Multivariate causality tests with simulation and application

Bai, Zhidong; Li, Heng; WONG, Wing Keung; Zhang, Bingzhi

Link to publication

Multivariate Causality Tests with Simulation and Application

Zhidong Bai

School of Mathematics and Statistics, Northeast Normal University

Department of Statistics and Applied Probability,

National University of Singapore

Heng Li

Department of Mathematics

Hong Kong Baptist University

Wing-Keung Wong

Department of Economics

Hong Kong Baptist University

Bingzhi Zhang

Department of BioStatistics

Columbia University

Corresponding author: Wing-Keung Wong, Department of Economics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. Tel: (852)-3411-7542, Fax: (852)-3411-5580, Email: awong@hkbu.edu.hk

Multivariate Causality Tests with Simulation and Application

Abstract   This paper extends the test established by Hiemstra and Jones (1994) to develop a nonlinear causality test in a multivariate setting. A Monte Carlo simulation is conducted to demonstrate the superiority of our proposed multivariate test over its bivariate counterpart. In addition, we illustrate the applicability of our proposed test to analyze the relationships among different Chinese stock market indices.

1 Introduction

It is an important issue to detect the causal relation among several time series, see, for example, Qiao, et al (2008, 2009) and Chiang, et al (2009). To examine whether past information of one series could contribute to the prediction of another series, linear Granger causality test (Granger, 1969) is developed to examine whether lag terms of one variable significantly explain another variable in a vector autoregressive regression model. However, the linear Granger causality test does not perform well in detecting nonlinear causal relationships. To circumvent the limitation, Baek and Brock (1992) develop a bivariate nonlinear Granger causality test to examine the remaining nonlinear predictive power of a residual series of a variable on the residual of another variable obtained from a linear model. Hiemstra and Jones (1994) have further modified the test. In this paper, we first discuss the linear causality test in the multivariate setting and thereafter extend the theory by developing a nonlinear causality test in the multivariate setting. In addition, we conduct simulation to demonstrate the superiority of our proposed multivariate test over its bivariate counterpart in the performance of both size and power. At last, we illustrate the applicability of our proposed test to analyze the relationships among different Chinese stock market indices.

1

2 Multivariate Linear Granger Causality Test

We first review the linear Granger causality test in the multivariate setting.

2.1 Vector Autoregressive Regression

To test the linear causality relationship between two vectors of stationary time series, $x_t = (x_{1,t}, \cdots, x_{n_1,t})'$ and $y_t = (y_{1,t}, \cdots, y_{n_2,t})'$, where there are $n_1 + n_2 = n$ series in total, one could construct the following vector autoregressive regression (VAR) model:

$$
\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} A_{x[n_1 \times 1]} \\ A_{y[n_2 \times 1]} \end{pmatrix} + \begin{pmatrix} A_{xx}(L)_{[n_1 \times n_1]} & A_{xy}(L)_{[n_1 \times n_2]} \\ A_{yx}(L)_{[n_2 \times n_1]} & A_{yy}(L)_{[n_2 \times n_2]} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_{x,t} \\ e_{y,t} \end{pmatrix}, \quad (1)
$$

where $A_{x[n_1 \times 1]}$ and $A_{y[n_2 \times 1]}$ are two vectors of intercept terms, $A_{xx}(L)_{[n_1 \times n_1]}$, $A_{xy}(L)_{[n_1 \times n_2]}$, $A_{yx}(L)_{[n_2 \times n_1]}$, and $A_{yy}(L)_{[n_2 \times n_2]}$ are matrices of lag polynomials, $e_{x,t}$ and $e_{y,t}$ are the corresponding error terms.

To test the linear causality relationship between x_t and y_t is equivalent to testing the following null hypotheses: H_0^1: $A_{xy}(L) = 0$ and H_0^2: $A_{yx}(L) = 0$. There are four different situations for the causality relationships between x_t and y_t in (1): (a) rejecting H_0^1 but not rejecting H_0^2 implies a *unidirectional causality* from y_t to x_t, (b) rejecting H_0^2 but not rejecting H_0^1 implies a *unidirectional causality* from x_t to y_t, (c) rejecting both H_0^1 and H_0^2 implies existence of *feedback* relations, and (d) not rejecting both H_0^1 and H_0^2 implies that x_t and y_t are not rejected to be *independent*.

To test H_0^1 and/or H_0^2, one may first obtain the residual covariance matrix Σ from the full model in (1) without imposing any restriction on the parameters, and compute the residual covariance matrix Σ_0 for the restricted model in (1) with the restriction on the parameters imposed by the null hypothesis, H_0^1 and/or H_0^2. Thereafter, one could use the F-test or the likelihood ratio statistic $(T - c)\left(log|\Sigma_0| - log|\Sigma| \right)$ (Sims, 1980) to test for H_0^1 and/or H_0^2 where T is the number of usable observations, c is the number of parameters estimated in the unrestricted system.

2.2 ECM-VAR model

If the time series are conintegrated, one should impose the error-correction mechanism (ECM) on the VAR to test for Granger causality between the variables of interest. In particular, when testing the causality relationship between two vectors of non-stationary time series, we let $X_t = (X_{1,t}, \cdots, X_{n_1,t})'$ and $Y_t = (Y_{1,t}, \cdots, Y_{n_2,t})'$, and let $x_{it} = \Delta X_{it}$ and $y_{it} = \Delta Y_{it}$ be the corresponding stationary differencing series such that there are $n_1 + n_2 = n$ series in total. If X_t and Y_t are cointegrated, then, instead of using the VAR in (1), one should adopt the following ECM-VAR model:

$$
\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} A_{x[n_1 \times 1]} \\ A_{y[n_2 \times 1]} \end{pmatrix} + \begin{pmatrix} A_{xx}(L)_{[n_1 \times n_1]} & A_{xy}(L)_{[n_1 \times n_2]} \\ A_{yx}(L)_{[n_2 \times n_1]} & A_{yy}(L)_{[n_2 \times n_2]} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}
$$
$$
+ \begin{pmatrix} \alpha_{x[n_1 \times 1]} \\ \alpha_{y[n_2 \times 1]} \end{pmatrix} \cdot ecm_{t-1} + \begin{pmatrix} e_{x,t} \\ e_{y,t} \end{pmatrix} , \tag{2}
$$

where ecm_{t-1} is lag one of the error correction term, $\alpha_{x[n_1 \times 1]}$ and $\alpha_{y[n_2 \times 1]}$ are the coefficient vectors for the error correction term ecm_{t-1}. Thereafter, one could test the null hypothesis $H_0 : A_{xy}(L) = 0$ and/or $H_0 : A_{yx}(L) = 0$ to identify strict causality relation using the LR test as discussed in Section 2.1.

2.3 Multivariate Nonlinear Causality Test

In this section, we will extend the nonlinear causality test for the bivariate setting developed by Hiemstra and Jones (1994) to the multivariate setting. To identify any nonlinear Granger causality relationship from any two series, say $\{x_t\}$ and $\{y_t\}$ in the bivariate setting, one has to first apply the linear model to $\{x_t\}$ and $\{y_t\}$ to identify their linear causal relationships and obtain the corresponding residuals, $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$. Thereafter, one has to apply a nonlinear Granger causality test to the residual series, $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$, of the two variables being examined to identify the remaining nonlinear causal relationships between their residuals. This is also true if one would like to identify existence of any nonlinear Granger causality relation between two vectors of time series, say $x_t = (x_{1,t}, \cdots, x_{n_1,t})'$ and $y_t = (y_{1,t}, \cdots, y_{n_2,t})'$ in the multivariate setting. One has to apply the VAR model

in (1) or the ECM-VAR model in (2) to the series to identify their linear causal relationships and obtain their corresponding residuals. Thereafter, one has to apply a nonlinear Granger causality test to the residual series. For simplicity, in this section we denote $X_t = (X_{1,t}, \cdots, X_{n_1,t})'$ and $Y_t = (Y_{1,t}, \cdots, Y_{n_2,t})'$ to be the corresponding residuals of any two vectors of variables being examined.

We first define the lead vector and lag vector of a time series, say $X_{i,t}$, as follows: for $X_{i,t}, i = 1, \cdots, n_1$, the m_{x_i}-length lead vector and the L_{x_i}-length lag vector of $X_{i,t}$ are:

$$X_{i,t}^{m_{x_i}} \equiv (X_{i,t}, X_{i,t+1}, \cdots, X_{i,\ t+m_{x_i}-1}), m_{x_i} = 1, 2, \cdots, t = 1, 2, \cdots,$$

$$X_{i,\ t-L_{x_i}}^{L_{x_i}} \equiv (X_{i,\ t-L_{x_i}}, X_{i,\ t-L_{x_i}+1}, \cdots, X_{i,\ t-1}), L_{x_i} = 1, 2, \cdots, t = L_{x_i} + 1, L_{x_i} + 2, \cdots,$$

respectively. We denote $M_x = (m_{x_1}, \cdots, m_{x_{n_1}})$, $L_x = (L_{x_1}, \cdots, L_{x_{n_1}})$, $m_x = max(m_{x_1}, \cdots, m_{x_{n_1}})$, and $l_x = max(L_{x_1}, \cdots, L_{x_{n_1}})$. The m_{y_i}-length lead vector, $Y_{i,t}^{m_{y_i}}$, the Ly_i-length lag vector, $Y_{i,t-L_{y_i}}^{L_{y_i}}$, of $Y_{i,\ t}$, and M_y, L_y, m_y, and l_y can be defined similarly.

Given m_x, m_y, L_x, L_y, and $e > 0$, we define the following four events:

(1) $\left\{\|X_t^{M_x} - X_s^{M_x}\| < e\right\} \equiv \left\{\|X_{i,t}^{M_{x_1}} - X_{i,s}^{m_{x_1}}\| < e, \text{ for any } i = 1, \cdots, n_1\right\}$;

(2) $\left\{\|X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x}\| < e\right\} \equiv \left\{\|X_{i,t-L_{x_i}}^{L_{x_i}} - X_{i,s-L_{x_i}}^{L_{x_i}}\| < e, \text{ for any } i = 1, \cdots, n_1\right\}$;

(3) $\left\{\|Y_t^{M_y} - Y_s^{M_y}\| < e\right\} \equiv \left\{\|Y_{i,t}^{m_{y_i}} - Y_{i,s}^{m_{y_i}}\| < e, \text{ for any } i = 1, \cdots, n_2\right\}$; and

(4) $\left\{\|Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y}\| < e\right\} \equiv \left\{\|Y_{i,t-L_{y_i}}^{L_{y_i}} - Y_{i,s-L_{y_i}}^{L_{y_i}}\| < e, \text{ for any } i = 1, \cdots, n_2\right\}$,

where $\| \cdot \|$ denotes the maximum norm which is defined as $\|X - Y\| = max(|x_1 - y_1|, |x_2 - y_2|, \cdots, |x_n - y_n|)$ for any two vectors $X = (x_1, \cdots, x_n)$ and $Y = (y_1, \cdots, y_n)$. The vector series $\{Y_t\}$ is said not to strictly Granger cause another vector series $\{X_t\}$ if

$$Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e \,\big|\, \|X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x}\| < e, \|Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y}\| < e,\right)$$

$$= Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e, \big|\, \|X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x}\| < e\right),$$

where $Pr(\cdot|\cdot)$ denotes conditional probability.

The test statistic for testing non-existence of nonlinear Granger causality can be obtained as follows:

$$\sqrt{n}\left(\frac{C_1\big(M_x + L_x, L_y, e, n\big)}{C_2\big(L_x, L_y, e, n\big)} - \frac{C_3\big(M_x + L_x, e, n\big)}{C_4\big(L_x, e, n\big)}\right) \tag{3}$$

where

$$C_1\big(M_x + L_x, L_y, e, n\big) \equiv \frac{2}{n(n-1)}\sum_{t<s}\sum\prod_{i=1}^{n_1} I\big(x_{i,t-L_{x_i}}^{m_{x_i}+L_{x_i}}, x_{i,s-L_x}^{m_{x_i}+L_{x_i}}, e\big) \cdot \prod_{i=1}^{n_2} I\big(y_{i,t-L_{y_i}}^{L_{y_i}}, y_{i,s-L_{y_i}}^{L_{y_i}}, e\big),$$

$$C_2\big(L_x, L_y, e, n\big) \equiv \frac{2}{n(n-1)}\sum_{t<s}\sum\prod_{i=1}^{n_1} I\big(x_{i,t-L_{x_i}}^{L_{x_i}}, x_{i,s-L_x}^{L_{x_i}}, e\big) \cdot \prod_{i=1}^{n_2} I\big(y_{i,t-L_{y_i}}^{L_{y_i}}, y_{i,s-L_y}^{L_{y_i}}, e\big),$$

$$C_3\big(M_x + L_x, e, n\big) \equiv \frac{2}{n(n-1)}\sum_{t<s}\sum\prod_{i=1}^{n_1} I\big(x_{i,t-L_x}^{m_{x_i}+L_{x_i}}, x_{i,s-L_{x_i}}^{m_{x_i}+L_{x_i}}, e\big),$$

$$C_4\big(L_x, e, n\big) \equiv \frac{2}{n(n-1)}\sum_{t<s}\sum\prod_{i=1}^{n_1} I\big(x_{i,t-L_{x_i}}^{L_{x_i}}, x_{i,s-L_{x_i}}^{L_{x_i}}, e\big),$$

$$I(x, y, e) = \begin{cases} 0, & \text{if } \|x-y\| > e \\ 1, & \text{if } \|x-y\| \le e \end{cases}, \quad \text{and}$$

$t, s = max\big(L_x, L_y\big) + 1, \cdots, T - m_x + 1, n = T + 1 - m_x - max\big(L_x, L_y\big).$

In this paper, we develop the following theorem:

Theorem 2.1. *To test the null hypothesis, H_0, that $\{Y_{1,t}, \cdots, Y_{n_2,t}\}$ does not strictly Granger cause $\{X_{1,t}, \cdots, X_{n_1,t}\}$, under the assumptions that the time series $\{X_{1,t}, \cdots, X_{n_1,t}\}$ and $\{Y_{1,t}, \cdots, Y_{n_2,t}\}$ are strictly stationary, weakly dependent, and satisfy the mixing conditions stated in Denker and Keller (1983), if the null hypothesis, H_0, is true, the test statistic defined in (3) is distributed as $N\big(0, \sigma^2(M_x, L_x, L_y, e)\big)$. When the test statistic in (3) is too far away from zero, we reject the null hypothesis. A consistent estimator of $\sigma^2(M_x, L_x, L_y, e)$ is $\hat{\sigma}^2(M_x, L_x, L_y, e) = \widehat{\nabla f(\theta)}^T \cdot \hat{\Sigma} \cdot \widehat{\nabla f(\theta)}$ in which each component $\Sigma_{i,j}$ $(i, j = 1, \cdots, 4)$, of the covariance matrix Σ is given by:*

$$\Sigma_{i,j} = 4 \cdot \sum_{k \ge 1} \omega_k E(A_{i,t} \cdot A_{j,t+k-1}) \quad,$$

5

$$
\omega_k = \begin{cases} 1 & \text{if } k = 1 \\ 2, & \text{otherwise} \end{cases},
$$

$$
A_{1,t} = h_{11}\left(x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}, e\right) - C_1(M_x + L_x, L_y, e),
$$

$$
A_{2,t} = h_{12}\left(x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}, e\right) - C_2(L_x, L_y, e),
$$

$$
A_{3,t} = h_{13}\left(x_{t-L_x}^{M_x+L_x}, e\right) - C_3(M_x + L_x, e), \quad \text{and}
$$

$$
A_{4,t} = h_{14}\left(x_{t-L_x}^{L_x}, e\right) - C_4(L_x, e),
$$

where $h_{1i}(\mathbf{z_t})$, $i = 1, \cdots, 4$, is the conditional expectation of $h_i(\mathbf{z_t}, \mathbf{z_s})$ given the value of $\mathbf{z_t}$ as follows:

$$
h_{11}\left(x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}, e\right) = E\left(h_1 \,\middle|\, x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}\right),
$$

$$
h_{12}\left(x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}, e\right) = E\left(h_2 \,\middle|\, x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}\right),
$$

$$
h_{13}\left(x_{t-L_x}^{M_x+L_x}, e\right) = E\left(h_3 \,\middle|\, x_{t-L_x}^{M_x+L_x}\right), \quad \text{and} \quad h_{14}\left(x_{t-L_x}^{L_x}, e\right) = E\left(h_4 \,\middle|\, x_{t-L_x}^{L_x}\right).
$$

A consistent estimator of $\Sigma_{i,j}$ elements is given by:

$$
\hat{\Sigma}_{i,j} = 4 \cdot \sum_{k=1}^{K(n)} \omega_k(n) \left[\frac{1}{2(n-k+1)} \sum_t \left(\hat{A}_{i,t}(n) \cdot \hat{A}_{j,t-k+1}(n) + \hat{A}_{i,t-k+1}(n) \cdot \hat{A}_{j,t}(n) \right) \right],
$$

$$
K(n) = [n^{1/4}], \quad \omega_k(n) = \begin{cases} 1 & \text{if } k = 1 \\ 2(1 - [(k-1)/K(n)]) & \text{otherwise} \end{cases},
$$

in which $\hat{A}_{i,t}$ is defined in the appendix for $i = 1, 2, 3, 4$ and a consistent estimator of $\nabla f(\theta)$ is:

$$
\widehat{\nabla f(\theta)} = \left[\frac{1}{\hat{\theta}_2}, -\frac{\hat{\theta}_1}{\hat{\theta}_2^2}, -\frac{1}{\hat{\theta}_4}, \frac{\hat{\theta}_3}{\hat{\theta}_4^2} \right]^T
$$

$$
= \left[\frac{1}{C_2\left(L_x, L_y, e, n\right)}, -\frac{C_1\left(m + L_x, L_y, e, n\right)}{C_2^2\left(L_x, L_y, e, n\right)}, -\frac{1}{C_4\left(L_x, e, n\right)}, \frac{C_3\left(M_x + L_x, e, n\right)}{C_4^2\left(L_x, e, n\right)} \right]^T.
$$

The prove is given in the appendix.

3   Monte Carlo Simulation

In this section, we present the Monte Carlo simulation to demonstrate the superiority of our proposed multivariate nonlinear Granger causality test over its bivariate counterpart in the performance of both size and power when the underlying series possess multivariate nonlinear Granger causality nature.

We have conducted simulations for a variety of time series possessing different multivariate nonlinear Granger causality relationships. All simulations show that our proposed multivariate nonlinear Granger causality test performs better in both size and power. For simplicity, we only present the results of the following equation:

$$X_t = \beta \, Y_{t-1} Z_{t-1} + \epsilon_t \tag{4}$$

where $\{Y_t\}$ and $\{Z_t\}$ are i.i.d. and mutually independent random variables generated from $N(0,1)$, $\{\epsilon_t\}$ is Gaussian white noise generated from $N(0,0.1)$. Under the model in (4), the variables $\{Y_t, Z_t\}$ nonlinear Granger cause $\{X_t\}$ if $\beta \neq 0$ and there is no Granger causality relationship if $\beta = 0$. The bivariate nonlinear Granger causality test could detect the bivariate causality relationships well but it may not be able to examine the causality relationships under multivariate settings including the one set in (4). Thus, we expect that our proposed multivariate test could perform better than its bivariate counterpart in this model setting. To justify our claim, we conduct a simulation with 1,000 Monte Carlo runs based on sample size of 50 and 100 observations for each β value. We set lead length $m = 1$ and the common lag length $L_x = L_y = L_z$ for all the cases being examined. A common scale parameter of $e = 1.5\sigma$ is used where $\sigma = 1$ denotes the standard deviation of standardized series. In the simulation of each replication, the values of the test statistics for different common lag lengths are compared with their asymptotic critical values at the 0.05 nominal significance level. The percentage of rejecting the null of $\beta = 0$ is reported in Tables 1 and 2 for sample size of 50 and 100, respectively.

Table 1 displays the simulation results of sample size 50 with the value of β varying from -0.5 to 0.5 and the common lag length being 1, 2, and 3 for all the cases under examination. When $\beta = 0$, that is, $\{X_t\}$ is independent of both $\{Y_t\}$ and $\{Z_t\}$ implying

that the null hypothesis is true, both bivariate and multivariate tests are conservative when the common lag length $= 1, 2$. When the common lag length is equal to 3, both tests have empirical sizes similar to the nominated significance level of 0.05. In short, Table 1 exhibits that (a) the average of the simulated size of multivariate test is closer to the nominated significance level of 0.05. When β is nonzero, our simulation shows that (b) the powers of both bivariate and multivariate tests perform better when lag length $= 1$ and their powers reduce when lag length increases, and (c) the power of our proposed multivariate test is much higher than that of its bivariate counterpart for any lag length being examined in our paper. We note that (b) is reasonable because only lag one of both Y and Z "cause" X in (4) while our findings in (a) and (c) show that our proposed multivariate test performs better than its bivariate counterpart in both size and power.

We turn to examine both size and power when we increase the sample size to 100. The results are displayed in Table 2. As the sample size is larger, we report results with longer lag length scale including $L_x = L_y = L_z = 1, 3$, and 5.[1] Comparing with sample size $= 50$, the simulation results show that our observations of (a), (b), and (c) for sample size $= 50$ still hold for sample size of 100, but, as expected, both size and power for both bivariate and multivariate tests have improved and our findings are consistent to show that our proposed multivariate test performs better than its bivariate counterpart in both size and power when sample size $= 100$.

4 Illustration

Qiao, et al (2008) have examined the bivariate linear and nonlinear Granger causality relationships between pairs of daily returns from five indices: (a) Shanghai A shares (SHA) and Shanghai B shares (SHB) from Shanghai Stock Exchange (SHSE), (b) Shenzhen A shares (SZA) and Shenzhen B shares (SZB) from Shenzhen Stock Exchange (SZSE), and (c) Hong Kong H shares (H) before and after February 19, 2001, the date the Chinese Government allowed domestic citizens to trade B shares from this date onwards[2].

---

[1]We have examined other lag lengths and the results are consistent with our findings. Thus, we skip reporting other lag lengths to save space.

[2]Readers may refer to Qiao, et al (2008) for detailed information on SHA, SZA, SHB, SZB, and H.

Table 1: Size and power comparison between bivariate and multivariate nonlinear Granger causality tests when sample size=50

Beta	-0.5	-0.4	-0.3	-0.2	-0.1	-0.05	0	0.05	0.1	0.2	0.3	0.4	0.5
Lags=1													
Bivariate	0.917	0.875	0.833	0.775	0.413	0.114	0.016	0.129	0.384	0.748	0.851	0.894	0.902
Multivariate	0.996	1.000	0.997	0.987	0.716	0.204	0.029	0.220	0.686	0.981	1.000	0.999	0.999
Lags=2													
Bivariate	0.645	0.655	0.624	0.516	0.240	0.100	0.034	0.126	0.255	0.509	0.598	0.647	0.644
Multivariate	0.983	0.979	0.956	0.870	0.427	0.144	0.032	0.165	0.468	0.881	0.956	0.979	0.982
Lags=3													
Bivariate	0.465	0.452	0.444	0.333	0.178	0.072	0.049	0.073	0.162	0.355	0.422	0.460	0.433
Multivariate	0.864	0.850	0.796	0.676	0.293	0.115	0.045	0.106	0.310	0.655	0.789	0.842	0.853
Bivariate	0.676	0.661	0.634	0.541	0.277	0.095	0.033	0.109	0.267	0.537	0.624	0.667	0.660
Multivariate	0.948	0.943	0.916	0.844	0.479	0.154	0.035	0.164	0.488	0.839	0.915	0.940	0.945

Note: The last two rows display the average values for the case of "Bivariate" and "Multivariate", respectively, for the corresponding value of β.

Table 2: Size and power comparison between bivariate and multivariate nonlinear Granger causality tests when sample size=100

Beta	-0.5	-0.4	-0.3	-0.2	-0.1	-0.05	0	0.05	0.1	0.2	0.3	0.4	0.5
Lags=1													
Bivariate	1.000	1.000	0.996	0.989	0.781	0.266	0.029	0.271	0.807	0.994	0.999	0.997	0.999
Multivariate	1.000	1.000	1.000	1.000	0.971	0.448	0.037	0.452	0.977	1.000	1.000	1.000	1.000
Lags=3													
Bivariate	0.872	0.842	0.823	0.740	0.394	0.122	0.046	0.107	0.406	0.700	0.812	0.836	0.846
Multivariate	1.000	0.999	0.993	0.973	0.636	0.145	0.048	0.169	0.604	0.963	0.990	0.997	1.000
Lags=5													
Bivariate	0.561	0.554	0.508	0.453	0.189	0.075	0.048	0.072	0.210	0.412	0.525	0.577	0.532
Multivariate	0.865	0.823	0.776	0.676	0.256	0.105	0.049	0.104	0.264	0.645	0.788	0.842	0.870
Bivariate	0.811	0.799	0.776	0.727	0.455	0.154	0.041	0.150	0.474	0.702	0.779	0.803	0.792
Multivariate	0.955	0.941	0.923	0.883	0.621	0.233	0.045	0.242	0.615	0.869	0.926	0.946	0.957

Note: The last two rows display the average values for the case of "Bivariate" and "Multivariate", respectively, for the corresponding value of β.

As they only apply the bivariate Granger causality test to study the issue, their results may not be able to capture any multivariate causality relationship among these indices. To bridge the gap, in this paper we first apply the traditional multivariate linear Granger causality test and thereafter apply our proposed multivariate nonlinear causality test to examine the existence of multivariate linear and nonlinear causality relationships in any of the following three groups for the segmented Chinese stock markets before and after February 19, 2001: (a) A-share (including SHA and SZA) and B-share (including SHB and SZB), (b) Shanghai stock market (including SHA and SHB) and Shenzhen stock market(including SZA and SZB), and (c) domestic stock market (A-share) and foreigner-invested stock market (including B-share and H-share). The studying period is from October 6, 1992 to December 31, 2007 and all data are taken from DataStream International. For easy comparison, we follow Qiao, et al (2008) to use February 19, 2001 as the cut-off point [3] so that the first sub-period is from October 6, 1992 to February 16, 2001 and the second sub-period is from February 19, 2001 to December 31, 2007.

We first adopt the VAR model in (1) or the ECM-VAR model in (2) to examine whether there is any multivariate linear Granger causality relationship among the indices in any of the groups mentioned above. We find that, in the first subperiod, there is only strong unidirectional linear causality from Shenzhen stock market to Shanghai stock market. On the other hand, in the second sub-period, we find that (a) there is unidirectional linear causality from B-share to A-share, and (b) there are strong feedback causality relationships between Shanghai stock market and Shenzhen stock market, and between domestic stock market and foreigner-invested stock market[4].

After checking the multivariate linear causalities, we apply our new proposed multivariate nonlinear Granger causality test to the error terms from the estimated VAR or ECM-VAR models to investigate whether there is any remaining undetected multivariate nonlinear relationship among the indices. We set the common lead length $m = 1$ and the common lag length to be 1 to 10. A common scale parameter of $e = 1.5\sigma$ is used,

---

[3]We choose February 19, 2001 as the cut-off point because as of that date, the Chinese government adopted a new policy that removed the restrictions on trading B shares by domestic citizens.

[4]The detailed results of the multivariate linear causality test are available upon request.

where $\sigma = 1$ denotes the standard deviation of standardized series. The values of the standardized test statistic of (3) are reported in Table 3. Under the null hypothesis of no multivariate nonlinear Granger causality, the test statistic is asymptotically distributed as $N(0,1)$. Therefore, very large or small values of the test statistic lead to the rejection of null hypothesis, or equivalently, indicate the existence of multivariate nonlinear causality.

The results show that, in the first subperiod, (a) there is weak nonlinear causality from B-share to A-share, and (b) there is strong bi-directional nonlinear causality between Shenzhen and Shanghai stock markets, since the test statistics of SZ→SH are significant at the 5% level for all lags, and the test statistics of SH→SZ are significant at the 5% level for the first half of lags. In the second sub-period, our results infer that (c) there exists strong feedback nonlinear causality between Shenzhen and Shanghai stock markets similar to the results of the first subperiod, (d) there is strong unidirectional nonlinear causality from A-share to B-share, since the test statistics of A→B are significant at the 5% level for some lags, and (e) there exists weak nonlinear causality from domestic stock market to foreigner-invested stock market, as there are two test statistics to be significant. Our findings infer that there are more nonlinear causality among the indices after the Chinese Government introduced the policy on February 19, 2001.

Appendix: Proof of Theorem 2.1

Before we prove the theorem, we first introduce the U-statistic (Kowalski and Tu, 2007) in the following definition, which is essential in the proof.

Definition 4.1. *Consider an i.i.d. sample of $p \times 1$ column vector of response $\mathbf{y_i}$ ($1 \leq i \leq n$). Let $\mathbf{h}(\mathbf{y}_1, \cdots, \mathbf{y}_m)$ be a symmetric vector-valued function with m arguments. A **one-sample, m-argument multivariate U-statistic vector** with **kernel vector h** is defined as:*

$$\mathbf{U_n} = \begin{pmatrix} n \\ m \end{pmatrix}^{-1} \sum_{(i_1, \cdots, i_m) \in C_m^n} \mathbf{h}(\mathbf{y}_{i_1}, \cdots, \mathbf{y}_{i_m}),$$

Table 3: Multiple Nonlinear Testing Results for China's Stock Markets

	First Sub-Period			Second Sub-Period		
lags	B→A	SZ→SH	B,H→A	B→A	SZ→SH	B,H→A
1	1.620*	4.499***	0.617	0.358	1.434*	1.214
2	1.413*	4.882***	0.344	0.947	3.361***	1.019
3	0.512	4.610***	-0.591	0.376	4.085***	0.609
4	0.546	4.179***	-0.418	-0.541	3.638***	-0.178
5	-0.187	3.971***	-0.800	-1.217	3.359***	-1.305
6	-0.704	3.523***	-1.024	-1.018	3.031***	-1.158
7	-1.663	2.953***	-1.465	-0.961	2.846***	-0.625
8	-1.858	2.863***	-1.249	-1.191	1.990**	-0.879
9	-1.859	2.375***	-1.058	-1.331	2.084**	-1.426
10	-2.186	1.887**	-1.054	-1.283	1.733**	-0.778
lags	A→B	SH→SZ	A→B,H	A→B	SH→SZ	A→B,H
1	-2.130	3.071***	-1.638	2.098**	2.611***	2.309**
2	-2.423	2.983***	-1.171	2.311**	2.792***	2.708***
3	-2.052	1.896**	-1.049	2.282**	2.599***	1.232
4	-1.901	2.381***	-1.263	1.673**	1.827**	0.224
5	-1.104	1.754**	-1.428	1.115	1.474*	-0.600
6	-1.201	1.505*	-1.461	1.649**	1.308*	-1.128
7	-1.325	1.091	-0.943	1.569*	0.907	-0.766
8	-0.961	0.793	-1.204	1.329*	0.933	-0.943
9	-1.177	0.843	-1.165	1.122	0.726	-0.774
10	-1.371	0.294	-0.636	1.222	0.913	-0.597

Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively. A includes SHA and SZA, B includes SHB and SZB, SH includes SHA and SHB, and SZ includes SZA and SZB. The first sub-period is from October 6, 1992 to February 16, 2001 while the second sub-period is from February 19, 2001 to December 31, 2007.

where $C_m^n = (i_1, \cdots, i_m), 1 \le i_1 < \cdots < i_m \le n$, denotes the set of all distinct combinations of m indices $(i_1, \cdots, i_m)$ from the integer set $1, 2, \cdots, n$.

Let $\Theta = E\big(\mathbf{h}(\mathbf{y}_1, \cdots, \mathbf{y}_m)\big)$. We have

$$
\begin{aligned}
E\big(\mathbf{h}\big) &= E\left(\binom{n}{m}^{-1} \sum_{(i_1, \cdots, i_m) \in C_m^n} \mathbf{h}\big(\mathbf{y}_{i_1}, \cdots, \mathbf{y}_{i_m}\big) \right) \\
&= \binom{n}{m}^{-1} \sum_{(i_1, \cdots, i_m) \in C_m^n} E\big(\mathbf{h}(\mathbf{y}_1, \cdots, \mathbf{y}_m)\big) = \Theta \,.
\end{aligned}
$$

Now, we proceed on to prove Theorem 2.1. We denote that

$$
C_1\big(M_x + L_x, L_y, e\big) \equiv Pr\left(\| X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \| < e, \| Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \| < e \right),
$$
$$
C_2\big(L_x, L_y, e\big) \equiv Pr\left(\| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e, \| Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \| < e \right),
$$
$$
C_3\big(M_x + L_x, e\big) \equiv Pr\left(\| X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \| < e \right), \text{ and}
$$
$$
C_4\big(L_x, e\big) \equiv Pr\left(\| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e \right).
$$

Then, we have

$$
\begin{aligned}
&Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e \big| \, \| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e, \| Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \| < e \right) \\
&= \frac{C_1\big(M_x + L_x, L_y, e\big)}{C_2\big(L_x, L_y, e\big)}, \\
&Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e \big| \, \| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e \right) = \frac{C_3\big(M_x + L_x, e\big)}{C_4\big(L_x, e\big)},
\end{aligned}
$$

and the strict Granger noncausality condition can be stated as

$$
H_0: \quad \frac{C_1\big(M_x + L_x, L_y, e\big)}{C_2\big(L_x, L_y, e\big)} - \frac{C_3\big(M_x + L_x, e\big)}{C_4\big(L_x, e\big)} = 0 \,.
$$

Instead of analyzing i.i.d. samples, we take vector samples from strictly stationary, weakly

dependent series which satisfy the mixing conditions of Denker and Keller(1983) such that

$$\mathbf{z_t} = \left(X_{1,t-L_{x_1}}^{M_{x_1}+L_{x_1}}, \cdots, X_{n_1,t-L_{x n_1}}^{M_{x_1}+L_{x_1}}, Y_{1,t-L_{y_1}}^{L_{y_1}}, \cdots, Y_{n_2,t-L_{y n_2}}^{L_{y n_2}} \right)$$

$$t = max\left(L_x, L_y\right), \cdots, n; \ n = T - m_x - max\left(L_x, L_y\right) + 1$$

contains $\sum_{i=1}^{n_1} m_{x_1} + \sum_{i=1}^{n_1} L_{x_i} + \sum_{i=1}^{n_2} L_{y_i}$ variables in each vector. For any given $\left(M_x, L_x, L_y, e\right)$, we denote $\Theta = \left(\theta_1, \ \theta_2, \ \theta_3, \ \theta_4\right)'$ where $\theta_1 \equiv C_1\left(M_x + L_x, L_y, e\right)$, $\theta_2 \equiv C_2\left(L_x, L_y, e\right)$, $\theta_3 \equiv C_3\left(M_x+L_x, e\right)$, and $\theta_4 \equiv C_4\left(L_x, e\right)$. We denote $\mathbf{U_n} = \left(U_{1n}, U_{2n}, U_{3n}, U_{4n}\right)'$ where $U_{1n} \equiv C_1\left(M_x + L_x, L_y, e, n\right)$, $U_{2n} \equiv C_2\left(L_x, L_y, e, n\right)$, $U_{3n} \equiv C_3\left(M_x + L_x, e, n\right)$, and $U_{4n} \equiv C_4\left(L_x, e, n\right)$. One could easily show that $\mathbf{U_n}$ is a one-sample, 2-argument 4-variable U-statistic vector with kernel vector $\mathbf{h}\left(\mathbf{z_t}, \mathbf{z_s}\right) = \left(h_1, h_2, h_3, h_4\right)'$, where

$$h_1 \equiv \prod_{i=1}^{n_1} I\left(X_{i,t-L_{x_i}}^{m_{x_i}+L_{x_i}}, X_{i,s-L_x}^{m_{x_i}+L_{x_i}}, e\right) \cdot \prod_{i=1}^{n_2} I\left(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_{y_i}}^{L_{y_i}}, e\right),$$

$$h_2 \equiv \prod_{i=1}^{n_1} I\left(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_x}^{L_{x_i}}, e\right) \cdot \prod_{i=1}^{n_2} I\left(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_y}^{L_{y_i}}, e\right),$$

$$h_3 \equiv \prod_{i=1}^{n_1} I\left(X_{i,t-L_x}^{m_{x_i}+L_{x_i}}, X_{i,s-L_{x_i}}^{m_{x_i}+L_{x_i}}, e\right), \quad h_4 \equiv \prod_{i=1}^{n_1} I\left(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_{x_i}}^{L_{x_i}}, e\right),$$

$$\mathbf{U_n} = \binom{n}{2}^{-1} \sum_{1 \leq t < s \leq n} \mathbf{h}\left(\mathbf{z_t}, \mathbf{z_s}\right).$$

Since function $I\left(x, y, e\right)$ is symmetric with respect to x and y, $\mathbf{h}\left(\mathbf{z_t}, \mathbf{z_s}\right)$ is symmetric with two arguments $\mathbf{z_t}$ and $\mathbf{z_s}$. In addition, we notice that

$$E\left(h_1\right) = Pr\left(\parallel X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \parallel < e\right) \cdot Pr\left(\parallel Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \parallel < e\right)$$

$$= Pr\left(\parallel X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \parallel < e, \ \parallel Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \parallel < e\right) = \theta_1$$

since residual series $\{X_t\}$ and $\{Y_t\}$ are assumed to be obtained from the VAR model are independent with each other. Similarly, we have $E\left(h_i\right) = \theta_i$, $i = 2, 3, 4$.

14

In the bivariate case, the sample vector is

$$\mathbf{z_t} = \left(X_{t-L_x}^{m_x+L_x}, \; Y_{t-L_y}^{L_y} \right) \quad t = 1, \cdots, n; n = T - m_x - max(L_x, L_y) + 1 \,,$$

in which it contains $m_x + L_x + L_y$ variables.

This change does not affect the asymptotic properties of the U-statistic vector used in the test. As in the bivariate case, the central limit theorem proved by Denker and Keller(1983) can be applied to the U-statistic vector $\mathbf{U_n}$; that is, under the assumption that series $\{x_{1,t}, \cdots, x_{n_1,t}, y_{1,t}, \cdots, y_{n_2,t}\}$ are strictly stationary, weakly dependent, and satisfying one of the mixing conditions of Denker and Keller, we have:

$$\sqrt{n}(\mathbf{U_n} - \Theta) \xrightarrow{d} N(0, \Sigma), \text{ as } n \to \infty \,,$$

where $\xrightarrow{d}$ means convergence in distribution, and Σ is the 4×4 covariance matrix of $\mathbf{U_n}$ containing $\{\Sigma_{i,j}, i, j = 1, \cdots, 4\}$. Furthermore, by Denker and Keller (1986), the sequence $\mathbf{U_n}$ of the U-statistics converges to Θ in probability. Now, we present our proposed test statistic as a function of $\mathbf{U_n}$ such that

$$\sqrt{n}f(\mathbf{U_n}) = \sqrt{n} \left(\frac{U_{1n}}{U_{2n}} - \frac{U_{3n}}{U_{4n}} \right) \,.$$

Under the null hypothesis that $\{Y_t\}$ does not strictly Granger cause $\{X_t\}$, we have $f(\Theta) = \theta_1/\theta_2 - \theta_3/\theta_4 = 0$. Thus, using the delta method (Serfling, 1980, pp.122-125), $\sqrt{n}[f(\mathbf{U_n}) - f(\Theta)]$ has the same limit distribution as $\sqrt{n}[\nabla f(\Theta)^T \cdot (\mathbf{U_n} - \Theta)]$, and hence, we have

$$\sqrt{n}[f(\mathbf{U_n}) - f(\Theta)] = \sqrt{n} \left(\frac{C_1(M_x + L_x, L_y, e, n)}{C_2(L_x, L_y, e, n)} - \frac{C_3(M_x + L_x, e, n)}{C_4(L_x, e, n)} \right)$$
$$\sim N \left(0, \nabla f(\Theta)^T \cdot \Sigma \cdot \nabla f(\Theta) \right) \,,$$

where $\nabla f(\Theta)$ is the derivative of f evaluated at Θ such that

$$\nabla f(\Theta) = \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \frac{\partial f}{\partial \theta_3}, \frac{\partial f}{\partial \theta_4} \right)^T = \left(\frac{1}{\theta_2}, \; \frac{-\theta_1}{\theta_2^2}, \; \frac{-1}{\theta_4}, \; \frac{\theta_3}{\theta_4^2} \right)^T \,.$$

Now, we turn to show that a consistent estimator of $\sigma^2(M_x, L_x, L_y, e)$ becomes

$$\hat{\sigma}^2(M_x, L_x, L_y, e) = \widehat{\nabla f(\theta)}^T \cdot \hat{\Sigma} \cdot \widehat{\nabla f(\theta)}\,.$$

Applying the results of Denker and Keller (1983) and Newey and West (1987), we obtain a consistent estimator of $\Sigma_{i,j}$ elements to be:

$$\hat{\Sigma}_{i,j} = 4 \cdot \sum_{k=1}^{K(n)} \omega_k(n) \left[\frac{1}{2(n-k+1)} \sum_t \left(\hat{A}_{i,t}(n) \cdot \hat{A}_{j,t-k+1}(n) + \hat{A}_{i,t-k+1}(n) \cdot \hat{A}_{j,t}(n) \right) \right],$$

$$K(n) = [n^{1/4}], \; \omega_k(n) = \begin{cases} 1 & \text{if } k = 1 \\ 2(1 - [(k-1)/K(n)]) & \text{otherwise} \end{cases},$$

$$\hat{A}_{1,t} = \frac{1}{n-1} \left(\sum_{s \neq t} \prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{m_{x_i}+L_{x_i}}, X_{i,s-L_x}^{m_{x_i}+L_{x_i}}, e) \cdot \prod_{i=1}^{n_2} I(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_{y_i}}^{L_{y_i}}, e) \right)$$
$$- C_1(M_x + L_x, L_y, e, n),$$

$$\hat{A}_{2,t} = \frac{1}{n-1} \left(\sum_{s \neq t} \prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_x}^{L_{x_i}}, e) \cdot \prod_{i=1}^{n_2} I(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_y}^{L_{y_i}}, e) \right)$$
$$- C_2(L_x, L_y, e, n),$$

$$\hat{A}_{3,t} = \frac{1}{n-1} \left(\sum_{s \neq t} \prod_{i=1}^{n_1} I(X_{i,t-L_x}^{m_{x_i}+L_{x_i}}, X_{i,s-L_{x_i}}^{m_{x_i}+L_{x_i}}, e) \right) - C_3(m + L_x, e, n),$$

$$\hat{A}_{4,t} = \frac{1}{n-1} \left(\sum_{s \neq t} \prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_{x_i}}^{L_{x_i}}, e) \right) - C_4(L_x, e, n),$$

$$t, s = max(L_x, L_y), \cdots, n \text{ and } n = T - m_x - max(L_x + L_y) + 1,$$

and a consistent estimator of $\nabla f(\theta)$ is:

$$\widehat{\nabla f(\theta)} = \left[1/\hat{\theta}_2, \; -\hat{\theta}_1/\hat{\theta}_2^2, \; -1/\hat{\theta}_4, \; \hat{\theta}_3/\hat{\theta}_4^2 \right]^T$$
$$= [1/C_2(L_x, L_y, e, n), -C_1(m + L_x, L_y, e, n)/C_2^2(L_x, L_y, e, n),$$
$$- 1/C_4(L_x, e, n), C_3(M_x + L_x, e, n)/C_4^2(L_x, e, n)]^T$$

Thus, a consistent estimator of $\sigma^2(M_x, L_x, L_y, e)$ is:

$$\hat{\sigma}^2(M_x, L_x, L_y, e) = \widehat{\nabla f(\theta)}^T \cdot \hat{\Sigma} \cdot \widehat{\nabla f(\theta)}$$

and the assertion of the theorem follows.

$\square$

References

[1] Baek, E.G., Brock, W.A. (1992), A General Test for Nonlinear Granger Causality : Bivariate Model, working paper, Korea Development Institute, University of Wisconsin-Madison.

[2] Chiang, T.C., Qiao, Z., Wong, W.K. (2009), New Evidence on the Relation between Return Volatility and Trading Volume, *Journal of Forecasting* 29(5), 502 - 515.

[3] Denker, M., Keller, G. (1983), On U-statistics and v. Mises' Statistics for weakly Dependent Processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 505-522.

[4] Granger, C.W.J. (1969), Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37(3), 424-438.

[5] Hiemstra, C., Jones, J.D. (1994), Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation, *Journal of Finance* 49(5), 1639-1664.

[6] Kowalski, J., Tu, X.M. (2007), *Modern Applied U-Statistics*, John Wiley & Sons, New York.

[7] Newey, W.K., West, K.D. (1987), A simple, positive semi-definite, heteroskedasticity and autocorrelation constistent covariance matrix, *Econometrica* 55(3) (1987) 703-708.

[8] Qiao, Z., Li, Y., Wong, W.K. (2008), Policy Change and Lead-lag relations among China's Segmented Stock Markets, *Journal of Multinational Financial Management* 18, 276-289.

[9] Qiao, Z., McAleer, M., Wong, W.K. (2009), Linear and Nonlinear Causality of Consumption Growth and Consumer Attitudes, *Economics Letters* 102(3), 161-164.

[10] Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

[11] Sims, C.A. (1980), Macroeconomics and Reality, *Econometrica* 48, 1-48.