Econometric Theory, **16**, 2000, 102–111. Printed in the United States of America.

COINTEGRATION AND DISTANCE BETWEEN INFORMATION SETS

Umberto Triacca
Istituto Nazionale di Statistica

This paper investigates Granger noncausality and the cointegrating relation between two time series in the Hilbert space framework. This framework allows us to analyze the relationship between cointegration and distance between two information sets. In particular, we prove that if two variables, X and Y, are cointegrated, then the distance between two information sets, concerning the differenced series ΔX and ΔY, must be less than the standard deviation of ΔX.

1. INTRODUCTION

A time series, X_t, is said to be integrated of order d (denoted $X_t \sim I(d)$) if it is a series that has a stationary, invertible, nondeterministic ARMA representation after differencing d times. If X_t and Y_t are both $I(d)$ and there exists a scalar α ($\neq 0$) so that $Z_t = X_t - \alpha Y_t \sim I(d-b)$, $b > 0$, then X_t and Y_t are said to be cointegrated of order d,b, denoted $(X_t, Y_t) \sim CI(d,b)$.

Now we suppose that two time series, X_t and Y_t, are both $I(1)$. The main purpose of this paper is the formalization of the following idea. If the information in the past and present of the variable $\Delta Y_t = (1-L)Y_t$, where L is the lag operator defined by $LY_t = Y_{t-1}$, organized in whatever form (i.e., considering any finite and infinite linear combination of the variables $\ldots, \Delta Y_{t-1}, \Delta Y_t, \Delta Y_{t+1}, \ldots$), is "too distant" from the past, present, and future of the variable ΔX_t (i.e., the set $\{\Delta X_t; t = 0, \pm 1, \pm 2, \ldots\}$), then X_t and Y_t are not cointegrated.

The paper is organized as follows. In Section 2 we state Granger's definition of noncausality in Hilbert space framework. In Sections 3 and 4 the main results are presented. Conclusions are given in Section 5.

2. GRANGER'S NONCAUSALITY IN HILBERT SPACE FRAMEWORK

In this section we focus on the causal relationship between two time series $\{X_t, X_{t-1}, \ldots\}$ and $\{Y_t, Y_{t-1}, \ldots\}$. The definition of causality that we consider is that of Granger (1969). The essence of Granger's concept of causality is that Y does not cause X if and only if the (minimum mean square error) linear predictor of X_{t+1} based on $X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots$ is equal to the linear predictor based on $X_t, X_{t-1}, \ldots$ alone.

I am indebted to Marco Lippi, Peter C.B. Phillips, and an anonymous referee for their useful comments on a previous version. Address correspondence to: Umberto Triacca, Istat, via Cesare Balbo 00184, Roma, Italy.

  0266-4666/00 $9.50

Granger's definition of noncausality can be formalized in terms of Hilbert space geometry.[1]

Let $L^2(\Omega, F, P)$ be the Hilbert space of square integrable random variable defined P-almost surely on the probability space (Ω, F, P). In $L^2(\Omega, F, P)$ the inner product and the norm are defined as, respectively, $\langle x, y\rangle = E(xy)$, $\|x\| = (E(x^2))^{1/2}$ $\forall x, y \in L^2(\Omega, F, P)$. If $\|x_n - x\| \to 0$ as $n \to \infty$, we say that $\{x_n\}$ converges in quadratic mean to a limit point x. A point is a limit point of a set M (subset of $L^2(\Omega, F, P)$) if it is a limit point of a sequence from M. In particular, M is said to be closed if it contains all its limit points. If S is an arbitrary subset of $L^2(\Omega, F, P)$, then the set of all $a_1 x_1 + \cdots + a_k x_k$ ($k = 1, 2, \ldots; a_1, \ldots, a_k$ arbitrary real numbers; $x_1, \ldots, x_k$ arbitrary elements of S) is called linear manifold spanned by S and is symbolized by $\mathrm{sp}(S)$. If we add to $\mathrm{sp}(S)$ all its limit points we obtain a closed set that we call the closed linear manifold spanned by S, symbolized by $\overline{\mathrm{sp}}(S)$. Two elements x, y of $L^2(\Omega, F, P)$ are called orthogonal if $\langle x, y\rangle = 0$, and we write $x \perp y$. If $S \subset L^2(\Omega, F, P)$ is any subset of $L^2(\Omega, F, P)$, then we write $x \perp S$ if $x \perp s$ for all $s \in S$; similarly, the notation $S \perp T$, for two subsets S and T of $L^2(\Omega, F, P)$, indicates that all elements of S are orthogonal to all elements of T. For two subsets S and T of $L^2(\Omega, F, P)$ it is well known that if $x \perp T$ $\forall x \in S$, then $\overline{\mathrm{sp}}(S) \perp T$. For a given $x \in L^2(\Omega, F, P)$ and a closed subspace S of $L^2(\Omega, F, P)$, we define the orthogonal projection of $x \in L^2(\Omega, F, P)$ on S, denoted by $(x|S)$, as the element of S such that $\|x - (x|S)\| \leq \|x - z\|$ for any $z \in S$. We observe that if M and N are orthogonal closed subspaces of $L^2(\Omega, F, P)$, then $M + N = \{m + n; m \in M, n \in N\}$ is a closed linear subspace of $L^2(\Omega, F, P)$ and $(x|M + N) = (x|M) + (x|N)$ for any $x \in L^2(\Omega, F, P)$.

We now consider a bivariate discrete stochastic process $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \ldots\}$, defined on (Ω, F, P), with finite second moments. Without loss of generality, we can suppose that $E(X_t) = E(Y_t) = 0$ $\forall t$. We denote by $H_{XY}(t)$, $H_X(t)$, and $H_Y(t)$ the closures with respect to mean square convergence of the linear manifolds generated, respectively, by subsets $\{X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots\}$, $\{X_t, X_{t-1}, \ldots\}$, and $\{Y_t, Y_{t-1}, \ldots\}$ in the Hilbert space $L^2(\Omega, F, P)$, that is,

$$H_{XY}(t) = \overline{\mathrm{sp}}(\{X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots\}),$$

$$H_X(t) = \overline{\mathrm{sp}}(\{X_t, X_{t-1}, \ldots\}),$$

$$H_Y(t) = \overline{\mathrm{sp}}(\{Y_t, Y_{t-1}, \ldots\}).$$

We observe that if $H_X(t)$ and $H_Y(t)$ are orthogonal, $H_X(t) \perp H_Y(t)$, then

$$H_{XY}(t) = H_X(t) + H_Y(t) = \{x + y; x \in H_X(t), y \in H_Y(t)\}.$$

The linear subspace $H_{XY}(t)$ represents the information "linearly organized" contained in the present and past of the process $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \ldots\}$. With respect to this information set we can give the following definition of Granger noncausality.

DEFINITION 1. *The variable Y does not Granger cause the variable X, iff*

$$(X_{t+1}|H_{XY}(t)) = (X_{t+1}|H_X(t)) \quad \forall t.$$

A variable Y Granger causes another variable X, given the information set $H_{XY}(t)$, if X_{t+1} can be better predicted using values Y_{t-s}, $s \geq 0$ than without past and present Y values, that is, Y Granger causes X if the information in $H_Y(t)$ is relevant in explaining future X values.

A similar definition may be found in Bruneau and Nicolaï (1994), Boudjellaba, Dufour, and Roy (1992), and Kohn (1981).

3. CAUSALITY AND DISTANCE

Let $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \ldots\}$ be a stationary process, defined on the probability space (Ω, F, P), with $E(X_t) = E(Y_t) = 0\ \forall t$, and let σ_X^2, σ_Y^2 denote the variance of X_t and Y_t, respectively. We consider furthermore the following subset of $L^2(\Omega, F, P)$:

$$I_X = \{X_t; t = 0, \pm 1, \pm 2, \ldots\}.$$

The quantity

$$d(I_X, H_Y(t)) = \inf\{\|x - y\|; x \in I_X, y \in H_Y(t)\}$$

is said distance between I_X and $H_Y(t)$. We observe that $d(I_X, H_Y(t)) \leq \sigma_X$.

We now can prove the following theorem.

THEOREM 1. *If $d(I_X, H_Y(t)) = \sigma_X$, then Y does not Granger cause X.*

Proof. If $d(I_X, H_Y(t)) = \sigma_X$, then $\|X_\tau - y\| \geq \sigma_X\ \forall y \in H_Y(t), \forall \tau$. We prove first that X_τ is orthogonal to all vectors in $H_Y(t)\ \forall \tau$. If $y = 0$, we have $\langle X_\tau, y \rangle = 0\ \forall \tau$. Now suppose $y \neq 0$. For a given $y \in H_Y(t)$, and for any scalar $\alpha \in R$, we have

$$\begin{aligned}\sigma_X^2 = \|X_\tau\|^2 \leq \|X_\tau - \alpha y\|^2 &= \langle X_\tau - \alpha y, X_\tau - \alpha y \rangle \\ &= \langle X_\tau, X_\tau \rangle - 2\alpha \langle X_\tau, y \rangle + \alpha^2 \langle y, y \rangle \\ &= \|X_\tau\|^2 - 2\alpha \langle X_\tau, y \rangle + \alpha^2 \|y\|^2 \quad \forall \tau.\end{aligned}$$

Therefore $0 \leq \alpha^2 \|y\|^2 - 2\alpha \langle X_\tau, y \rangle\ \forall \tau$. Because $y \neq 0$, we can choose $\alpha = \langle X_\tau, y \rangle / \|y\|^2\ \forall \tau$ so that

$$0 \leq (\langle X_\tau, y \rangle^2 / \|y\|^2) - 2\langle X_\tau, y \rangle^2 / \|y\|^2 = -\langle X_\tau, y \rangle^2 / \|y\|^2 \quad \forall \tau.$$

https://doi.org/10.1017/S0266466600161055 Published online by Cambridge University Press

It follows that $\langle X_\tau, y\rangle = 0\ \forall\tau$. Thus $X_\tau \perp H_Y(t)\ \forall\tau$. Because $H_X = \overline{\text{sp}}\{X_t;\ t = 0,\pm1,\pm2,\dots\}$, this implies $H_X \perp H_Y(t)$. Therefore, we have

$$
\begin{aligned}
(X_{t+1}|H_{XY}(t)) &= (X_{t+1}|H_X(t) + H_Y(t)) \\
&= (X_{t+1}|H_X(t)) + (X_{t+1}|H_Y(t)) \\
&= (X_{t+1}|H_X(t)) \quad \forall t.
\end{aligned}
$$

That is, Y does not Granger cause the variable X. ■

Theorem 1 asserts that when the distance between any linear combination (finite or infinite) of an element of the set $\{Y_t, Y_{t-1},\dots\}$ and any element of the set I_X is equal to σ_X, then the variable Y does not Granger cause the variable X.

We now consider the distance between I_X and $H_Y = \overline{\text{sp}}\{Y_t;\ t = 0,\pm1,\pm2,\dots\}$, that is, the quantity $d(I_X, H_Y) = \inf\{\|x-y\|;\ x\in I_X,\ y\in H_Y\}$.

THEOREM 2. *If $d(I_X, H_Y) = \sigma_X$, then Y does not Granger cause X and X does not Granger cause Y.*

Proof. By using the same arguments we used in proving Theorem 1, we can show that if $d(I_X, H_Y) = \sigma_X$, then $X_t \perp H_Y\ \forall t$. Because $H_X = \overline{\text{sp}}\{X_t;\ t = 0,\pm1,\pm2,\dots\}$, this implies $H_X \perp H_Y$. Therefore, we have

$$
\begin{aligned}
(X_{t+1}|H_{XY}(t)) &= (X_{t+1}|H_X(t) + H_Y(t)) \\
&= (X_{t+1}|H_X(t)) + (X_{t+1}|H_Y(t)) \\
&= (X_{t+1}|H_X(t))) \quad \forall t
\end{aligned}
$$

and

$$
\begin{aligned}
(Y_{t+1}|H_{XY}(t)) &= (Y_{t+1}|H_X(t) + H_Y(t)) \\
&= (Y_{t+1}|H_X(t)) + (Y_{t+1}|H_Y(t)) \\
&= (Y_{t+1}|H_Y(t)) \quad \forall t.
\end{aligned}
$$

That is, Y does not Granger cause the variable X and X does not Granger cause the variable Y. ■

Theorem 2 states that the condition $d(I_X, H_Y) = \sigma_X$ is sufficient to exclude any Granger causal link between X and Y. This could appear surprising; indeed, we could expect that such a condition would be sufficient only to exclude causality from Y to X. We would also expect to need two conditions—namely, $d(I_X, H_Y) = \sigma_X$ and $d(I_Y, H_X) = \sigma_Y$, where $I_Y = \{Y_t;\ t = 0,\pm1,\pm2,\dots\}$—to obtain the thesis of Theorem 2. But this would be redundant because the following theorem applies.

THEOREM 3. *$d(I_X, H_Y) = \sigma_X$, iff $d(I_Y, H_X) = \sigma_Y$.*

https://doi.org/10.1017/S0266466600161055 Published online by Cambridge University Press

Proof. If $d(I_X, H_Y) = \sigma_X$, then $H_X \perp H_Y$; thus, by the Pythagorean theorem, we have

$$\|y - x\|^2 = \|y\|^2 + \|x\|^2 \quad \forall x \in H_X, \qquad \forall y \in H_Y.$$

In particular, we have $\|Y_t - x\|^2 = \|Y_t\|^2 + \|x\|^2 \; \forall x \in H_X, \forall t$, and this implies that $\|Y_t - x\| \geq \|Y_t\| = \sigma_Y \; \forall x \in H_X, \forall t$, that is, $d(I_Y, H_X) = \sigma_Y$. In a similar way we can prove that $d(I_Y, H_X) = \sigma_Y$ implies $d(I_X, H_Y) = \sigma_X$. ■

In this section we have proved the following statements.

1. If the distance between the set $I_X = \{X_t; t = 0, \pm 1, \pm 2, \dots\}$ and the linear space $H_Y(t)$ reaches the limit value σ_X, then Y does not Granger cause X;
2. if this same value is reached by the distance between the set $I_X = \{X_t;\ t = 0, \pm 1, \pm 2, \dots\}$ and the linear space H_Y, then Y does not Granger cause X and X does not Granger cause Y.

In other terms, we have that, if Y Granger causes X, then $d(I_X, H_Y(t)) < \sigma_X$, and if Y Granger causes X or X Granger causes Y, then $d(I_X, H_Y) < \sigma_X$. Therefore, the condition $d(I_X, H_Y) < \sigma_X$ can be interpreted as a condition of contiguity: the existence of a Granger causal link between X and Y indicates that processes $\{X_t\}$ and $\{Y_t\}$ are contiguous, that is, $d(I_X, H_Y) < \sigma_X$.

Moreover, we note that, if $\sigma_X = 0$, then the variable X is causally prior with respect to any variable, and this is consistent with the fact that the Granger's approach to causality excludes all nonstochastic variables.

Finally we observe that we have not at all considered the distance between the sets $I_X = \{X_t; t = 0, \pm 1, \pm 2, \dots\}$ and $I_Y = \{Y_t; t = 0, \pm 1, \pm 2, \dots\}$, because even if this distance were greater than or equal to σ_X, the information contained in I_Y could be organized so as to be useful for the prediction of the future of X. Let us consider, in fact, the process $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \dots\}$ defined as

$$X_t = \alpha Y_{t-1} + \varepsilon_{x_t} \qquad \alpha < 0$$

$$Y_t = \varepsilon_{y_t}$$

where ε_{x_t} and ε_{y_t} are independent white noise, with $E(\varepsilon_{x_t}^2) = E(\varepsilon_{y_t}^2) = 1$.[2] In this case Y Granger causes X; further, because $d(I_X, I_Y) = \sqrt{\alpha^2 + 2}$ and $\sigma_X = \sqrt{\alpha^2 + 1}$, we have $d(I_X, I_Y) \geq \sigma_X$. Thus the condition $d(I_X, I_Y) \geq \sigma_X$ does not imply the noncausality between X and Y.

4. COINTEGRATION AND DISTANCE

We now abandon the assumption that $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \dots\}$ is a stationary process. Suppose that X_t and Y_t are both $I(1)$ and that they are generated by a vector autoregressive process of order p,

$$A(L)\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix} \tag{1}$$

where

$$A(L) = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix}$$
$$= I - A_1 L - \cdots - A_p L^p$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} L - \cdots - \begin{bmatrix} a_{11,p} & a_{12,p} \\ a_{21,p} & a_{22,p} \end{bmatrix} L^p$$

and $(e_{xt}, e_{yt})'$ is a bivariate white noise process with mean zero and nonsingular covariance matrix.

In the process (1) Y does not Granger cause X iff $a_{12,i} = 0$ for $i = 1, \ldots, p$. Conversely X does not Granger cause Y iff $a_{21,i} = 0$ for $i = 1, \ldots, p$.

We observe that using the first-difference operator Δ, (1) can be reparameterized as an error correction model:

$$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = \sum_{i=1}^{p-1} \Gamma_i \begin{bmatrix} \Delta X_{t-i} \\ \Delta Y_{t-i} \end{bmatrix} + \Pi \begin{bmatrix} X_{t-p} \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix},$$

where

$$\Gamma_i = -\left(I - \sum_{j=1}^{i} A_j \right) \quad \text{for } i = 1, \ldots, p-1 \quad \text{and} \quad \Pi = -\left(I - \sum_{j=1}^{p} A_j \right).$$

The cointegration literature focuses on the parameter matrix Π because, assuming that all roots of the implicit VAR polynomial $(\det[A(z)])$ lie outside the unit circle or equal to 1, the nonstationarity character of the analyzed series, $(X_t, Y_t)'$, is determined by Π. If Π has full rank all variables in the system are stationary. On the contrary, when $\Pi = \mathbf{0}$, the variables are not cointegrated, and the process may be represented as a VAR in first differences. If Π is not zero and has rank 1, the variables are cointegrated.

A well known result is that cointegration implies Granger causality in at least one direction (Granger, 1988); if $(X_t, Y_t) \sim CI(1,1)$, then Y Granger causes X or X Granger causes Y.

Now, because $(\Delta X_t, \Delta Y_t)'$ is supposed to be a zero mean purely nondeterministic stationary process, there exists a multivariate Wold representation, as follows:

$$(1 - L) \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \Psi(L) \begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix}, \tag{2}$$

where

$$\Psi(L) = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix}$$

https://doi.org/10.1017/S0266466600161055 Published online by Cambridge University Press

with $\Psi_{ij}(L) = \Psi_{ij,0} + \Psi_{ij,1}L + \Psi_{ij,2}L^2 + \cdots$ and $\Psi_{11,0} = \Psi_{22.0} = 1$, $\Psi_{12,0} = \Psi_{21.0} = 0$. We remember that ΔY does not Granger cause ΔX iff $\Psi_{12,i} = 0$ for $i = 1, 2, \ldots$. Conversely ΔX does not Granger cause ΔY iff $\Psi_{21,i} = 0$ for $i = 1, 2, \ldots$. Thus we can prove the following result.

THEOREM 4. *If ΔY does not Granger cause ΔX and ΔX does not Granger cause ΔY, then Y does not Granger cause X and X does not Granger cause Y.*

Proof. If ΔY does not Granger cause ΔX and ΔX does not Granger cause ΔY, then $\Psi_{12}(L) = \Psi_{21}(L) = 0$. Premultiplying (2) by $A(L)$ results in

$$(1-L)A(L)\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = A(L)\Psi(L)\begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix}. \tag{3}$$

Substituting (1) in (3), we have

$$(1-L)\begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix} = A(L)\Psi(L)\begin{bmatrix} e_{xt} \\ e_{yt} \end{bmatrix}. \tag{4}$$

Now, equation (4) has to hold for all realization of $(e_{xt}, e_{yt})'$, which requires that $(1-L)I$ and $A(L)\Psi(L)$ represent the identical polynomial in L. This means that

$$(1-z)I = A(z)\Psi(z) \tag{5}$$

for all values of z. Thus, because by hypothesis ΔY does not Granger cause ΔX and ΔX does not Granger cause ΔY, we have that

$$\begin{bmatrix} 1-L & 0 \\ 0 & 1-L \end{bmatrix} = \begin{bmatrix} A_{11}(L)\Psi_{11}(L) & A_{12}(L)\Psi_{22}(L) \\ A_{21}(L)\Psi_{11}(L) & A_{22}(L)\Psi_{22}(L) \end{bmatrix}$$

and $A_{12}(L)\Psi_{22}(L) = A_{21}(L)\Psi_{11}(L) = 0$ implies that $A_{12}(L) = A_{21}(L) = 0$. ■

To this point we can finally establish a result that places in relation the distance between the differentiated processes, ΔX_t and ΔY_t, and the condition of cointegration between the processes X_t and Y_t.

We remember first of all that $\{(\Delta X_t, \Delta Y_t)', t = 0, \pm1, \pm2, \ldots\}$ is a stationary process, defined on the probability space (Ω, F, P), with $E(\Delta X_t) = E(\Delta Y_t) = 0\ \forall t$; let $\sigma^2_{\Delta X}$ denote the variance of ΔX_t. We consider the following subsets of $L^2(\Omega, F, P)$:

$$I_{\Delta X} = \{\Delta X_t;\ t = 0, \pm1, \pm2, \ldots\}$$

and

$$H_{\Delta Y} = \overline{\mathrm{sp}}\{\Delta Y_t;\ t = 0, \pm1, \pm2, \ldots\}.$$

The distance between $I_{\Delta X}$ and $H_{\Delta Y}$ is the quantity

$$d(I_{\Delta X}, H_{\Delta Y}) = \inf\{\|x - y\|;\ x \in I_{\Delta X},\ y \in H_{\Delta Y}\}.$$

Now, we can prove the following theorem.

THEOREM 5. *If $d(I_{\Delta X}, H_{\Delta Y}) = \sigma_{\Delta X}$, then X_t and Y_t are not cointegrated.*

Proof. If $d(I_{\Delta X}, H_{\Delta Y}) = \sigma_{\Delta X}$, then, by Theorem 2, ΔY does not Granger cause ΔX and ΔX does not Granger cause ΔY. It follows, by Theorem 4, that Y does not Granger cause X and X does not Granger cause Y. Thus, because cointegration implies Granger causality in at least one direction, we can conclude that X_t and Y_t are not cointegrated. ■

Theorem 5 asserts that if X and Y are cointegrated, then the distance between at least one linear combination (finite or infinite) of elements of the set $\{\dots, \Delta Y_{t-1}, \Delta Y_t, \Delta Y_{t+1}, \dots\}$ and at least one element of the set $I_{\Delta X}$ must be less than the standard deviation of ΔX.

5. CONCLUSION AND INTERPRETATION OF THE RESULTS

In this paper we have investigated Granger noncausality and the cointegrating relation between two nonstationary time series in the Hilbert space framework.

We have shown that when the information in the past and present of the variable Y, organized in whatever form (i.e., considering any finite and infinite linear combination of the variables $Y_t, Y_{t-1}, \dots$), is "too distant" from the past, present, and future of the variable X (i.e., the set $\{X_t;\ t = 0, \pm 1, \pm 2, \dots\}$), then Y does not Granger cause the variable X. It is interesting to see how this result is related to noncausality in the usual sense. The notion of Granger causality stipulates that a variable Y causes another variable X if the past and present values of Y can be used to predict X more accurately than simply using the past and present values of X. Now, if $d(I_X, H_Y(t)) = \sigma_X$, then $d(I_X(t+), H_Y(t)) = \sigma_X$, where $I_X(t+) = \{X_{t+1}, X_{t+2}, \dots\}$. This means that the information in the past and present of the variable Y, organized in whatever form (i.e., the set $H_Y(t)$), is "too distant" from the information in $I_X(t+)$ and hence the information in $H_Y(t)$ is not relevant in explaining future X values (i.e., Y does not Granger cause X).

Theorem 2 asserts that in a bivariate framework the existence of a Granger causal link between X and Y requests that the processes $\{X_t\}$ and $\{Y_t\}$ are contiguous, that is, $d(I_X, H_Y) < \sigma_X$. It is worth emphasizing that, in a multivariate framework, Granger causality does not require that every cause is contiguous with its effects; however, in a bivariate framework, it points out only direct causal links between contiguous processes. This is the problem of noncausality due to omitted variables (Lutkepohl, 1982).

We also have proved that if X_t and Y_t are both $I(1)$ and if the information in the past and present of the variable $\Delta Y_t = (1 - L)Y_t$, organized in whatever form (i.e., considering any finite and infinite linear combination of the variables $\dots, \Delta Y_{t-1}, \Delta Y_t, \Delta Y_{t+1}, \dots$) is "too distant" from the past, present, and future of the variable ΔX_t (i.e., the set $\{\Delta X_t;\ t = 0, \pm 1, \pm 2, \dots\}$), then X_t and Y_t are not cointegrated. This is a surprising result: according to it a condition concerning the differenced series implies a condition (noncointegration) concerning the linear combinations of the series (in levels). The intuition behind it is the following. Cointegration is the statistical implication of the existence of a long-run relationship between the economic variables. When two variables, X and Y, are cointegrated there is a tendency for the variables to move together to maintain a log-run equilibrium; therefore a cointegration relationship cannot exist between X and Y if the information in the past, present, and future of ΔY, organized in whatever form, is "too distant" from the past, present, and future of ΔX.

In general differencing the data causes loss of information about the relationship between the levels of the variables; however, according to Theorem 5, the cointegration property can be detected from differenced variables. This seems to mean that we need not information of nominal variables but only that of the differenced variables to see the variables are not cointegrated.

Finally, we note that Theorem 5 connects to earlier work on efficient estimation of cointegrating regressions by way of lead and lag formulations (see Phillips and Loretan, 1991; Saikkonen, 1991; Stock and Watson, 1993).

NOTES

1. Details on Hilbert spaces can be found in Halmos (1951), Brockwell and Davis (1987), and Caines (1987).

2. We remember that for the process $\{(X_t, Y_t)', t = 0, \pm 1, \pm 2, \ldots\}$ defined as

$$X_t = \alpha Y_{t-1} + \varepsilon_{x_t},$$

$$Y_t = \varepsilon_{y_t},$$

where ε_{x_t} and ε_{y_t} are independent white noise, with $E(\varepsilon_{x_t}^2) = E(\varepsilon_{y_t}^2) = 1$, we have

$$d(I_X, I_Y) = \begin{cases} \sqrt{\alpha^2 + 2} & \text{if } \alpha < 0 \\ \sqrt{2} & \text{if } \alpha = 0 \\ \sqrt{\alpha^2 - 2\alpha + 2} & \text{if } \alpha > 0 \end{cases}.$$

REFERENCES

Boudjellaba, H., J.M. Dufour, & R. Roy (1992) Testing causality two vectors in multivariate autoregressive moving average models. *Journal of the American Statistical Association* 87, 1082–1090.

Brockwell, P.J. & R.A. Davis (1987) *Time Series: Theory and Methods.* New York: Springer.

Bruneau, C. & J.-P. Nicolaï (1994) Amèlioration de la prèvision et causalitè entre deux sèries d'un système multivarié autorègressif stationaire. *Annals d'Economie et de Statistique* 36, 1–22.

Caines, P.E. (1987) *Linear Stochastic System.* New York: Wiley.

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

Granger, C.W.J. (1988) Some recent development in a concept of causality. *Journal of Econometrics* 39, 199–211.

Halmos, P.R. (1951) *Introduction to Hilbert Space and the Theory of Spectral Multiplicity.* New York: Chelsea Publishing Company.

Kohn, R. (1981) A characterization of Granger-Sims exogeneity. *Economics Letters* 8, 129–133.
Lutkepohl, H. (1982) Non-causality due to omitted variables. *Journal of Econometrics* 19, 367–378.
Phillips, P.C.B. & M. Loretan (1991) Estimating long run economic equilibria. *Review of Economic Studies* 58, 407–436.
Saikkonen, P. (1991) Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.