TO: REQUESTORS OF BAEK AND BROCK'S "A GENERAL TEST FOR NONLINEAR GRANGER CAUSALITY: BIVARIATE MODEL"

FROM: W.A. BROCK

I apologize for the coffee stained (with various random handwritings in the margins) condition of the typescript for this paper. It was in the process of being revised for a journal when both authors got caught up in other lines of work after their joint piece

Baek, E., and Brock, W., (1992), "A Nonparametric Test for Independence of a Multivariate Time Series," STATISTICA SINICA, 2,137-156,

was published.

Furthermore the excellent papers

Hiemstra, C., and Jones, J., (1994), "Testing for Linear and Nonlinear Granger Causality in the Stock-Volume Relation," JOURNAL OF FINANCE, 49, 1639-1664.

Abhyankar, A. (1994), "Linear and Nonlinear Granger Causality: Evidence from the FT-SE 100 Index Futures and Cash Markets," Department of Accountancy and Finance, University of Stirling, SCOTLAND.,

have gone beyond our work in this area.

Furthermore Dr. R. Antoniewicz,

Antoniewicz, R., (1992), "A Causal Relationship Between Stock Prices and Volume," Division of Research and Statistics, Board of Governors, U.S. Federal Reserve System, Washington, D.C.

has developed another approach to nonlinear causality testing based upon the bootstrap-based specification test of Brock, Lakonishok, LeBaron, J. FINANCE, Dec., 1992.

We suggest contacting Professors A. Abhyankar, C. Hiemstra (at University of Strathclyde, Glasgow, SCOTLAND) and Dr. R. Antoniewicz for more work in this area.

A General Test for Nonlinear Granger Causality: Bivariate Model

by

Ehung Gi Baek
Korea Development Institute
P.O. Box 113, Cheong-Ryang, Seoul 130-012, KOREA
and
Department of Economics
Iowa State University
Ames, IA 50011-1070

William A. Brock
Department of Economics
University of Wisconsin-Madison
Madison, Wisconsin 53706

May, 1991, Revised January, 1992

Abstract: In this paper we develop a general test for nonlinear Granger causality in a bivariate context. The test is based upon a measure of local correlation called the correlation integral. A small Monte Carlo study is conducted to evaluate the performance of the test. The test is applied to money and income. Some evidence of nonlinear causality was found after the size of the test was corrected.

We would like to thank Craig Hiemstra, Bruce Mizrach, James Ramsey, referees of this Journal, ~~for comments~~ and other conference participants at the UCLA conferences on nonlinear dynamics and econometrics, for helpful comments on this paper. In addition W.A. Brock would like to express thanks to the NSF Grant #, the Wisconsin Graduate School & Wisconsin Alumni Research Foundation, and the Vilas Trust for essential financial support. None of the above are responsible for errors.

1. INTRODUCTION

The purpose of this study is to develop a statistical tool to detect unidirectional causal orderings between two, possibly nonlinearly related, variables. "Causality" here is used in the sense of Granger as exposited by Geweke (1984, p. 1102). We apply our version of a nonlinear Granger causality test to money and income using the data of Stock and Watson (1989). Before we explain the new test proposed in this paper we must give enough background on linear causality tests to place the new test into context.

We use notation as follows: $\epsilon>0$, $\mathbb{R}^k$ denotes the vector space of k-dimensional real vectors, $\{Z_t\}$ denotes a stochastic process, i.e., a sequence of random variables from a common probability triple to the real line, $\mathbb{R}^1\equiv\mathbb{R}$. Here $t=1,2,\ldots$. Bold letters will denote vectors (possibly infinite dimensional) of real numbers or random variables.

Let $\{X_t\}$, $\{Y_t\}$ be two scalar valued strictly stationary, ergodic stochastic processes. Let $\mathbf{Z}_t\equiv\{Z_t,Z_{t-1},\ldots\}$ denote the history to time t for any stochastic process $\{Z_t\}$. All we need to expound our contribution to causality testing is to give an heuristic version of the original Granger (1969) definition (in the linear context):

<u>Definition</u>: $\{Y_t\}$ causes $\{X_t\}$ if $\mathbf{Y}_t$ helps linearly predict X_{t+1} in addition to $\mathbf{X}_t$. This idea is operationalized by regressing X_{t+1} on both $\mathbf{X}_t,\mathbf{Y}_t$ and testing whether the coefficients of $\mathbf{Y}_t$ are all zero. □

A host of issues such as (i) how to select the lag length, (ii) how to select the background information set $\{U_t\}$ against which the incremental predictive content of Y_t for X_{t+1} is to be measured, (iii) which theoretically equivalent form of the linear Granger test should one use, i.e., which of these have the best size and power performance, (iv) how does one deal with problems in implementation caused by different methods of detrending. We refer the reader to Granger (1969), Geweke, Meese, and Dent (1983), Sims (1972), Stock and Watson (1989), Sims, Stock, and Watson (1990), for discussion of these issues. Dealing with such issues in our nonlinear context is beyond the scope of this paper. Turn now to explanation of our nonlinear causality test. We need a pedagogical example.

Consider the following pair of processes,

$$X_t = \beta Y_{t-q} X_{t-p} + \epsilon_t, \tag{1}$$

where $\{\epsilon_t\},\{Y_t\}$ are mutually independent and each are Independently and Identically Distributed (IID) Gaussian with zero mean and unit variance. We shall refer to a pair of stochastic processes, such as $\{\epsilon_t\},\{Y_t\}$, which are each IID but are also mutually independent at all leads and lags as "IIDI." The point of example (1) is this. If one linearly regresses X_{t+1} on X_t,Y_t one will find zero coefficients on both X_t,Y_t. Yet it is obvious that Y_t incrementally helps predict X_{t+1} given X_t. It is easy to detect this incremental nonlinear predictability by appropriate nonlinear regression. Rather than nonlinear regression we take a general approach here based upon something called the "correlation integral" which is a measure of local spatial correlation of temporal series which are "embedded" in an appropriate "embedding" space. The next section exposits this idea.

The paper is organized as follows. Section one contains the introduction. The second section defines the correlation integral and gives some of its properties. Section three reports a small Monte Carlo study on the size performance of our test. Section four applies the test to money and income data. Finally, in Section five, we summarize and conclude.

2. CORRELATION INTEGRAL

Let $\{Z_t\}$ denote an $\mathbb{R}^k$-valued stochastic process. Put k=1, for the moment, and define $W_t \equiv Z_{t,m} \equiv (Z_{t-1},\ldots,Z_{t-m})$ and call W_t an "m-history of Z starting at t." Put, $\epsilon>0$,

$$C_m(\epsilon,T) \equiv (1/T^*) \Sigma\Sigma\, I_\epsilon(W_t,W_s), \tag{2}$$

where $I_\epsilon(a,b) \equiv \text{Max}\{|a_i-b_i|<\epsilon,\ i=1,2,\ldots,m\}$, $a,b \in \mathbb{R}^m$. Here the double sum, $\Sigma\Sigma$ is over $\{(s,t)|1\leq s<t\leq T\}$, $T^* \equiv T(T-1)/2$ for the U-statistic form; $T^* \equiv T^2$, and $\Sigma\Sigma$ is over $\{(s,t)|1\leq s,t\leq T\}$ for the V-statistic form. Intuitively C_m measures the fraction of pairs of m-histories whose distances are less than ϵ. For the case when k=1, i.e., Z_t takes values in the real line, $\mathbb{R}$, we would expect C_m to scale with the power m as ϵ increases if $\{Z_t\}$ is IID. If $\{Z_t\}$ is generated by a one dimensional deterministic chaos (such as $Z_t=4Z_{t-1}(1-Z_{t-1})$) then one would expect C_m to scale with the power one as ϵ increases. In general the elasticity of C_m with respect to ϵ is a measure of the number of active degrees of freedom at ϵ of $\{Z_t\}$. See Brock and Baek (1991) or Brock, Hsieh, and LeBaron (1991) for a general discussion of

the properties of C_m and references to the literature.

The theory of U and V statistics is expounded by Denker and Keller (1983). Intuitively U and V statistics are statistics that behave much like averages. Denker and Keller (1983) show, under conditions of weak dependence, stationarity, and ergodicity the following, for both the U-statistic and V-statistic form of C_m:

$$C_m(\epsilon,T) \longrightarrow C_m(\epsilon),$$
$$T^{1/2}[C_m(\epsilon,T) - C_m(\epsilon)] \xrightarrow{d} N(0,V_m(\epsilon)) \text{ as } T\to\infty \qquad (3)$$

where convergence is in distribution. Here N(0,V) denotes the normal distribution with mean 0 and variance, V.

<u>Maintained Assumptions</u>: We shall assume weak dependence, strict stationarity, and ergodicity of all data generating processes as in Denker and Keller (1983, p. 507, for stochastic processes; 1986, p. 76, for deterministic, possibly chaotic, processes). We call this maintained "regularity" assumption "(DK)" for reference purposes. Denker and Keller (1983) gives three versions of weak dependence which are strong enough to deliver their results. The reader is free to use whichever of the three she wishes. Only the U-statistic form is treated here. The V-statistic form is similar. □

The quantity, $C_m(\epsilon)$, that appears in (3) is easy to interpret. It is given by,

$$C_m(\epsilon) = \Pr\{|W_t - W_s| < \epsilon\} \equiv \iint I_\epsilon(w_t,w_s)dF(w_t)dF(w_s), \qquad (4)$$
$$\equiv \iint I_\epsilon(w,w')dF(w)dF(w')$$

where $|.|$ denotes the max norm on R^m, "Pr" denotes "probability," $F(w) \equiv Pr\{W_t \equiv W \leq w\}$. Note that the probability is computed as a double integral and that $F(w)$ is the cumulative distribution of an m-dimensional vector for embedding in dimension m. An obvious generalization would replace ϵ by an m-vector, e, which may be useful when units vary across the components of W.

Return now to the pair of stochastic processes, $\{X_t\}$, $\{Y_t\}$, recall the definition of an m-history, $Z_{t,m} \equiv (Z_{t-1},\ldots,Z_{t-m})$, and consider the following probability

$$Pr\{|X_t - X_s| < \epsilon_1 \text{ given } |X_{t,p} - X_{s,p}| < \epsilon_1,\ |Y_{t,q} - Y_{s,q}| < \epsilon_2\}. \tag{5}$$

Think of this probability as measuring the probability that $|X_t - X_s| < \epsilon_1$, given the pair of p-histories of X differ by less than ϵ_1 and the pair of q-histories of Y differ by less than ϵ_2, i.e., $|X_{t,p} - X_{s,p}| < \epsilon_1$, and, $|Y_{t,q} - Y_{s,q}| < \epsilon_2$. Note that (5) is computed as a ratio in (7) below where the entries in (7) are computed as in (4). This motivates the following definition.

Definition: Y fails to nonlinearly Granger cause X if

$$Pr\{|X_t - X_s| < \epsilon_1 \text{ given } |X_{t,p} - X_{s,p}| < \epsilon_1,\ |Y_{t,q} - Y_{s,q}| < \epsilon_2\}$$
$$= Pr\{|X_t - X_s| < \epsilon_1 \text{ given } |X_{t,p} - X_{s,p}| < \epsilon_1\}. \ \square \tag{6}$$

Note that the definition says, given ϵ_1 and ϵ_2, q lags of Y does not incrementally help predict next period's value of X given p lags of X. If $X_t = G(X_{t,p}, Y_{t,q})$ for some deterministic continuous function G then it is easy to see why the event, "$Y_{t,q}$ close to $Y_{s,q}$", would help nonlinearly incrementally predict "X_t close to X_s". The definition is motivated by a hope that at least part of the deterministic intuition may pass to the case where G is a stochastic function, especially if the conditional variance of G given $X_{t,p}, Y_{t,q}$ is small enough relative to the unconditional variance of X_t. Of course this definition depends on the number of lags of future prediction which is one here; the ϵ_1, ϵ_2; the lags in the histories of X and Y i.e., p and q. Turn now to operationalization of testing the null hypothesis H_o:Y fails to nonlinearly Granger cause X.

First rewrite each side of (6) as a ratio of unconditional probabilities. Second, for a sample of length T, estimate each numerator and denominator with correlation integrals. Third, take the difference of the ratio of statistics of step two, multiply it by $T^{1/2}$ and use the delta method (Serfling (1980, p. 124) together with the Denker-Keller (1983, 1986) theory to work out the limit distribution which is normal with mean zero and variance V under the null hypothesis H_o. Fourth, replace the variance V with a consistent estimator $\hat{V}$, divide by the square root of $\hat{V}$ to get a statistic which is asymptotically distributed N(0,1) under the null, H_o.

Carrying out the first step we obtain,

$$\Pr\{|X_t-X_s|<\epsilon_1 \text{ given } |X_{t,p}-X_{s,p}|<\epsilon_1,\ |Y_{t,q}-Y_{s,q}|<\epsilon_2\}$$

$$= \frac{\Pr\{|X_t-X_s|<\epsilon_1,\ |X_{t,p}-X_{s,p}|<\epsilon_1,\ |Y_{t,q}-Y_{s,q}|<\epsilon_2\}}{\Pr\{|X_{t,p}-X_{s,p}|<\epsilon_1,\ |Y_{t,q}-Y_{s,q}|<\epsilon_2\}}$$

$$\equiv C(x',x,y;e)/C(x,y;e), \qquad (7)$$

where we put $x' \equiv |X_t - X_s| < \epsilon_1$, $x \equiv |X_{t,p} - X_{s,p}| < \epsilon_1$, $y \equiv |Y_{t,q} - Y_{s,q}| < \epsilon_2$, $e \equiv (\epsilon_1, \epsilon_2)$ to lighten notation. Restating H_o in this new notation we obtain,

$$H_o: \ C(x',x,y;e)/C(x,y;e) - C(x',x;e)/C(x;e) = 0. \tag{8}$$

Now replace each "C" in (8) by a correlation integral estimator, $C(.;e,T)$, where

$$C(x',x,y;e,T) \equiv \Sigma\Sigma \ I\{|X_t - X_s| < \epsilon_1, |X_{t,p} - X_{s,p}| < \epsilon_1, |Y_{t,q} - Y_{s,q}| < \epsilon_2\}/T^*, \tag{9}$$

where $I\{A\}$ denotes the indicator function of event A which is one when A is true and is zero otherwise. Estimators for the other probabilities in (8) are defined analogously.

<u>Proposition 2.1</u>: Assume H_o and DK. Then,

$$T^{1/2}[C(x',x,y;e,T)/C(x,y;e,T) - C(x',x;e,T)/C(x;e,T)] \xrightarrow{d} N(0,V) \text{ as } T \to \infty \tag{10}$$

where convergence is in distribution.

Proof: Put

$$G[C(x',x,y;e,T),C(x;e,T),C(x',x;e,T),C(x,y;e,T)] \equiv G[U_1,U_2,U_3,U_4] \equiv C(x',x,y;e,T)/C(x,y;e,T) - C(x',x;e,T)/C(x;e,T). \tag{11}$$

We have a function G[.] of four U statistics. Let θ_i denote the mean of U_i. It follows from Denker and Keller (1986, Theorem 1, p. 75) that, U_i converges in probability to θ_i for each i. Under H_o, $G[\theta_1,\theta_2,\theta_3,\theta_4] \equiv G[\Theta]=0$. Put $U \equiv (U_1,U_2,U_3,U_4)$. Therefore, applying the delta method (Serfling (1980, p. 124)),

$$T^{1/2}\{G[U] - G[\Theta]\} \tag{12}$$

has the same asymptotic distribution as

$$T^{1/2}\nabla G(\Theta)\cdot(U - \Theta), \tag{13}$$

where $\nabla G(\Theta)$ denotes the gradient of G evaluated at Θ and "·" denotes dot product. By Denker and Keller (1983, p. 507, Equation (1)) each U-statistic has the representation

$$U_i - \theta_i = 2/T \,\Sigma\, (h_{1i}(.) - \theta_i) + R_i,\ T^{1/2}R_i \to 0 \text{ as } T \to \infty, \tag{14}$$

where convergence is in distribution. Here $h_{1i}(.)$ is the conditional expectation of the summand (called the "kernel" of the U-statistic) of the U-statistic on the subscript t argument. For example, in our case of,

$$G(U) \equiv U_1/U_4 - U_3/U_2, \tag{15}$$

$$h_{11}(.) \equiv h_{11}(X_t,X_{t,p},Y_{t,q}) \equiv h_{11}(Z_t)$$

$$\equiv E[I\{|X_t - X_s| < \epsilon_1, |X_{t,p} - X_{s,p}| < \epsilon_1, |Y_{t,q} - Y_{s,q}| < \epsilon_2\};\ \text{given}\ X_t, X_{t,p}, Y_{t,q}]. \tag{16}$$

The other h_{1i} are defined analogously. Let G_i denote $\partial G/U_i$ evaluated at Θ. Since $T^{1/2}R_i \to 0$ as $T \to \infty$, use (14) to rewrite an asymptotically equivalent form of (13) as follows,

$$(2/T^{1/2}) \sum_t \{ \sum_i G_i(h_{1i}(Z_t) - \theta_i)\} \xrightarrow{d} N(0,V) \text{ as } T \to \infty, \tag{17}$$

where strict stationarity implies the asymptotic variance V is given by

$$\begin{aligned} V &= \lim_{T\to\infty} E\{(2/T^{1/2}) \sum_t \{ \sum_i G_i(h_{1i}(Z_t)-\theta_i)\}\}^2 \\ &= 4E\{J(Z_1)^2 + 2 \sum_{k>1} J(Z_1)J(Z_k)\}, \end{aligned} \tag{18}$$

where $J(Z_k) \equiv \sum_i G_i(h_{1i}(Z_k) - \theta_i)$. The infinite series in (18) converges by the weak dependence assumed in the maintained regularity hypothesis, DK. □

It is now time to discuss practical problems of implementing a test of nonlinear Granger causality based upon Proposition 2.1. The problems are several. First, a choice of ϵ_1, ϵ_2 and the lags p, q must be made. Since the units of X and Y will typically differ, we have dealt with this problem in the application to money and income in Section 4 below by dividing each series by its estimated standard deviation. This normalizes each series to have standard deviation unity. However, this procedure introduces an estimated parameter problem, i.e., the estimated standard deviation, which could change the asymptotic distribution of the test statistic under the null. More will be said about this problem when we give the application below.

After normalizing each series by the standard deviation we choose the ϵ's so that $\epsilon_1=\epsilon_2$ and ϵ is between 1/2 and 3/2. This particular range of ϵ was motivated by the Monte Carlo studies done on the one dimensional case by Hsieh and LeBaron in Brock, Hsieh, LeBaron (1991). Hsieh and LeBaron showed that this range of ϵ gave good results on size and power for a particular experimental design which was motivated by applications to economics and finance.

Second, there are many test statistics besides (10) that one could use to test H_o. For example, if one had a particular point alternative H_a in mind then one could design a test of H_o that maximized power against H_a. The study of such problems lies beyond the scope of this paper.

Third, Dechert (1989) has shown that a closely related one dimensional version of our test is not consistent. That is to say there are alternatives to H_o against which the test has zero power. We suspect that Dechert's counterexample could be adapted to show that our test is not consistent against all alternatives to H_o.

Fourth, under the general maintained hypothesis, DK, of Denker and Keller (1983), the variance formula is given by an absolutely convergent infinite series. The issue arises of how to cope with an infinite number of terms in the construction of a consistent estimator for the variance under H_o. Strictly speaking, one must let the number of terms in the series grow slowly enough as the sample size grows in order to obtain a consistent estimator. Furthermore, the estimator must be positive on small samples in order to be a variance estimator.

In order to remain within the scope of the current paper we dealt with these problems by constructing test statistics with simple variance formulae. The test statistics are constructed in such a way as to have power against a wide class of alternatives to H_o but to minimize power against the subclass in H_o of processes that only depend upon past X's, not past Y's, but give, at the same time, simple variance formulae.

Consider the following equivalent statement of H_o,

$$H_o: \; C(x',x,y;e)C(x;e)-C(x',x;e)C(x,y;e) = 0. \tag{19}$$

Equation (19) suggests the following corollary of Proposition 2.1,

<u>Corollary 2.2</u>: Assume H_o and DK. Then,

$$T^{1/2}\{C(x',x,y;e,T)C(x;e,T)-C(x',x;e,T)C(x,y;e,T)\} \xrightarrow{d} N(0,V)$$
$$\text{as } T \to \infty. \tag{20}$$

Proof: Follow the same type of argument as in Proposition 2.1. □

Although removing ratios in the statistic helps avoid numerical instability problems the variance formula is still an infinite series under H_o. There are three strategies that one can use to find an estimator to approximate the variance, V, in (20). The first is to construct consistent estimators for each term in the expansion (18) and let the number of estimated terms grow slowly enough with respect to the sample size to achieve consistent estimation of V. The second is to ignore the intertemporal terms and simply construct a consistent estimator for the first term in (18). The third is to ignore the cross dependence in the first term of (18) and construct an estimator for it under the assumption that $\{X_t\}$, $\{Y_t\}$ are mutually independent and independently and identically distributed (IIDI). Note that we have listed the approximations in order of coarseness.

The third approximation for V in Corollary 2.5 in this paper is used for an application to money - income causal relationship. The intuition follows. Since the numerator of the test statistic in (10) is multiplied by $T^{1/2}$ we placed the most importance at getting a faithful approximation to H_o for the numerator of the test statistic that was close to zero if H_o was true. The motivation for choosing a coarse approximation for V was a crude attempt to optimize the following tradeoff. The third level of approximation gives a theoretically inaccurate formula, but one that contains few terms to estimate. This is to be traded against a formula that was theoretically more accurate but contained more terms to estimate.[1)]

Let us define notations, h_{1i} and θ_i, to derive the variance formula V in (20) under the second level approximation. The unit of the scale parameter, ϵ, for standardized series, $\{X_t\}$ and $\{Y_t\}$, is commonly standard deviation of each series. Put

$$h_{11}(X_t, X_{t,p}, Y_{t,q}) \equiv E[I\{|X_t - X_s| < \epsilon, |X_{t,p} - X_{s,p}| < \epsilon, |Y_{t,q} - Y_{s,q}| < \epsilon\}; \text{ given } X_t, X_{t,p}, Y_{t,q}],$$

$$h_{12}(X_{t,p}) \equiv E[I\{|X_{t,p}-X_{s,p}|<\epsilon; \text{ given } X_{t,p}],$$
$$h_{13}(X_t,X_{t,p}) \equiv E[I\{|X_t-X_s|<\epsilon, |X_{t,p}-X_{s,p}|<\epsilon; \text{ given } X_t,X_{t,p}],$$
$$h_{14}(X_{t,p},Y_{t,q}) \equiv E[I\{|X_{t,p}-X_{s,p}|<\epsilon, |Y_{t,q}-Y_{s,q}|<\epsilon\}; \text{ given } X_{t,p},Y_{t,q}],$$

and $\theta_i \equiv E[h_{1i}(.)]$ for each i. Omitting the arguments of the conditional expectations, $h_{1i}(\cdot)$, we devise the useful notations for the second moment calculations of $h_{1i}(\cdot)$. Under H_o of no Granger causality from Y to X,

$$E(h_{11})^2 \equiv K(x'x)K(xy,x^{-1}),\ E(h_{12})^2 \equiv K(x),\ E(h_{13})^2 \equiv K(x'x),$$
$$E(h_{14})^2 \equiv K(xy),\ E(h_{11}h_{12}) \equiv B(x'x,xy),\ E(h_{11}h_{13}) \equiv K(x'x)B(xy,x^{-1}),$$
$$E(h_{11}h_{14}) \equiv K(xy)B(x'x,x^{-1}),\ E(h_{12}h_{13}) \equiv B(x'x,x),$$
$$E(h_{12}h_{14}) \equiv B(xy,x), \text{ and } E(h_{13}h_{14}) \equiv B(x'x,xy)$$

where,

$$C(x) = \int[\int I\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})]dF(X_{t,p})$$
$$C(x'x) = \int[\int I_\epsilon(X_t,X_s)I\epsilon(X_{t,p},X_{s,p})dF(X_s,X_{s,p})]dF(X_t,X_{t,p})$$
$$C(xy) = \int[\int I_\epsilon(X_{t,p},X_{s,p})I\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})]dF(X_{t,p},Y_{t,q})$$
$$K(x) = \int[\int I_\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})]^2 dF(X_{t,p}),$$
$$K(x'x) = \int[\int I_\epsilon(X_t,X_s)I_\epsilon(X_{t,p},X_{s,p})dF(X_s,X_{s,p})]^2 dF(X_t,X_{t,p}),$$
$$K(xy,x^{-1}) = \int[\int I_\epsilon(X_{t,p},X_{s,p})I_\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})/\int I_\epsilon(X_{t,p},X_{s,p}) dF(X_{s,p})]^2 dF(X_{t,p},Y_{t,q}),$$
$$K(xy) = \int[\int I_\epsilon(X_{t,p},X_{s,p})I_\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})]^2 dF(X_{t,p},Y_{t,q}),$$
$$B(x'x,xy) = \int[\{\int I_\epsilon(X_t,X_s)I_\epsilon(X_{t,p},X_{s,p})dF(X_s,X_{s,p})\} \{\int I_\epsilon(X_{t,p},X_{s,p})I_\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})\}]dF(X_t,X_{t,p},Y_{t,q}),$$
$$B(xy,x^{-1}) = \int[\{\int I_\epsilon(X_{t,p},X_{s,p})I_\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})\}/ \{\int I_\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})\}]dF(X_{t,p},Y_{t,q})$$
$$B(xy,x) = \int[\{\int I_\epsilon(X_{t,p},X_{s,p})I_\epsilon(Y_{t,q},Y_{s,q})dF(X_{s,p},Y_{s,q})\}$$

$$\{\int I_\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})\}]dF(X_{t,p},Y_{t,q}),$$

$$B(x'x,x) = \int[\{\int I_\epsilon(X_t,X_s)I_\epsilon(X_{t,p},X_{s,p})dF(X_s,X_{s,p})\}$$
$$\{\int I_\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})\}]dF(X_t,X_{t,p})$$

$$B(x'x,x^{-1}) = \int[\{\int I_\epsilon(X_t,X_s)I_\epsilon(X_{t,p},X_{s,p})dF(X_s,X_{s,p})\}/$$
$$\{\int I_\epsilon(X_{t,p},X_{s,p})dF(X_{s,p})\}]dF(X_t,X_{t,p})$$

where $F(\cdot)$ is the unconditional cumulative distribution of the corresponding random variables. The notation " $'$ " is for the current forecasting horizon so that the variable x or y is automatically its lagged variables.

Corollary 2.3: Assume H_o and DK. Then the variance V in (20) under the second level approximation

$$\begin{aligned}V = 4[&C(x'x)^2C(xy)^2K(x)/C(x)^2 + 4C(x'x)C(xy)B(x'x,xy)\\ &- 2C(x'x)^2C(xy)B(xy,x)/C(x) - 2C(x'x)C(xy)^2B(x'x,x)/C(x)\\ &+ C(x)^2K(x'x)K(xy,x^{-1}) - 2C(x)C(x'x)K(xy)B(x'x,x^{-1})\\ &- C(x)C(xy)K(x'x)B(xy,x^{-1}) + C(x'x)^2K(xy) + C(xy)^2K(x'x)]. \qquad (21)\end{aligned}$$

Proof: We know that $J(Z_1) = \sum_{i=1}^{4} G_i(h_{1i}(Z_1)-\theta_i)$ where $(G_1,G_2,G_3,G_4) = (\theta_2,\theta_1,-\theta_4,-\theta_3)$, then the first term in (18) approximates the variance

$$\begin{aligned}V &\approx 4E[J(Z_1)^2]\\ &= -16(\theta_1\theta_2-\theta_3\theta_4)^2 + 4[\theta_1^2E(h_{12})^2 + 2\theta_1\theta_2E(h_{11}h_{12}) - 2\theta_1\theta_3E(h_{12}h_{14}) -\\ &\quad 2\theta_1\theta_4E(h_{12}h_{13}) + \theta_2^2E(h_{11})^2 - 2\theta_2\theta_3E(h_{11}h_{14}) - 2\theta_2\theta_4E(h_{11}h_{13}) +\\ &\quad \theta_3^2E(h_{14})^2 + 2\theta_3\theta_4E(h_{13}h_{14}) + \theta_4^2E(h_{13})^2]. \qquad (22)\end{aligned}$$

By definition of θ_i, θ_2=C(x), θ_3=C(x′x), θ_4=C(xy). Since $\theta_1\theta_2 = \theta_3\theta_4$ under H_o, θ_1=C(x′x)C(xy)C(x)$^{-1}$. Finally, replace θ_i and the second moments of h_{1i} with C(·)'s, K(·)'s, and B(·)'s in (22) to obtain (21). □

If we care about intertemporal terms due to weak dependence in (18), the number of terms will increase geometrically according to the parameter k. For instance, the variance formula up to k=2 will be

$$\begin{aligned} V = & -48(\theta_1\theta_2-\theta_3\theta_4)^2 + 4[\theta_1^2 E(h_{12})^2 + 2\theta_1\theta_2 E(h_{11}h_{12}) - 2\theta_1\theta_3 E(h_{12}h_{14}) - \\ & 2\theta_1\theta_4 E(h_{12}h_{13}) + \theta_2^2 E(h_{11})^2 - 2\theta_2\theta_3 E(h_{11}h_{14}) - 2\theta_2\theta_4 E(h_{11}h_{13}) + \\ & \theta_3^2 E(h_{14})^2 + 2\theta_3\theta_4 E(h_{13}h_{14}) + \theta_4^2 E(h_{13})^2] + 8[\theta_1^2 E(h_{12}h_{12}^*) + \\ & \theta_1\theta_2\{E(h_{11}h_{12}^*) + E(h_{12}h_{11}^*)\} - \theta_1\theta_3\{E(h_{12}h_{14}^*) + E(h_{14}h_{12}^*)\} - \\ & \theta_1\theta_4\{E(h_{12}h_{13}^*)+E(h_{13}h_{12}^*)\} + \theta_2^2 E(h_{11}h_{11}^*) - \theta_2\theta_3\{E(h_{11}h_{14}^*) + \\ & E(h_{14}h_{11}^*)\} - \theta_2\theta_4\{E(h_{11}h_{13}^*) + E(h_{13}h_{11}^*)\} + \theta_3^2 E(h_{14}h_{14}^*) + \\ & \theta_3\theta_4\{E(h_{13}h_{14}^*) + E(h_{14}h_{13}^*)\} + \theta_4^2 E(h_{13}h_{13}^*)], \end{aligned} \tag{23}$$

where * denotes the conditional probabilities of $h_{1i}(.)$ for the one-period ahead, i.e. $h_{11}^*=h_{11}(X_{t+1},X_{t+1,p},X_{t+1,q})$, $h_{12}^*=h_{12}(X_{t+1,p})$, $h_{13}^*=h_{13}(X_{t+1},X_{t+1,p})$ $h_{14}^*=h_{14}(X_{t+1,p},X_{t+1,q})$. Define all the second moment of h_{1i} for the current and one-period ahead in the following way:

$$\begin{aligned} & E(h_{11}h_{11}^*) = B(x'x,xy,x''x',x'y,x^{-1},x'^{-1}), \quad E(h_{12}h_{12}^*) = B(x,x'), \\ & E(h_{13}h_{13}^*) = B(x'x,x''x'), \quad E(h_{14}h_{14}^*) = B(xy,x'y'), \\ & E(h_{11}h_{12}^*+h_{12}h_{11}^*) = B(x'x,xy,x',x^{-1}) + B(x''x',x'y',x,x'^{-1}), \\ & E(h_{11}h_{13}^*+h_{13}h_{11}^*) = B(x'x,xy,x''x',x^{-1}) + B(x''x',x'y',x'x,x'^{-1}), \\ & E(h_{11}h_{14}^*+h_{14}h_{11}^*) = B(x'x,xy,x'y',x^{-1}) + B(x''x',x'y',xy,x'^{-1}), \end{aligned}$$

$E(h_{12}h^*_{13}+h_{13}h^*_{12}) = B(x,x''x') + B(x',x'x)$,
$E(h_{12}h^*_{14}+h_{14}h^*_{12}) = B(x,x'y')+B(x',xy)$,
$E(h_{13}h^*_{14}+h_{14}h^*_{13}) = B(x'x,x'y') + B(x''x',xy)$

where $B(\cdot)$s - the expected values of multiplication of the conditional probabilities - can be easily defined in a similar way to other $B(\cdot)$s previously defined for Corollary 2.3. The notation " $''$ " is for one-period ahead forecasting horizon.

Corollary 2.4: Let V_1 be the variance in (21) which is the first term in (18). The variance V in (20) when p=q=1 under H_o becomes,

$$V = V_1 + V_2 \quad \text{where}$$
$$\begin{aligned} V_2 = 8[&C(x'x)^2C(xy)^2B(x,x')/C(x)^2 + C(x'x)C(xy)\{B(x'x,xy,x',x^{-1}) \\ &+ B(x''x',x'y',x,x'^{-1})\} - C(x'x)^2C(xy)\{B(x,x'y')+B(x',xy)\}/C(x) \\ &- C(x'x)C(xy)^2\{B(x,x''x')+B(x',x'x)\}/C(x) + C(x)^2\{B(x'x,xy,x''x', \\ &x'y,x^{-1},x'^{-1})\} - C(x)C(x'x)\{B(x'x,xy,x'y',x^{-1}) + B(x''x',x'y',xy, \\ &x'^{-1})\} - C(x)C(xy)\{B(x'x,xy,x''x',x^{-1}) + B(x''x',x'y',x'x,x'^{-1})\} + \\ &C(x'x)^2B(xy,x'y') + C(x'x)C(xy)\{B(x'x,x'y')+B(x''x',xy)\} + C(xy)^2 \\ &B(x'x,x''x')]. \end{aligned} \tag{24}$$

Proof: Under H_o, $\theta_1\theta_2=\theta_3\theta_4$ holds and replace all the second moment of $h_{1i}(\cdot)$ and $h^*_{1i}(\cdot)$ with $B(\cdot)$s defined above to get (24). □

The complicated variance formula V in (21) or (24) is dramatically simplified under the third level approximation. Corollary 2.5 below gives the results for, (a) p=q=1; (b) general

$p,q<\infty$.

Corollary 2.5: Under the third level approximation (i.e., $\{X_t\}$, $\{Y_t\}$ IIDI) we have, (a), for p=q=1,

$$V = 4[C(x)]^2K(x)[K(x)-C(x)^2][K(y)-C(y)^2], \tag{25}$$

(b), for $p,q<\infty$,

$$V = 4[C(x)]^{2p}K(x)^p[K(x)-C(x)^2][K(y)^q-C(y)^{2q}]. \tag{26}$$

Proof: We will show only (a) here because (b) can be shown analogously. Under the assumption of IIDI,

$C(x'x)=C(x)^2$, $C(xy)=C(x)C(y)$, $K(x'x)=K(x)^2$, $K(xy)=K(x)K(y)$,
$K(xy,x^{-1})=K(y)$, $B(x,x'x)=K(x)C(x)$, $B(x,xy)=K(x)C(y)$, $B(x'x,x^{-1})=C(y)$,
$B(xy,x^{-1})=C(y)$, $B(x'x,xy)=K(x)C(x)C(y)$, $B(x,x')=C(x)^2$, $B(x,x''x')=C(x)^3$,
$B(x',x'x)=K(x)C(x)$, $B(x,x'y')=C(x)^2C(y)$, $B(x',xy)=C(x)^2C(y)$,
$B(x'x,x''x')=K(x)C(x)^2$, $B(xy,x'y')=C(x)^2C(y)^2$,
$B(x'x,x'y')=K(x)C(x)C(y)$, $B(x''x',xy)=C(x)^3C(y)$,
$B(x'x,xy,x',x^{-1})=K(x)C(x)C(y)$, $B(x''x',x'y',x,x'^{-1})=C(x)^3C(y)$,
$B(x'x,xy,x''x',x^{-1})=K(x)C(x)^2C(y)$, $B(x''x',x'y',x'x,x'^{-1})=K(x)C(x)^2C(y)$,
$B(x'x,xy,x'y',x^{-1})=K(x)C(x)C(y)^2$, $B(x''x',x'y',xy,x'^{-1})=C(x)^3C(y)^2$,
$B(x'x,xy,x''x',x'y,x^{-1},x'^{-1})=K(x)C(x)^2C(y)^2$.

Replacing $B(\cdot)$s, $C(\cdot)$s, and $K(\cdot)$s in (21) and (24) with much simpler expressions above, we show that

$V_1 = 4[C(x)]^2 K(x)[K(x)-C(x)^2][K(y)-C(y)^2]$ and $V_2 = 0$. Therefore the variance V under IIDI assumption is exactly same as that in (25). □

If $\{X_t\}$, $\{Y_t\}$ are IIDI we have,

$$C(x',x,y;e)C(x;e) - C(x',x;e)C(x,y;e) =$$
$$C(x',x,y;e) - C(x';\epsilon_1)C(x;\epsilon_1)C(y;\epsilon_2) = 0. \qquad (27)$$

Tests of IIDI can be based upon the test statistic,

$$T^{1/2}S(T) \equiv T^{1/2}\{C(x',x,y;e,T)-C(x';\epsilon_1,T)C(x;\epsilon_1,T)C(y;\epsilon_2,T)\}. \qquad (28)$$

Under IIDI, Baek and Brock (1992) show (28) is asymptotically normal with mean 0 and variance V. Note that tests based upon

$$S \equiv C(x',x,y;e) - C(x';\epsilon_1)C(x;\epsilon_1)C(y;\epsilon_2), \qquad (29)$$

are not appropriate for testing H_o (cf. (8) or (19)) because H_o can be true while S is not equal to zero because of temporal dependence in $\{X_t\}$ even though past Y's do not help predict future X's.

3. SIZE PERFORMANCE

The size performance of the nonlinear Granger causality test based upon (24) is studied by a small Monte Carlo experiment. The size is calculated under the assumption that $\{X_t\}$, $\{Y_t\}$ is IIDI.[2)] Hence the experiment we report should be taken as suggestive only. We generated two series of 320 N(0,1) pseudo-random numbers by using the

IMSL Fortran Subroutine. We calculated the test statistic R=2500 times using the formula[3)]

$$T^{1/2}[C(x',x,y)C(x) - C(x',x)C(x,y)]/V^{1/2} \qquad (30)$$

where V in (26) is consistently estimated from the given sample.

Figure 1 gives a sample histogram for p=3, q=3, ϵ=1, R=2500.[4)] All of the histograms appear unimodal and bell-shaped. We looked at size, sample statistics, and sample quantiles for p=q ranging from 1 to 6 and ϵ=1 and 1.5. The choice of ϵ values was motivated by the findings of Hsieh and LeBaron in Brock, Hsieh, and Lebaron (1991) that the related BDS (Brock, Dechert, and Scheinkman (1987)) statistic had best size and power performance for ϵ in [0.5,1.5].

Turn now to Table 1 for the size of the test statistic. Note that, as p=q increases the size increases. Therefore we must use size-corrected critical values for practical use. For small values of ϵ, the size is biased upward dramatically for higher values of p=q. This appears to be caused by scarcity of the number of pairs of m-histories that are within ϵ in distance of each other for small ϵ. Because of this problem we recommend use of ϵ=1.5 rather than ϵ=1. A serious issue is raised by the choice of ϵ and the lags p, q. Presumably if one had an alternative to the null which one wished maximal power against, then one could choose ϵ, p, q to maximize power against that alternative. The optimal choice of ϵ, p, q is beyond the scope of this paper.

The second table, Table 2, exhibits the quality of the approximation to normality. Even though sample means, medians, and standard deviations are very close to the standard normal distribution, all test statistics are skewed to the right, and except for a few cases, they are also leptokurtic. The test statistic approximates the normal distribution better for $\epsilon=1.5$ than for $\epsilon=1$. In Table 3, we correct the quantiles of the nonlinear Granger causality statistic for practical inference. Since size is very biased for certain parameter values, we recommend use of these size corrected critical values. Note that the problem of bias does not arise from poor accuracy of our estimates of $C(x;\epsilon)$, $C(y;\epsilon)$, $K(x;\epsilon)$, $K(y;\epsilon)$. Table 4 reports simulated values of these quantities and shows that they are close to the theoretical values reported by Hsieh and Lebaron in Brock, Hsieh, LeBaron (1991).

Another strategy for dealing with the problems caused by poor quality asymptotics is to use the bootstrap to bootstrap the distribution of the test statistic under the null. Versions of bootstrap technology are now available for general stationary observations. See Kunsch (1989), and Leger, Politis, Romano (1991). Turn now to an application to money and income.

4. AN APPLICATION TO MONEY AND INCOME

In this part of the paper we apply the nonlinear Granger causality test to money and income. The data consists of monthly observations from January, 1959 to December, 1985, on seasonally adjusted nominal M1 and industrial production from Stock and Watson (1989). We thank Stock and Watson for giving us this data. Following Stock and Watson, we take logarithms (denoted by m and y), and estimate the bivariate Vector AutoRegression (VAR) given below,

$$y_t = \underset{(2.34)}{0.21} + \underset{(2.00)}{1.60\times10^{-4}}t + \underset{(24.56)}{1.395}\, y_{t-1} - \underset{(-4.50)}{0.441}\, y_{t-2} + \underset{(0.87)}{0.088}\, y_{t-3}$$

$$- \underset{(-0.60)}{0.061}\, y_{t-4} - \underset{(-0.62)}{0.062}\, y_{t-5} + \underset{(0.92)}{0.052}\, y_{t-6} + \underset{(2.22)}{0.287}\, m_{t-1}$$

$$- \underset{(-1.44)}{0.291}\, m_{t-2} + \underset{(1.43)}{0.289}\, m_{t-3} - \underset{(-1.69)}{0.342}\, m_{t-4} + \underset{(1.57)}{0.317}\, m_{t-5}$$

$$- \underset{(-2.16)}{0.279}\, m_{t-6} + u_{yt} \tag{31}$$

$$m_t = \underset{(2.30)}{0.088} + \underset{(2.73)}{9.61\times10^{-5}}t + \underset{(2.15)}{0.054}\, y_{t-1} - \underset{(-1.54)}{0.066}\, y_{t-2} - \underset{(-0.32)}{0.015}\, y_{t-3}$$

$$- \underset{(-0.33)}{0.014}\, y_{t-4} + \underset{(1.66)}{0.071}\, y_{t-5} - \underset{(-1.47)}{0.037}\, y_{t-6} + \underset{(21.92)}{1.245}\, m_{t-1}$$

$$- \underset{(-3.93)}{0.349}\, m_{t-2} + \underset{(2.49)}{0.221}\, m_{t-3} - \underset{(-4.13)}{0.367}\, m_{t-4} + \underset{(4.11)}{0.366}\, m_{t-5}$$

$$- \underset{(-2.28)}{0.130}\, m_{t-6} + u_{mt}. \tag{32}$$

Here the F-test statistic=3.10 [2.09] with significance level 0.0060 [0.0542] for the restriction of all zero money [output] coefficients.

The issue we wish to treat in our paper concerns predictive content of one series for the other beyond linear predictability. Hence, we apply the nonlinear Granger causality test to the residuals of the VAR model (31), (32). The above VAR was an attempt to extract the linear structure. Application of the nonlinear causality test to estimated residuals of linear models raises three issues. First, we may have miss-specified the linear model even if the data generating process was linear. We need to be precise about what we mean by "linear."

We define a data generating process to be linear if it is generated by a Wold-like representation with IID innovations. Note that it is IID innovations not just uncorrelated innovations that gives the definition content. Although we shall not deal with it in this paper, we want to point out that IID-linearity is not the only useful definition of linearity. Another definition is MDS-linearity where "IID" is replaced by "MDS" and "MDS" stands for Martingale Difference Sequence. We attempted to deal with the linear miss-specification problem by using a likelihood ratio test to select the optimal lag length.

A second, and possibly more serious issue, is the use of estimated residuals in the nonlinear causality test. This causes an estimated parameters problem. Even if the correct linear model is estimated and all of the potential incremental predictability lies in the innovations of this linear model and even if H_o is true, the asymptotic distribution of the test statistic under H_o will typically be distorted.

Baek and Brock (1992) have shown that a test of the null hypothesis of IIDI model innovations based on (27) has the same asymptotic distribution whether estimated residuals or true residuals are used of IID-linear models provided the data generating process is IID-linear and the correct IID-linear model is estimated. But that result does not apply to a test of H_o (cf. (8) or (20)) on the residuals of an estimated linear model. It is beyond the scope of this paper to deal with it here.

A third issue is how to interpret the results of a rejection of the null hypothesis when the test is applied to estimated residuals. The problems are several. For example, the rejection may be due to neglected nonstationarity or other kinds of dependence which may be of no or little direct use for forecasting purposes such as heteroscedasticity.

Put $X \equiv u_{yt}$, $Y \equiv u_{mt}$. We first test whether money residuals help predict income residuals, i.e.,

$$H_o(u_m \not\Rightarrow u_y): \quad C(x',x,y;e)C(x;e) - C(x',x;e)C(x,y;e) = 0, \tag{33}$$

second, we reverse the roles of X and Y to test whether income residuals help predict money residuals. The residual series were normalized by their estimated standard deviations before applying the tests.

We use the notation $H_o(u_m \not\Rightarrow u_y)$, for the null hypothesis: $\{u_m\}$ does not incrementally help predict $\{u_y\}$. Recall that $\{u_m\}$, $\{u_y\}$ are residuals of the VAR (31), (32) so incremental predictive power of one residual series for another measured by test statistic (20) is, by definition, incremental nonlinear predictive power.

Table 5: Nonlinear Granger causality test for Money and Income

1959:12 - 1985:12

VAR Lag = 6, ϵ = 1.0

(p,q)	(1,1)	(2,2)	(3,3)	(4,4)	(5,5)	(6,6)
money → income	2.06 (0.0330)	2.42 (0.0284)	2.27 (0.0772)	2.23 (0.1584)	2.70 (0.2016)	1.72 (0.3436)
income → money	0.13 (0.4276)	-1.70 (0.0668)	-3.13 (0.0140)	-2.18 (0.1432)	-4.01 (0.0980)	3.98 (0.2012)

VAR Lag = 6, ϵ = 1.5

(p,q)	(1,1)	(2,2)	(3,3)	(4,4)	(5,5)	(6,6)
money → income	0.82 (0.2024)	1.18 (0.1388)	0.95 (0.1880)	1.86 (0.0608)	1.47 (0.1164)	0.57 (0.3296)

income → money	-0.13 (0.4536)	-1.08 (0.1352)	-1.86 (0.0328)	-1.92 (0.0380)	-3.27 (0.0020)	-2.12 (0.0532)

Notes: P-valuses, which are based on Table 3, are in parenthesis. The null hypothesis of the first row is that money does not Granger cause income and the second row's null hypothesis is that income does not Granger cause money.

Note that we are testing for incremental nonlinear predictability above and beyond linear predictability. When we apply a two-tail test, the test statistics demonstrate that the first null hypothesis, money does not incrementally nonlinearly Granger cause income, is significant at 10% level at (1,1) and (2,2) variable lags for ϵ=1.0 since their p-values are 0.03 and 0.0284. Moreover the second null hypothesis that income does not incrementally nonlinearly Granger cause money becomes significant at (3,3) for ϵ=1.0, and at (3,3), (4,4) and (5,5) for ϵ=1.5 since their p-values are 0.014, 0.0328, 0.038, and 0.002 respectively.

The results in the Table above are consistent with the possibility that the bivariate VAR does not capture all the structure between money and income. A possible reason may be variation in the strength of the causal relation across the business cycle.[5)] However, we wish to empasize, as was stressed earlier, that rejection of the null hypothesis is consistent with neglected nonstationarity and neglected heteroscedasticity as well as potentially useful nonlinear predictability. Given this caveat the results above suggest that it might be worthwhile to see if improved prediction of money and income can be done by a nonlinear model.

We empasize that the test was applied to normalized estimated residuals of an estimated bivariate VAR which contained estimated trends. This introduces at least three estimated parameters problems which may contaminate the asymptotic distribution of the test statistic even if we correctly detrended and correctly identified the VAR and even if the correct VAR has IIDI innovations.

One way to attempt to correct for this problem is to bootstrap the distribution of the test statistic under the bivariate model we fitted by resampling the residuals of that model. In this way a 5%, say, rejection region could be constructed under the null hypothesis that the data was generated by our bivariate VAR with IIDI innovations against the alternative of a nonlinear model. Of course this procedure relies on correct identification of the VAR if the data is truly VAR.

The bootstrap could also be used to correct for the estimated parameters problem caused by normalizing each residual series by its estimated standard deviation.

Finally the estimated parameters problem could be dealt with in the usual way by using a Taylor series argument to correct the asymptotic distribution. That strategy was used by Baek and Brock (1992) on a simpler problem. These interesting questions must await further research.

5. SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FUTURE RESEARCH

In this paper we used a measure of local spatial correlation, called the correlation integral, in order to build a statistical test for nonlinear (Granger) causal orderings between two variables. The test could be implemented on raw data or, if there is a strong presumption of a linear relationship, such as VAR's in aggregative macro econometrics, the test could be implemented on residuals of fitted linear models.

We executed a small Monte Carlo study of this test and found serious bias in the size. Therefore we recommend use of size-adjusted critical values or use of the bootstrap to bootstrap the distribution of the statistic under the null. The bootstrap is a way of dealing with the estimated parameters problem that arises in application of the test to residuals of VAR's.

We applied our test to the residuals of a bivariate VAR fit to money and income data and found some evidence consistent with nonlinear predictability in these residuals.

Given the increasing speed of modern computers we feel that, in time, our methods will become useful for nonlinear causality testing. However, more work needs to be done on techniques to improve size and power performance before the techniques advocated in this paper are of practical use.

FOOTNOTES

1. Craig Hiemstra (1992) has done preliminary work that suggests our neglect of the intertemporally dependent terms hurts the size performance of the test a lot. To resolve this problem, we use the size adjusted critical values in the application section.

2. Hiemstra pointed out that asymptotic normality based critical values from the IIDI assumption lead to a serious size problem. To avoid

this criticism, we obtain and critical values from the quantiles table (Table 3) for the money-income ~~causality~~ applications.

3. It costs about 230 U.S. dollars to run one Fortran simulation program on a WYLBUR mainframe for a given ϵ when the number of replications is 2500 for a sample size of 320. The GAUSS programming language was used to do the money and income application. This application took about 4 minutes for p=q=6 and took less than 2 minutes for p=q=1 on a 20-MHz 386 PC.

4. Histograms for a range of values of p, q, ϵ may be had by writing the authors at the Korean address.

5. After we completed this work, Simon Potter gave us a copy of Holmes and Tufte (1991). They show that conclusions drawn on causality relations between real GNP and the nominal monetary base are dependent on business cycle asymmetry. More precisely "several measures of the monetary base are prima facie causes of real GNP only during contractions. These results suggest that studies of the efficacy of monetray policy should explicitly account for business cycle asymmetry." (Holmes and Tufte, (1991) abstract). It may be possible that their findings lie behind our results, but this requires more investigation.

REFERENCES

Baek, E. and W. Brock (1992), 'A nonparametric test for independence of a multivariate time series', Statistica Sinica, 2, 137-156.

Brock, W., W. Dechert, and J. Scheinkman (1987), 'A test for independence based on the correlation dimension', SSRI paper #8702, Department of Economics, University of Wisconsin-Madison.

Brock, W. and E. Baek (1991), 'Some theory of statistical inference for nonlinear science', Review of Economic Studies, 58, 697-716.

Brock, W., D. Hsieh and B. LeBaron (1991), Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence, MIT Press, Cambridge, Mass.

Dechert, W. (1989), 'An application of chaos theory to deterministic and stochastic observations', Department of Economics, The University of Houston.

Denker, M. and G. Keller (1983), 'On U-statistics and von Mises's statistics for weakly dependent processes', Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 64, 505-522.

Denker, M. and G. Keller (1986), 'Rigorous statistical procedures for data from dynamical systems', Journal of Statistical Physics, 44, 67-93.

Geweke, J., R. Meese and W. Dent (1983), 'Comparing alternative tests of causality in temporal systems', Journal of Econometrics, 21, 161-194.

Geweke, J (1984), 'Inference and causality in economic time series', in Griliches Z. and M. Intriligator (ed.), Handbook of econometrics, Volume 2, North-Holland, New York.

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

Hiemstra, C. (1992), 'Detection and description of nonlinear dynamics using correlation integral based estimators', Department of Economics, Loyola College.

Holmes, J. and D. Tufte (1991), 'Output, prices, and the monetary base over the business cycle', Department of Economics, SUNY at Buffalo and Department of Economics, University of Alabama.

Kunsch, H. (1989), 'The jackknife and the bootstrap for general stationary observations', Annals of Statistics, 17, 1217-1241.

Leger, C., D. Politis and J. Romano (1991), 'Bootstrap technology and applications', Department of Statistics, Stanford University.

Sims, C. (1972), 'Money, income and causality', American Economic Review, 62, 540-552.

Sims, C., J. Stock and M. Watson (1990), 'Inference in linear time series models with some unit roots', Econometrica, 58, 113-144.

Stock, J. and M. Watson (1989), 'Interpreting the evidence on money-income causality', Journal of Econometrics, 40, 161-181.

Table 1: Size of nonlinear Granger causality statistic

Standard Normal Distribution: Sample size = 320

p=1, q=1	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	0.56	0.52	1.00
% < -1.96	2.08	2.08	2.50
% < -1.64	4.84	4.36	5.00
% > 1.64	6.92	5.80	5.00
% > 1.96	3.60	3.12	2.50
% > 2.33	1.80	1.40	1.00

p=2, q=2	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	1.96	0.92	1.00
% < -1.96	4.52	2.12	2.50
% < -1.64	7.56	4.48	5.00
% > 1.64	9.20	7.08	5.00
% > 1.96	6.12	3.52	2.50
% > 2.33	3.60	1.84	1.00

p=3, q=3	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	5.32	1.52	1.00
% < -1.96	9.20	2.80	2.50
% < -1.64	13.00	5.72	5.00
% > 1.64	14.44	7.24	5.00
% > 1.96	11.04	4.56	2.50
% > 2.33	7.16	2.24	1.00

p=4, q=4	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	12.76	1.88	1.00
% < -1.96	16.68	3.68	2.50
% < -1.64	20.56	6.20	5.00
% > 1.64	21.48	8.40	5.00
% > 1.96	18.12	4.76	2.50
% > 2.33	14.48	2.32	1.00

p=5, q=5	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	22.52	1.96	1.00
% < -1.96	26.56	4.20	2.50
% < -1.64	29.72	7.88	5.00
% > 1.64	31.00	9.76	5.00
% > 1.96	27.28	5.84	2.50
% > 2.33	23.84	3.56	1.00

p=6, q=6	ϵ=1.0	ϵ=1.5	N(0,1)
% < -2.33	31.12	3.80	1.00
% < -1.96	34.28	6.36	2.50
% < -1.64	36.88	10.16	5.00
% > 1.64	34.80	11.32	5.00
% > 1.96	32.64	7.52	2.50
% > 2.33	30.00	4.92	1.00

Notes: IMSL subroutine DRNNOA was called to generate two sets of 320 standard normal random numbers,and RNSET was called to set initial seeds. Total number of replications is 2500. The parameters, p and q, denote lags of the first and the second variable for nonlinear Granger causlity test. Epsilon is the the scale parameter. Standard normal distribution is denoted by N(0,1).

Explain tables.

Table 2: Distribution of nonlinear Granger causality statistic

Standard Normal Distribution: Sample size = 320

p=1, q=1	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.00	0.02	0.00
Median	-0.04	0.01	0.00
Std dev	1.03	1.00	1.00
Skewkess	0.28	0.20	0.00
	(0.00)	(0.00)	
Kurtosis	3.05	3.06	3.00
	(0.63)	(0.51)	

p=2, q=2	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.01	0.03	0.00
Median	-0.03	-0.01	0.00
Std dev	1.19	1.04	1.00
Skewness	0.22	0.20	0.00
	(0.00)	(0.00)	
Kurtosis	3.01	3.22	3.00
	(0.96)	(0.03)	

p=3, q=3	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.03	0.03	0.00
Median	0.00	-0.01	0.00
Std dev	1.51	1.22	1.00
Skewkess	0.16	0.10	0.00
	(0.00)	(0.05)	
Kurtosis	3.02	3.08	3.00
	(0.83)	(0.39)	

p=4, q=4	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.07	0.04	0.00
Median	0.01	-0.03	0.00
Std dev	2.11	1.13	1.00
Skewness	0.23	0.08	0.00
	(0.00)	(0.11)	
Kurtosis	3.38	3.20	3.00
	(0.00)	(0.04)	

p=5, q=5	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.09	0.04	0.00
Median	-0.01	0.00	0.00
Std dev	3.36	1.22	1.00
Skewkess	0.19	0.25	0.00
	(0.00)	(0.00)	
Kurtosis	3.94	3.41	3.00
	(0.00)	(0.00)	

p=6, q=6	ϵ=1.0	ϵ=1.5	N(0,1)
Mean	0.10	0.04	0.00
Median	-0.14	0.01	0.00
Std dev	5.19	1.38	1.00
Skewness	0.42	0.21	0.00
	(0.00)	(0.00)	
Kurtosis	4.48	3.50	3.00
	(0.00)	(0.00)	

Notes: IMSL subroutine DRNNOA was called to generate two sets of 320 standard normal random numbers,and RNSET was called to set initial seeds. Total number of replications is 2500. The parameters, p and q, denote lags of the first and the second variable for nonlinear Granger causlity test. Epsilon is the the scale parameter. Standard normal distribution is denoted by N(0,1). Significance level is reported in parenthesis for skewness and kurtosis.

Table 3: Quantiles of nonlinear Granger causality statistic

Standard Normal Distribution: Sample Size = 320

p=1, q=1	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-2.18	-2.15	-2.33
2.5%	-1.88	-1.86	-1.96
5.0%	-1.63	-1.57	-1.64
10.0%	-1.29	-1.23	-1.28
25.0%	-0.74	-0.67	-0.67
75.0%	0.67	0.68	0.67
90.0%	1.39	1.30	1.28
95.0%	1.81	1.71	1.64
97.5%	2.14	2.60	1.96
99.0%	2.60	2.60	2.33

p=2, q=2	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-2.56	-2.31	-2.33
2.5%	-2.23	-1.92	-1.96
5.0%	-1.88	-1.61	-1.64
10.0%	-1.48	-1.25	-1.28
25.0%	-0.79	-0.68	-0.67
75.0%	0.78	0.70	0.67
90.0%	1.58	1.39	1.28
95.0%	2.06	1.79	1.64
97.5%	2.52	2.16	1.96
99.0%	3.01	2.70	2.33

p=3, q=3	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-3.24	-2.52	-2.33
2.5%	-2.81	-2.07	-1.96
5.0%	-2.36	-1.70	-1.64
10.0%	-1.91	-1.32	-1.28
25.0%	-1.03	-0.70	-0.67
75.0%	1.00	0.70	0.67
90.0%	2.04	1.43	1.28
95.0%	2.65	1.89	1.64
97.5%	3.16	2.29	1.96
99.0%	3.68	2.62	2.33

p=4, q=4	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-4.49	-2.60	-2.33
2.5%	-3.93	-2.20	-1.96
5.0%	-3.26	-1.75	-1.64
10.0%	-2.61	-1.34	-1.28
25.0%	-1.33	-0.70	-0.67
75.0%	1.39	0.78	0.67
90.0%	2.84	1.52	1.28
95.0%	3.66	1.93	1.64
97.5%	4.30	2.29	1.96
99.0%	5.20	2.79	2.33

p=5, q=5	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-7.72	-2.68	-2.33
2.5%	-6.32	-2.20	-1.96
5.0%	-5.48	-1.84	-1.64
10.0%	-3.98	-1.49	-1.28
25.0%	-2.10	-0.76	-0.67
75.0%	2.19	0.81	0.67
90.0%	4.34	1.62	1.28
95.0%	5.69	2.11	1.64
97.5%	6.78	2.57	1.96
99.0%	8.47	3.16	2.33

p=6, q=6	ϵ=1.0	ϵ=1.5	N(0,1)
1.0%	-11.22	-3.14	-2.33
2.5%	-9.71	-2.58	-1.96
5.0%	-7.91	-2.18	-1.64
10.0%	-6.08	-1.65	-1.28
25.0%	-3.11	-0.89	-0.67
75.0%	3.08	0.92	0.67
90.0%	6.52	1.75	1.28
95.0%	8.92	2.32	1.64
97.5%	11.43	2.88	1.96
99.0%	13.95	3.62	2.33

Notes: IMSL subroutine DRNNOA was called to generate two sets of 320 standard normal random numbers,and RNSET was called to set initial seeds. Total number of replications is 2500. The parameters, p and q, denote lags of the first and the second variable for nonlinear Granger causlity test. Epsilon is the the scale parameter. Standard normal distribution is denoted by N(0,1).

Figure 1

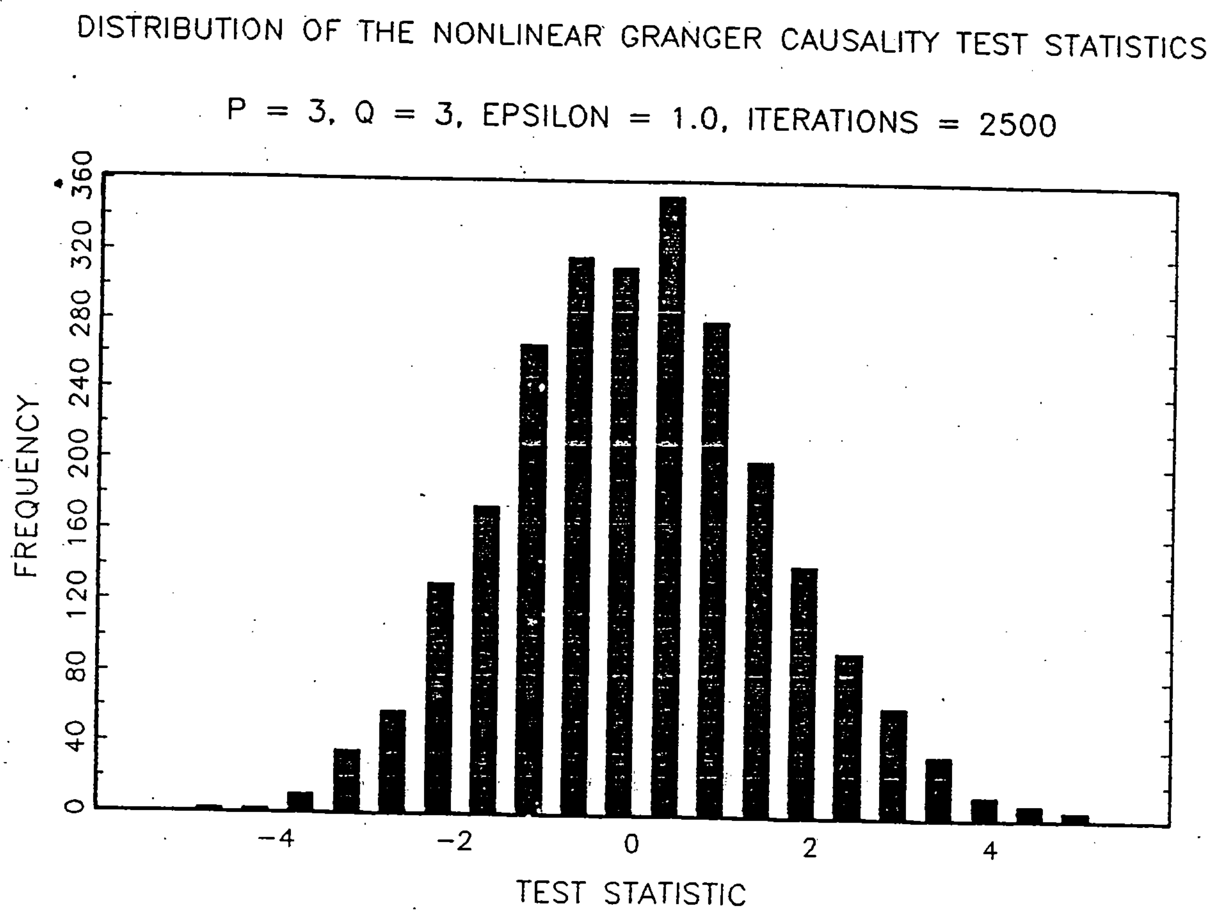
DISTRIBUTION OF THE NONLINEAR GRANGER CAUSALITY TEST STATISTICS
P = 3, Q = 3, EPSILON = 1.0, ITERATIONS = 2500
FREQUENCY
0
40
80
120
160
200
240
280
320
360
−4
−2
0
2
4
TEST STATISTIC

Table 4: Simulated Values of $C(x;\epsilon)$, $C(y:\epsilon)$, $K(x;\epsilon)$ and $K(y;\epsilon)$

	$\epsilon = 1.0$	$\epsilon = 1.5$
$C(x;\epsilon)$	0.5195 (0.0068)	0.7104 (0.0053)
$C(y;\epsilon)$	0.5196 (0.0066)	0.7104 (0.0052)
$K(x;\epsilon)$	0.2984 (0.0101)	0.5365 (0.0100)
$K(y;\epsilon)$	0.2985 (0.0099)	0.5366 (0.0099)

Notes: IMSL subroutine DRNNOA was called to generate two sets of 320 standard normal random numbers,and RNSET was called to set initial seeds. Total number of replications is 2500.