. Another reason is probably due to the inherent unreliability of the procedure of restricting the maximum lag order of each variable to be identical. There is no particular reason that every variable should enter into the equation with identical lag length. If we do not restrict the lag length to be identical, the above disparity may be reconciled. For instance, a close scrutiny of the interest rate-money pair estimates revealed that the coefficients of the lagged interest rate in the money equation are highly insignificant for orders higher than 5. Testing the model

$$(33)\quad \begin{pmatrix} (1-L)\log BR \\ (1-L)^2\log M2 \end{pmatrix} = \begin{pmatrix} \psi_{11}^{*\,12}(L) & \psi_{13}^{*\,12}(L) \\ \psi_{31}^{*\,5}(L) & \psi_{33}^{*\,12}(L) \end{pmatrix} \begin{pmatrix} (1-L)\log BR \\ (1-L)^2\log M2 \end{pmatrix} + \begin{pmatrix} a^* \\ c^* \end{pmatrix} + \begin{pmatrix} \xi_1^* \\ \xi_3^* \end{pmatrix}$$

against the unconstrained 12th order model, we have chi square value of 3.48 with seven degrees of freedom, indicating the acceptance of (33) as the maintained hypothesis. Taking (33) as the maintained hypothesis, we test $\sigma^2(z|\overline{Z}, \overline{Y}) < \sigma^2(z|\overline{Z})$, which has chi square value of 11.61 with five degrees of freedom, indicating the rejection of $y \not\Rightarrow z$. The test of $\sigma^2(y|\overline{Y}, \overline{Z}) < \sigma^2(y|\overline{Y})$ has chi square value of 5.42 with 12 degrees of freedom, indicating the acceptance of $z \not\Rightarrow y$. The conclusion now is consistent with what one would expect from various patterns of causality.

As we can see from the above analysis, whether the conclusions are consistent with the population properties of various patterns of causality depend critically upon the order of lags chosen and the size of the test. To partially alleviate the problems associated with arbitrary choice of the significance level and the shortage of degrees of freedom when unconstrained vector autoregressive process is used as the maintained hypothesis, we use

Akaike's (1969a, b) final prediction error (FPE) criterion to respecify the model.

The FPE is defined as the (asymptotic) mean square prediction error,

(34) $\text{FPE of } y_t = E(y_t - \hat{y}_t)^2$,

where $\hat{y}_t$ is the predicator of y_t ,

(35) $\hat{y}_t = \hat{\psi}^m_{11}(L)y_t + \hat{\psi}^n_{12}(L)x_t + \hat{\psi}^r_{13}(L)z_t + \hat{a}$.

The superscripts m , n , and r denote the order of lags in $\psi_{11}(L)$, $\psi_{12}(L)$, and $\psi_{13}(L)$. $\hat{\psi}^m_{11}(L)$, $\hat{\psi}^n_{12}(L)$, $\hat{\psi}^r_{13}(L)$, and $\hat{a}$ are least squares estimates obtained when we treat the observations from $-Q + 1$ to 0 as fixed $\{t: t = -Q + 1, \ldots, 0, 1, \ldots, T\}$, $m, n, r \leq Q$. Akaike (1969a) defines the estimate of FPE in this case as

(36) $$\text{FPE}_y(m,n,r) = \frac{T+m+n+r+1}{T-m-n-r-1} \cdot \frac{\sum_{t=1}^{T} (y_t-\hat{y}_t)^2}{T} .$$

The use of FPE criterion fits in nicely with the idea of evaluating the predictability in terms of mean square prediction error. The criterion tries to balance the risk due to the bias when a lower order (than the true order) is selected and the risk due to the increase in variance when a higher order is selected by choosing the specification which gives the smallest FPE. Shibata (1976) has derived the asymptotic distribution and risks of Akaike's statistic. Monte Carlo studies by Geweke and Messe (1979), Quandt and Trussell (1979) have demonstrated its good properties in finite samples. (For additional discussion on the properties of the FPE criterion, see Hsiao (1979b).)

We note that in a multiple AR model, least squares applied to each equation is consistent and asymptotically normally distributed. Therefore, we may ignore the correlations in the innovations for the moment and apply Akaike's (1969) FPE criterion to each equation to determine the order of lags in ψ_{ij} .

If Q is the maximum order of $\psi_{ij}(L)$, then one way to select the order of $\psi_{ij}(L)$ is to let the order of each $\psi_{ij}(L)$ vary between 0 and Q . For a system of p variables, this means there will be $(Q+1)^p$ combinations of $\psi_{ij}(L)$ for the i-th equation. In the case where $p = 3$ and $Q = 13$, we will have to compute 2746 FPE's for each of these three equations. To reduce the computation burden to less than (usually substantially than) $[p + (p-1)](Q+1)$ (in this case 70) we use a squential procedure suggested by Hsiao (1979a) to identify the system.

The smallest FPE's for one, two, and three dimensional analysis with $Q = 13$ are reported in Table 4. After checking the omitted variables' effects as described in Hsiao (1979a)[9], we tentatively choose the following specifications:

$$(37)\quad \begin{bmatrix} (1-L)^2\log GNP \\ (1-L)^2\log M2 \\ (1-L)\log BR \end{bmatrix} = \begin{bmatrix} \psi_{11}^{6}(L) & 0 & \psi_{13}^{6}(L) \\ 0 & \psi_{22}^{12}(L) & \psi_{23}^{4}(L) \\ \psi_{31}^{5}(L) & 0 & \psi_{33}^{3}(L) \end{bmatrix} \begin{bmatrix} (1-L)^2\log GNP \\ (1-L)^2\log M2 \\ (1-L)\log BR \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

To further check the adequacy of specification (37) a sequence of likelihood ratio tests were carried out by deliberately over-fitting and under-fitting (37). The results are reported in Tables 5 and 6.[10] They do not seem to indicate any serious problem with our specification. Hence,

Table 4

The Optimal Lag Order and the FPE's

of the Controlled Variable*

Controlled Variable	First Manipulated Variable	Second Manipulated Variable	FPE x 10^{-4}
BR (3)	-	-	131.4
GNP (6)	-	-	1.701
M2 (12)	-	-	1.333
BR (3)	M2 (1)	-	133.5
BR (3)	GNP (5)	-	131.353
GNP (6)	M2 (1)	-	1.716
GNP (6)	BR (6)	-	1.633
M2 (12)	GNP (2)	-	1.332
M2 (12)	BR (4)	-	1.284
BR (3)	GNP (5)	M2 (1)	132.7
GNP (6)	BR (6)	M2 (2)	1.662
M2 (12)	BR (4)	GNP (2)	1.299

* The number in parenthesis indicates the order of lags of each variable.

Table 5: Likelihood Ratio Tests of (37) Against Higher Order Autoregressive Processes

Maximum Order Fitted for	Model 1	Model 2	Model 3	Model 4
ψ_{11}	6	6	6	6
ψ_{12}	0	6	0	0
ψ_{13}	6	6	6	6
ψ_{21}	2	2	0	4
ψ_{22}	12	12	12	12
ψ_{23}	4	4	4	4
ψ_{31}	5	5	5	5
ψ_{32}	0	0	4	4
ψ_{33}	3	3	3	3
Degrees of Freedom	2	8	4	8
Likelihood Ratio Statistic	3.09	7.23	1.496	8.452

Table 6: Likelihood Ratio Tests of (37) Against Lower Order Autoregressive Processes

Maximum Order Fitted for	Model 1	Model 2	Model 3
ψ_{11}	6	6	6
ψ_{12}	0	0	0
ψ_{13}	0	6	6
ψ_{21}	0	0	0
ψ_{22}	12	12	12
ψ_{23}	4	0	4
ψ_{31}	5	5	0
ψ_{32}	0	0	0
ψ_{33}	3	3	3
Degrees of Freedom	6	4	5
Likelihood Ratio Statistic	15.926*	12.182*	11.91*

*Significant at 5% level

we choose (37) as the time series model and report the full information estimates in Table 7.

Specification (37) indicates that GNP <=> BR , and BR => M2 . According to Theorem 1, we also conclude that equation (37) indicates that M2 neither causes GNP, nor causes BR, and according to Theorem 4, GNP causes M2 indirectly. We then check the validity of these assertions by comparing the FPE's of the bivariate and trivariate analysis. Table 4 indicates that for the bivariate analysis, GNP => M2, GNP => BR, and BR => M2, Yet in the trivariate analysis, GNP no longer appears in the M2 equation. However, the direct causality from GNP to BR, and BR to M2 still hold. This is precisely our definition of indirect causality. On the other hand, the bivariate and trivariate analysis treating either income or the interest rate as the output variable show that using money as an input variable increases the FPE. According to Corollary 2 they indicate that money neither causes income nor causes the interest rate. As one can see these results do confirm our assertion.

Although different procedures may lead to different specifications, at least in this particular case they tend to point out the qualitative conclusion that the bank rate was the primary driving force for income and money, and to a lesser degree responded to income changes. The results are consistent with the Bank of Canada's primary policy concern of trying to maintain "appropriate exchange rates" and "appropriate credit conditions". To the extent to which the Bank primarily aims to regulate the structure of interest rates and not the money supply (e.g. see Courchene (1977) and Rasminsky (1967)), and Canadian interest rates cannot be independent of world (essentially U.S.) rates, movements in the money stock and interest rates

Table 7

Maximum Likelihood Estimates of (37)*

		$(1-L)^2\log GNP$	$(1-L)^2\log M2$	$(1-L)\log r$
$(1-L)^2\log GNP$	(-1)	-.751 (-7.232)		1.748 (1.896)
	(-2)	-.663 (-5.105)		2.433 (2.173)
	(-3)	-.368 (-2.660)		3.351 (2.756)
	(-4)	-.471 (-3.651)		3.020 (2.673)
	(-5)	-.211 (-1.612)		2.890 (3.149)
	(-6)	-.201 (-1.871)		
$(1-L)^2\log M2$	(-1)		-.440 (-4.216)	
	(-2)		-.401 (-3.489)	
	(-3)		-.116 (-.967)	
	(-4)		-.717 (-5.982)	
	(-5)		-.353 (-2.826)	
	(-6)		-.432 (-3.305)	
	(-7)		-.210 (-1.581)	
	(-8)		-.463 (-3.693)	
	(-9)		-.173 (-1.578)	
	(-10)		-.208 (-1.885)	
	(-11)		-.162 (-1.483)	
	(-12)		-.243 (-2.382)	

continued ...

Table 7 continued

(I-L)log BR	(-1)	-.0007 (-.068)	-.009 (-.962)	.207 (2.205)
	(-2)	.001 (.130)	-.008 (-.855)	-.016 (-.162)
	(-3)	-.010 (-.948)	.024 (2.847)	-.221 (-2.638)
	(-4)	-.024 (-2.415)	-.022 (-2.581)	
	(-5)	.012 (1.257)		
	(-6)	-.033 (-3.493)		
Standard Error of the regression		.011	.009	.102

* The number in parenthesis is (large sample) t-statistics.

can be expected to respond more to movements in nominal income.

The fact that nominal income does not respond to changes in M2 might be due to instability in M2 rather than representing a contradiction of the monetarists' position that nominal income is closely related to broad aggregates of financial assets. The term structure of deposit rates has been less stable in Canada than in the U.S. With the existence of a great number of close substitutes issued by near-banks as well as by other financial institutions, structural shifts in intermediation can occur, the effect of which might be to merely alter the charter banks' share but not the total value of these interest-bearing liabilities. A concrete example of this occurred in early 1972. In this sense, M2 might not be a good proxy for the broad aggregates of financial assets in Canada. In fact, the Bank has also maintained that M2 is an unsatisfactory definition of money in Canada (Bank of Canada annual report 1974).

6. Conclusion

In modelling economic time series data there is usually difficulty in discriminating finely among several forms which appear consistent with the information in the data. In fact, in finite sample it often happens that if we do not restrict the parametric specifications, the problem of collinearity, shortage of degrees of freedom, etc. would so confound the interpretation of the model that we do not know what to make of it (Ando (1977), Klein (1977)). Engle (1978), Hendry (1974), Granger and Newbold (1977), Wallis (1974), Zellner and Palm (1974) and many others have suggested approaches to blend traditional econometric and time series analysis to construct better econometric models. In this paper we examine the possibility of extracting information using a pure time series approach. We use

Granger's (1969) notion of causality to derive various propositions about population properties of various causal events. These propositions can be used to provide an interpretation for multivariate time series models as well as a means in developing and checking the empirical implications of various variants of models and in ruling out a number of variants as being inconsistent with the prior theorems. We have applied these concepts to the time series analysis of Canadian money, income, and interest rate data. We have shown that if we are able to use the degrees of freedom more efficiently a pure statistical analysis of economic time series data is capable of yielding useful information.

As Granger (1980) has remarked that the concept of causality and procedures to fit time series model are topics in which individual tastes predominate. It would be improper to try to force research workers to accept definitions and procedures with which they feel uneasy. There is clearly a need for more discussion of this and other definitions of causality and for more explorations of various multiple time series modeling procedures. This paper represents a preliminary attempt to analyze multivariate time series data. The empirical result may or may not stand for further scrutiny. However the topic in my view is of sufficient importance and interest to justify further work in this area rather than brushing aside as ad hoc or not serious.

Footnotes

* This paper is a revision of the paper entitled "Time Series Modelling and Causal Ordering of Canadian Money, Income, and Interest Rate". Research support by the National Science Foundation Grant SOC77-14944-A01 and SES80-07576 at the Institute for Mathematical Studies in the Social Sciences, Stanford University, and by the Social Sciences and Humanities Research Council of Canada Grant 410-78-0031 and 80-0080 is gratefully acknowledged. I am indebted to C.W. Keng and K.Y. Tsui for computational assistance. I also wish to thank T. Amemiya, T.W. Anderson, J.D. Bossons, V. Bencivenga and J. Carr for helpful comments.

1. Note that the theorems in Section 3 may hold under weaker conditions than this (e.g. Hosoya (1977)). We use the covariance stationarity condition for simplicity of exposition because the weaker conditions would be unfamiliar to many readers, and because the FPE formula used below depends on the stationarity condition.
2. In fact, a more natural definition of no causality would be to follow Caines and Chan (1975) by specifying $\Phi(L)$ and $\Psi(L)$ as lower block triangular. However, we have been defining causality as a reduction in forecasting variance with respect to a given information set. To switch the definitions in terms of the forms of AR or MA operator would seem to be inconsistent. In any case, these two ways of defining no causality are equivalent, as stated in Theorem 1, Section 3.
3. Hosoya (1977) has proved the Granger non-causality in nonstationary cases. The result does not seem easily generalizable to the proof here because in non-stationary case the process is no longer a Cauchy sequence

in Hilbert space, thus making it very difficult to operate between different dimensions.

4. We note that if $\phi_{13\ell} = 0$ for all ℓ and $\phi_{23\ell} \neq 0$ for some ℓ, then we may switch the roles of $H_w(t-1)$ and $H_v(t-1)$ and construct an MA representation with $\phi^*_{12\ell} = 0$ for all ℓ under certain conditions.

5. Equivalently, we can assume $\phi_{12\ell} = 0$ for all ℓ and $\phi_{13\ell} \neq 0$, $\phi_{32\ell} \neq 0$ for some ℓ; then condition (6) should be replaced by the equivalent condition of

$$[\phi_{31}(L)\ \phi_{33}(L)] = C(L)[\phi_{11}(L)\ \phi_{13}(L)]$$

6. We are compelled to use seasonally adjusted M2 and income data together with seasonally unadjusted bank rate. For one thing the seasonally unadjusted M2 data are not available. For another there is no seasonal pattern in interest rate. We realize that use of seasonlly adjusted data with seasonally unadjusted data may pose problems (e.g. see Sims (1974), Wallis (1974)). We take comfort in that the empirical analysis performed here is meant to test the feasibility of the methodology, not in providing an answer to the money-income causality debate. (Presumably this is also the reason Sims (1980) analyzed his model using series which were seasonally adjusted together with others which were not.) We may also argue that most macroeconomic aggregates are appraised by economic agents in their adjusted forms regardless of whether or not the seasonal adjustment procedure actually eliminates the seasonal components without distorting the remainder. If this conjecture is correct, the use of seasonally unadjusted data will create bias. Even

if this conjecture is incorrect, there are reasons to believe that the bias due to the use of adjusted data may be minimal. Sims (1974) has demonstrated that if other-than-seasonal components are used and if the model involves unconstrained, long lag distribution which is the case in this paper, then the bias due to the use of adjusted data may be minor. Furthermore, we are using an AR form to test causality rather than using Sims' (1972) method; in this case the bias in our results may be negligible. This can be seen by considering the following simple example. Suppose the true relation of the variable x is of the form

$$\text{(A.1)} \qquad B(L)x_t^a + C(L)y_t^a = u_t \quad ,$$

where x^a , y^a are the seasonally unadjusted data, and u_t are white noises. Suppose the data are adjusted by different filters such that

$$\text{(A.2)} \qquad x_t^s = \eta(L)x_t^a \quad ,$$

$$y_t^s = \gamma(L)y_t^a \quad ,$$

(The special case of either $\gamma(L) \equiv 1$ or $\eta(L) \equiv 1$ means there is no adjustment in the corresponding variable.) Then (A.1) becomes

$$\text{(A.3)} \qquad D(L)x_t^s + F(L)y_t^s = u_t \quad ,$$

where

$$D(L) = \frac{B(L)}{\eta(L)} \quad , \qquad F(L) = \frac{C(L)}{\gamma(L)} \quad ,$$

As one can, if $C(L)$ is identically zero (i.e., y does not cause x), so is $F(L)$. On the other hand, if $C(L)$ is not identically zero, neither is $F(L)$ (i.e., y causes x).

7. One should note that by taking differences the final estimats of a model cannot provide a long run equilibrium analysis since presumably in equilibrium the rate of change is a constant, hence the second difference is zero (also see Sims (1980)).

8. In choosing the appropriate order, we have followed Anderson's (1971) suggestion by selecting a small significance level for high values of q and relatively large significance level for low values of q .

9. We modify the procedure described in step 5' of Hsiao (1979a) by letting ψ_{ij} vary between 0 and Q rather than between 0 and the order previously chosen because depending on the sign of the coefficients and the condition among the variables the omitted effects may operate in either directions.

10. We do not drop those variables which are insignificant (e.g. see McClave (1975, 78)). We take the position that if a higher order coefficient is nonzero, then all the lower order ones will be nonzero too. Otherwise, our model selection procedure would favor the causality relationship. Since the sample variability is such that there is a high probability that a variable will be significant even if it does not cause another variable. Furthermore, dropping one variable would affect all the t-statistics of other coefficients due to the correlation among them.

REFERENCES

Akaike, H. (1969a), "Statistical Predictor Identification", Annals of the Institute of Statistical Mathematics, 21, 203-217.

Akaike, H. (1969b), "Fitting Autoregressions for Prediction", Annals of the Institute of Statistical Mathematics, 21, 243-7.

Anderson, T.W. (1971), The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc.

Ando, A. (1977), "A Comment" in New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, 209-212.

Ansley, C., W.A. Spivey and W.J. Wrobleski (1977), "On the Structure of Moving Average Processes", Journal of Econometrics, 6, 121-134.

Bank of Canada (1974), Annual Report, Ottawa.

Barth, J.R. and J.T. Bennett (1974), "The Role of Money in the Canadian Economy: an Empirical Test", Canadian Journal of Economics, 7, 306-11.

Brunner, K. and A.H. Meltzer (1966), "A Credit Market Theory of the Money Supply and an Explanation of Two Puzzles in U.S. Monetary Policy", Essays in Honor of Macro Fanno, Padova, Italy.

Cagan, P. (1965), Determinant and Effects of Changes in the Stock of Money, 1875-1960, New York: Columbia University Press.

Caines, P.E. (1976), "Weak and Strong Feedback Free Processes", IEEE Trans. Automatic Control, AC-21, 1137-1139.

Caines, P.E. and C.W. Chan (1975), "Feedback between Stationary Stochastic Process", IEEE Trans. Automatic Control, AC-20, 498-508.

Courchene, T.J. (1977), Money, Inflation, and the Bank of Canada: An Analysis of Canadian Monetary Policy from 1970 to Early 1975, C.D. Howe Research Institute.

Engle, R. (1978), "Testing Price Equations for Stability Across Spectral Frequency Bands", Econometrica, 46, 869-881.

Feigl, H. (1953), "Notes on Causality", in H. Feigl and M. Brodbeck (editors), Readings in the Philosophy of Science, New York: Appleton-Century-Crofts, Inc., 408-418.

Friedman, M. (1961), "The Lag in Effect of Monetary Policy", The Journal of Political Economy, 69, 447-466.

Friedman, M. (1970), "A Theoretical Framework for Monetary Analysis", Milton Friedman's Monetary Framework (ed. by R.J. Gordon), Chicago: Chicago University Press, 1-62.

Geweke, J. (1978), "Testing the Exogeneity Specification in the Complete Dynamic Simultaneous Equation Model", Journal of Econometrics, 7, 163-186.

Geweke, J. and R. Messe (1979), "Estimating Distributed Lag of Unknown Order", paper presented at Econometric Society 1979 North America Summer Meeting.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", Econometrica, 37, 424-38.

Granger, C.W.J. (1980), "Testing for Causality: A Personal View", forthcoming in the Journal of Economic Dynamics and Control.

Granger, C.W.J. and P. Newbold (1977), Forecasting Economic Time Series, New York: Academic Press.

Hannan, E.J. (1969), "The Identification of Vector Mixed Autoregressive-Moving Average Systems", Biometrika, 56, 223-225.

Hendry, D.F. (1974), "Stochastic Specification in an Aggregate Demand Model of the United Kingdom", Econometrica, 42, 547-558.

Hosoya, Y. (1977), "On the Granger Condition for Non-Causality", Econometrica, 45, 1735-36.

Hsiao, C. (1979a), "Autoregressive Modelling of Canadian Money and Income Data", Journal of the American Statistical Association, 74, 553-560.

Hsiao, C. (1979b), "Causality Tests in Econometrics", Journal of Economic Dynamics and Control, 1, 321-46.

Klein, L. (1977), "Comments on Sargent and Sims' 'Business Cycle Modelling Without Pretending Too Much a Priori Economic Theory'", New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, 203-208.

McClave, J. (1975), "Subset Autoregression", Technometrics, 17, 213-220.

McClave, J. (1978), "Estimating the Order of Autoregressive Models: The Max χ^2 method", Journal of the American Statistical Association, 73, 122-128.

Masani, P. (1966), "Recent Trends in Multivariate Prediction Theory", Multivariate Analysis I (ed. P.R. Krishnaiah), New York: Academic Press Inc. 351-382.

McElroy, F.W. (1978), "A Simple Method of Causal Ordering", International Economic Review, 19, 1-24.

Pierce, D.A., and L.D. Haugh (1977), "Causality in Temporal Systems: Characterizations and Survey", Journal of Econometrics, 5, 265-293.

Quandt, R.E. and T.J. Trussell (1979), "Some Empirical Evidence on Model Selection Rules", Research memo no. 248, Econometric Research Program, Princeton University.

Rasminsky, L. (1967), "Central Banking in the Canadian Financial System", opening address by the Governor of the Bank of Canada to the 20th International Banking Summer School, Queen's University.

Rozanov, Yu, A. (1967), Stationary Random Process, San Francisco: Holden-Day.

Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criteria", Biometrika, 63, 117-126.

Simon, H.A. (1953), "Causal Ordering and Identifiability", in Studies in Econometric Method, edited by W.C. Hood and T.C. Koopmans, Cowles Commission Monograph 14, New York.

Sims, C.A. (1972), "Money, Income and Causality", American Economic Review, 62, 540-555.

Sims, C.A., (1974), "Seasonality in Regression", Journal of the American Statistical Association, 69, 618-626.

Sims, C.A. (1977), "Exogeneity and Causal Orderings in Macroeconomic Models", in New Methods of Business Cycle Research, Minneapolis Federal Reserve Bank, 23-44.

Sims, C.A. (1980), "Macro-Economics and Reality", Econometrica, 1-48.

Skoog, G. (1976), "Causality Characterizations: Bivariate, Trivariate and Multivariate Propositions", Staff report no. 14, Federal Reserve Bank of Minneapolis.

Tobin, J. (1970), "Money and Income: Post Hoc Ergo Propter Hoc?", Quarterly Journal of Economics, 84, 301-371

Wallis, K.F. (1974), "Seasonal Adjustment and Relations Between Variables", Journal of the American Statistical Association, 69, 18-34

Wallis, K. (1977), "Multiple Time Series Analysis and the Final Form of Econometric Models", Econometrica, 45, 1481-97.

Wold, H. (1960), "A Generalization of Causal Chain Models", Econometrica, 28, 443-463.

Zellner, A. (1977), "Comments on Time Series Analysis and Causal Concepts in Business Cycle Research", in New Methods of Business Cycle Research, Minneapolis Federal Reserve Bank, 167-174.

Zellner, A. (1978), "Causality and Econometrics", Paper presented at a joint Universtiy of Rochester - Carnegie-Mellon Universtiy Conference.

Zellner, A. and F. Palm (1974), "Time Series Analysis and Simultaneous Equation Econometric Models", Journal of Econometrics, 2, 17-54.