2010

Multivariate linear and nonlinear causality tests

Zhidong Bai

Wing Keung Wong
Hong Kong Baptist University, awong@hkbu.edu.hk

Bingzhi Zhang

Citation
Bai, Zhidong, Wing Keung Wong, and Bingzhi Zhang. "Multivariate linear and nonlinear causality tests." *Mathematics and Computers in Simulation* 81.1 (2010): 5-17.

Multivariate Linear and Non-Linear Causality Tests

Zhidong Bai

School of Mathematics and Statistics, Northeast Normal University

Department of Statistics and Applied Probability,

National University of Singapore

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

Acknowledgments

The second author would like to thank Professors Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. This research is partially supported by Northeast Normal University, the National University of Singapore, Hong Kong Baptist University, and Columbia University.

Multivariate Linear and Non-Linear Causality Tests

Abstract

The traditional linear Granger test has been widely used to examine the linear causality among several time series in bivariate settings as well as multivariate settings. Hiemstra and Jones [19] develop a nonlinear Granger causality test in bivariate settings to investigate the nonlinear causality between stock prices and trading volume. This paper extends their work by developing a non-linear causality test in multivariate settings.

Keywords: linear Granger causality, nonlinear Granger causality, U-statistics

1 Introduction

It is an important issue to detect the causal relation among several time series and it starts with two series, see, for example, Chiang, et al. [6], Leung and Wong [23], Qiao, et al. [31, 32], and the references therein for more discussion. To examine whether past information of one series could contribute to the prediction of another series, Granger causality test (Granger [16]) is developed to examine whether lag terms of one variable significantly explain another variable in a vector autoregressive regression model.

Linear Granger causality test can be used to detect the causal relation between two time series. However, the linear Granger causality test does not perform well in detecting nonlinear causal relationships. To circumvent this limitation, Baek and Brock [2] develop a nonlinear Granger causality test which is modified by Hiemstra and Jones [19] later on to study the bivariate non-linear causality relationship between two series by examining the remaining nonlinear predictive power of a residual series of a variable on the residual of another variable obtaining from a linear model. Nevertheless, the multivariate causal relationships are important but it has not been well studied, especially for nonlinear causality relationship. Thus, it is important to extend the Granger causality test to nonlinear causality test in multivariate settings.

In this paper, we first discuss linear causality tests in multivariate settings briefly and thereafter develop a non-linear causality test in multivariate settings. For any n variables involved in the causality test, we discuss a n-equation vector autoregressive regression (VAR) model to conduct the linear Granger test, and test for the significance of relevant coefficients across equations using likelihood ratio test. If those coefficients are significantly different from zero, the linear causality relationship is identified. We then extend the nonlinear Granger test from bivariate settings to multivariate settings. We notice that the bivariate nonlinear Granger test is developed by mainly applying the properties of U-statistic developed by Denker and Keller [7, 8]. Central limit theorem can be applied to the U-statistic whose arguments are strictly stationary, weakly dependent and satisfy mixing conditions of Denker and Keller [7, 8]. When we extend the test to multivariate settings, we find that the properties of the U-statistic for bivariate settings could also be used in the development of our proposed test statistic in multivariate settings, which is also a function of U-statistic.

The paper is organized as follows. We begin in next section to review the literature. Section 3 introduces definitions and notations and state some basic properties for the bivariate linear and nonlinear Granger causal tests. In Section 4, we first discuss the linear Granger causality tests in multivariate settings and thereafter develop the nonlinear Granger causality tests in multivariate settings. Section 5 gives a summary of our paper. Detailed proof will be given in the appendix.

2 Literature Review

Wiener [38] and Granger [16] first introduce the concept of causality which becomes a fundamental theory for analyzing dynamic relationships between variables of interest. Sims [37] provides a variant. Other substantial review and discussion in this area include Pierce and Haugh [30], Newbold [28], and Geweke [15]. Granger [16] first defines causality to be one period ahead predictability of a variable by another variable, which is also called causality in mean. Lütkepohl [24], Tjϕstheim [34], and others further extend this concept to vectors of variables. Thereafter, Lütkepohl [25] and Dufour and Renault [12] define causality in terms of predictability at any number of periods ahead.

In addition, Hosoya [20] extends the theory to analyze causality for stationary short-range dependent processes which do not necessarily follow a vector autoregressive model (VAR) whereas Lütkepohl and Poskitt [26] extend the theory by using a $VAR(\infty)$ model. On the other hand, Hidalgo [18] proposes to examine the test by covering long-range dependence while Saidi and Roy [35] extend the tests by deriving optimal rank-based tests for noncausality in the sense of Granger between two multivariate time series.

The Granger causality test introduced by Granger [16] is based on the assumption of linear relationships between variables and is, therefore, not able to explore the nonlinear causal relationship. However, Scheinkman and LeBaron [33], Brock, et al. [3], and others find evidence of significant nonlinear dependence in stock returns. In addition, there are many evidences of nonlinear causal relationships between variables in economics and finance. For example, Hsieh [21] notices that the nonlinear structure in stock price movements is motivated by asset behaviour that follows nonlinear models. Hiemstra and Jones [19] document evidence of significant nonlinear interaction between stock returns and trading volume.

To circumvent the limitation of the linear Granger causality test, Brock, et al. [4] propose a test that is based on the concept of correlation integral (Grassberger and Procaccia [17]). Moreover, Brock, et al. [5] develop an estimator of spatial probabilities across time to examine the identically and independently distributed assumption on the error term. In addition, Baek and Brock [2] propose a nonlinear Granger causality test to deal with the nonlinearity issue. The test is further modified by Hiemstra and Jones [19].

3 Bivariate Granger Causality Test

In this section, we will review the definitions of linear and nonlinear causality and discuss the linear and non-linear Granger causality tests to identify the causality relationships between two variables.

3.1 Bivariate Linear Granger Causality Test

Granger causality test is designed to detect causal direction between two time series by examining a correlation between the current value of one variable and the past values of another variable. Based on Granger's definition of causality, Y is strictly Granger causing X if the conditional distribution of X_t, given the past observations $X_{t-1}, X_{t-2}, \cdots$ and $Y_{t-1}, Y_{t-2}, \cdots$, differs from the conditional distribution of X_t, given the past observations

$X_{t-1}, X_{t-2}, \cdots$ only. Intuitively, Y is a Granger cause of X if adding past observations of Y to the information set increases the knowledge on the distribution of current values of X. More precisely, the **linear Granger causality** is conducted based on the following two-equation model:

Definition 3.1. *In a two-equation model:*

$$x_t = a_1 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{i=1}^{p} \beta_i y_{t-i} + \varepsilon_{1t} \tag{1a}$$

and

$$y_t = a_2 + \sum_{i=1}^{p} \gamma_i x_{t-i} + \sum_{i=1}^{p} \delta_i y_{t-i} + \varepsilon_{2t} \,, \tag{1b}$$

*where all $\{x_t\}$ and $\{y_t\}$ are stationary variables, p is the optimal lag in the system, and ε_{1t} and ε_{2t} are the disturbances satisfying the regularity assumptions of the classical linear normal regression model. The variable $\{y_t\}$ is said not to **Granger cause** $\{x_t\}$ if $\beta_i = 0$ in (1a), for any $i = 1, \cdots, p$. In other words, the past values of $\{y_t\}$ do not provide any additional information on the performance of $\{x_t\}$. Similarly, $\{x_t\}$ does not **Granger cause** $\{y_t\}$ if $\gamma_i = 0$ in (1b), for any $i = 1, \cdots, p$.*

It is well-known that one can test for linear causal relations between $\{x_t\}$ and $\{y_t\}$ by testing the following null hypotheses separately:

H_0^1: $\beta_1 = \cdots = \beta_p = 0$, and H_0^2: $\gamma_1 = \cdots = \gamma_p = 0$.

From testing these hypotheses, we have four possible testing results:

(1) If both Hypotheses H_0^1 and H_0^2 are accepted, there is no linear causal relationship between $\{x_t\}$ and $\{y_t\}$.

(2) If Hypothesis H_0^1 is accepted but Hypothesis H_0^2 is rejected, then there exists linear causality running unidirectionally from $\{x_t\}$ to $\{y_t\}$.

(3) If Hypothesis H_0^1 is rejected but Hypothesis H_0^2 is accepted, then there exists linear causality running unidirectionally from $\{y_t\}$ and $\{x_t\}$.

(4) If both Hypotheses H_0^1 and H_0^2 are rejected, then there exist feedback linear causal relationships between $\{x_t\}$ and $\{y_t\}$.

Several statistics could be used to test the above hypotheses. One of the most commonly used statistics is the standard reduced-versus-full-model F-test to test the hypothesis $H_0^1 : \beta_1 = \cdots = \beta_p = 0$ in (1a) and to identify the linear causal relationship from $\{y_t\}$ to $\{x_t\}$. Similarly, one could test the second null hypothesis $H_0^2 : \gamma_1 = \cdots = \gamma_p = 0$, and identify the linear causal relationship from $\{x_t\}$ to $\{y_t\}$.

3.2 Bivariate Nonlinear Causality Test

The general test for nonlinear Granger causality is first developed by Baek and Brock [2] and, later on, modified by Hiemstra and Jones [19]. As the linear Granger test is inefficient in detecting any nonlinear causal relationship, to examine the nonlinear Granger causality relationship between a pair of series, say $\{x_t\}$ and $\{y_t\}$, one has to first apply the linear models in (1a) and (1b) to $\{x_t\}$ and $\{y_t\}$ for identifying their linear causal relationships and obtaining their corresponding residuals, $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$. Thereafter, one has to apply a non-linear Granger causality test to the residual series, $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$, of the two variables, $\{x_t\}$ and $\{y_t\}$, being examined to identify the remaining nonlinear causal relationships between their residuals.

We first state the definition of nonlinear Granger causality and discuss the corresponding test developed by Hiemstra and Jones as follows:

Definition 3.2. *For any two strictly stationary and weakly dependent series $\{X_t\}$ and $\{Y_t\}$, the m-length lead vector of X_t and L_x-length lag vector of X_t are defined as*

$$\begin{aligned} X_t^m &\equiv \big(X_t, X_{t+1}, \cdots, X_{t+m-1}\big),\ m = 1, 2, \cdots,\ t = 1, 2, \cdots \\ X_{t-L_x}^{L_x} &\equiv \big(X_{t-L_x}, X_{t-L_x}, \cdots, X_{t-1}\big), L_x = 1, 2, \cdots, t = L_x + 1, L_x + 2, \cdots. \end{aligned}$$

The m-length lead vector, Y_t^m, and the L_x-length lag vector, $X_{t-L_x}^{L_x}$, of Y_t can be defined similarly. Series $\{Y_t\}$ ***does not strictly Granger cause*** *another series $\{X_t\}$ non-*

linearly if and only if

$$
\begin{aligned}
&Pr\left(\|X_t^m - X_s^m\| < e \,\middle|\, \| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e, \| Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \| < e\right) \\
=&Pr\left(\|X_t^m - X_s^m\| < e \,\middle|\, \| X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \| < e\right) \ ,
\end{aligned}
$$

where $Pr(\cdot\,|\,\cdot)$ denotes conditional probability and $\|\cdot\|$ denotes the maximum norm which is defined as $\|X - Y\| = max\big(|x_1 - y_1|, |x_2 - y_2|, \cdots, |x_n - y_n|\big)$ for any two vectors $X = \big(x_1, \cdots, x_n\big)$ and $Y = \big(y_1, \cdots, y_n\big)$.

Under Definition 3.2, the non-linear Granger causality test statistic is given by

$$
\sqrt{n}\left(\frac{C_1\big(m+L_x, L_y, e, n\big)}{C_2\big(L_x, L_y, e, n\big)} - \frac{C_3\big(m+L_x, e, n\big)}{C_4\big(L_x, e, n\big)}\right) \ , \tag{1}
$$

where

$$
\begin{aligned}
&C_1\big(m+L_x, L_y, e, n\big) \equiv \frac{2}{n(n-1)} \sum\sum_{t<s} I\big(x_{t-L_x}^{m+L_x}, x_{s-L_x}^{m+L_x}, e\big) \cdot I\left(y_{t-L_y}^{L_y}, y_{s-L_y}^{L_y}, e\right) \ , \\
&C_2\big(L_x, L_y, e, n\big) \equiv \frac{2}{n(n-1)} \sum\sum_{t<s} I\left(x_{t-L_x}^{L_x}, x_{s-L_x}^{L_x}, e\right) \cdot I\left(y_{t-L_y}^{L_y}, y_{s-L_y}^{L_y}, e\right) \ , \\
&C_3\big(m+L_x, e, n\big) \equiv \frac{2}{n(n-1)} \sum\sum_{t<s} I\left(x_{t-L_x}^{m+L_x}, x_{s-L_x}^{m+L_x}, e\right) \ , \\
&C_4\big(L_x, e, n\big) \equiv \frac{2}{n(n-1)} \sum\sum_{t<s} I\left(x_{t-L_x}^{L_x}, x_{s-L_x}^{L_x}, e\right) \ , \ \text{and} \\
&I(x, y, e) = \begin{cases} 0, & \text{if } \|x - y\| > e \\ 1, & \text{if } \|x - y\| \le e \end{cases} \ .
\end{aligned}
$$

The test statistic, see Hiemstra and Jones [19], possesses the following property:

Theorem 3.1. *For given values of m, L_x, L_y, and $e > 0$ defined in Definition 3.2, under the assumptions that both $\{X_t\}$ and $\{Y_t\}$ are strictly stationary, weakly dependent, and satisfy the conditions stated in Denker and Keller [7], if $\{Y_t\}$ does not strictly Granger* *cause $\{X_t\}$, then the test statistic defined in (1) is distributed as $N\big(0, \sigma^2(m, L_x, L_y, e)\big)$ asymptotically, and the estimator of the variance $\sigma^2(m, L_x, L_y, e)$ is given in their appendix.*

4 Multivariate Granger Causality Test

In this section, we first discuss the linear Granger causality tests in multivariate settings and, thereafter, develop the non-linear Granger causality tests from bivariate settings to multivariate settings.

4.1 Multivariate Linear Granger Causality Test

We first discuss the linear Granger causality test from bivariate settings to multivariate settings.

4.1.1 Vector Autoregressive Regression

The linear Granger test is applied in the vector autoregressive regression (VAR) scheme. For $t = 1, \cdots, T$, the n-variable VAR model is represented as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix} = \begin{pmatrix} A_{10} \\ A_{20} \\ \vdots \\ A_{n0} \end{pmatrix} + \begin{pmatrix} A_{11}(L) & A_{12}(L) & \dots & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \dots & A_{2n}(L) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}(L) & A_{n2}(L) & \dots & A_{nn}(L) \end{pmatrix} \begin{pmatrix} y_{1,\ t-1} \\ y_{2,\ t-1} \\ \vdots \\ y_{n,\ t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{nt} \end{pmatrix}, \tag{2}$$

where $(y_{1t}, \cdots, y_{nt})$ is the vector of n stationary time series at time t, L is the backward operation in which $Lx_t = x_{t-1}$, A_{i0} are intercept parameters, $A_{ij}(L)$ are polynomials in the lag operator L such that $A_{ij}(L) = a_{ij}(1)L + a_{ij}(2)L^2 + \cdots + a_{ij}(p)L^p$, and $e_t = (e_{1t}, \cdots, e_{nt})'$

is the disturbance vector obeying the assumption of the classical linear normal regression model.

In practice, it is common to set all the equations in VAR to possess the same lag length for each variable. So a uniform order p will be chosen for all the lag polynomials $A_{ij}(L)$ in the VAR model according to a certain criteria such as Akaike's information and Schwartz Bayesian criteria. Along with the Gauss-Markov assumptions satisfied for the error terms, ordinary least square estimation (OLSE) is appropriate to be used to estimate the model as it is consistent and efficient. However, long lag length for each variable will consume large number of degrees of freedom. For example, in the model stated in equation 2, there will be $n(np+1)$ coefficients including n intercept terms, n variances and $n(n-1)/2$ covariances to be estimated. When the available sample size T is not large enough, including too many regressors will make the estimation inefficient, and thus, cause the test unreliable. To circumvent this problem, one could adopt a near-VAR model and seemingly unrelated regressions estimation technique instead of applying OLSE to estimate the equations simultaneously. We skip the discussion of the near-VAR model and seemingly unrelated regressions estimation in this paper and, for simplicity, we only use OLSE to estimate the parameters in the VAR model to identify the causality relationship among vectors of different time series.

4.1.2 Multiple Linear Granger Causality Hypothesis and Likelihood Ratio Test

To test the linear causality relationship between two vectors of different 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 n-equation VAR as follows:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} A_{x[n_1\times 1]} \\ A_{y[n_2\times 1]} \end{pmatrix} + \begin{pmatrix} A_{xx}(L)_{[n_1\times n_1]} & A_{xy}(L)_{[n_1\times n_2]} \\ A_{yx}(L)_{[n_2\times n_1]} & A_{yy}(L)_{[n_2\times n_2]} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} e_x \\ e_y \end{pmatrix} , \qquad (3)$$

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

Similar to the bivariate case, there are four different situations for the existence of linear causality relationships between two vectors of time series x_t and y_t in (3):

(1) There exists a *unidirectional causality* from y_t to x_t if $A_{xy}(L)$ is significantly different from zero[1] and, at the same time, $A_{yx}(L)$ is not significantly different from zero;

(2) there exists a *unidirectional causality* from x_t to y_t if $A_{yx}(L)$ is significantly different from zero and, at the same time, $A_{xy}(L)$ is not significantly different from zero;

(3) there exist *feedback* relations when both $A_{xy}(L)$ and $A_{yx}(L)$ are significantly different from zero; and

(4) x_t and y_t are not rejected to be *independent* when both $A_{xy}(L)$ and $A_{yx}(L)$ is not significantly different from zero.

We note that one could consider one more situation as follows:

(5) x_t and y_t are rejected to be *independent* when either $A_{xy}(L)$ and $A_{yx}(L)$ is significantly different from zero. This is the same situation as either (1), (2) or (3) is true.

Testing the above statements is equivalent to testing the following null hypotheses:

(1) H_0^1: $A_{xy}(L) = 0$,

(2) H_0^2: $A_{yx}(L) = 0$, and

(3) both H_0^1 and $H_0^2 :\ A_{xy}(L) = 0$ and $A_{yx}(L) = 0$.

One may first obtain the residual covariance matrix Σ from the full model in (3) by using OLSE for each equation without imposing any restriction on the parameters, and

[1] We said $A_{xy}(L)$ is significantly different from zero if there exists any term in $A_{xy}(L)$ which is significantly different from zero.

compute the residual covariance matrix Σ_0 from the restricted model in (3) by using OLSE for each equation with the restriction on the parameters imposed by the null hypothesis, H_0^1, H_0^2, or both H_0^1 and H_0^2. Thereafter, besides using the F-test, one could use a similar approach as in Sims [37] to obtain the following likelihood ratio statistic $(T - c)(\ log|\Sigma_0| - log|\Sigma|\ )$ where T is the number of usable observations, c is the number of parameters estimated in each equation of the unrestricted system, and $log|\Sigma_0|$ and $log|\Sigma|$ are the natural logarithms of the determinants of restricted and unrestricted residual covariance matrices, respectively. When the null hypothesis is true, this test statistic has a asymptotic χ^2 distribution with the degrees of freedom equal to the number of restrictions on the coefficients in the system. For example, when we test $H_0 : A_{xy}(L) = 0$, one should let c equal to $np + 1$, and there are $n_2 \times p$ restrictions on the coefficients in the first n_1 equations of the model. Hence, the corresponding test statistic $(T - (np + 1))(\ log|\Sigma_0| - log|\Sigma|\ )$ asymptotically follows χ^2 with $n_1 \times n_2 \times p$ degrees of freedom. The conventional bivariate causality test is a special case of the multivariate causality test when $n_1 = n_2 = 1$. Besides using the F test, one could also use the likelihood ratio test to identify the linear causality relationship for any two variables in bivariate settings.

4.1.3 ECM-VAR model

Engle and Granger [13] show that if two non-stationary variables are cointegrated, we need to specify a model with a dynamic error correction representation when testing for Granger causality. This means that the traditional VAR model is augmented with a one-period lagged error correction term that is obtained from the cointegrated model.

Consider $(Y_{1t}, \cdots, Y_{nt})$ to be a vector of n non-stationary time series and assume that cointegration (Engle and Granger [13], Franses and McAleer [14]) exists with residual vector $vecm_t$. Let $y_{it} = \Delta Y_{it}$ for $i = 1, \cdots, n$ be the corresponding stationary differencing series. In this situation, one should not use the VAR model as stated in (2), one should impose the error-correction mechanism (ECM) on the VAR to test for Granger causality between these variables and obtain the ECM-VAR framework as follows:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix} = \begin{pmatrix} A_{10} \\ A_{20} \\ \vdots \\ A_{n0} \end{pmatrix} + \begin{pmatrix} A_{11}(L) & A_{12}(L) & \dots & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & \dots & A_{2n}(L) \\ \vdots & \vdots & & \vdots \\ A_{n1}(L) & A_{n2}(L) & \dots & A_{nn}(L) \end{pmatrix} \begin{pmatrix} y_{1,\ t-1} \\ y_{2,\ t-1} \\ \vdots \\ y_{n,\ t-1} \end{pmatrix}$$
$$+ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{+n} \end{pmatrix} \cdot ecm_{t-1} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{nt} \end{pmatrix}, \tag{4}$$

where ecm_{t-1} is the error correction term. In particular, in this paper, we consider to test the causality relationship between two non-stationary vector time series, $X_t = (X_{1,t}, \cdots, X_{n_1,t})'$ and $Y_t = (Y_{1,t}, \cdots, Y_{n_2,t})'$ in which $x_{it} = \Delta X_{it}$ and $y_{it} = \Delta Y_{it}$ are the corresponding stationary differencing series, where there are $n_1 + n_2 = n$ series in total. If X_t and Y_t are cointegrated with residual vector $vecm_t$, then, instead of using the n-equation VAR in (3), one should adopt the n-equation VAR in the following:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} A_{x[n_1\times 1]} \\ A_{y[n_2\times 1]} \end{pmatrix} + \begin{pmatrix} A_{xx}(L)_{[n_1\times n_1]} & A_{xy}(L)_{[n_1\times n_2]} \\ A_{yx}(L)_{[n_2\times n_1]} & A_{yy}(L)_{[n_2\times n_2]} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}$$
$$+ \begin{pmatrix} \alpha_{x[n_1\times 1]} \\ \alpha_{y[n_2\times 1]} \end{pmatrix} \cdot ecm_{t-1} + \begin{pmatrix} e_x \\ e_y \end{pmatrix}, \tag{5}$$

where $a_{x[n_1\times 1]}$ and $a_{y[n_2\times 1]}$ are two vectors of intercept terms, $A_{xx}(L)_{[n_1\times n_1]}$, $A_{xy}(L)_{[n_1\times n_2]}$, $A_{yx}(L)_{[n_2\times n_1]}$, $A_{yy}(L)_{[n_2\times n_2]}$ are matrices of lag polynomials, and $\alpha_{x[n_1\times 1]}$ and $\alpha_{y[n_2\times 1]}$ are the coefficient vectors for the error correction term ecm_{t-1}. Thereafter, one should test the null hypothesis $H_0 : A_{xy}(L) = 0$ or $H_0 : A_{yx}(L) = 0$ to identify strict causality relation using the LR test as discussed in Section 4.1.2. The lag length was selected using the two-stage procedure suggested in Abdalla and Murinde [1]. Following the Abdalla and Murinde [1] approach, the optimal lag length was selected through maximizing the value of the R^2.

4.2 Multivariate Nonlinear Causality Test

In this section, we will extend the nonlinear causality test for a bivariate setting developed by Hiemstra and Jones [19] to a multivariate setting.

4.2.1 Multivariate Nonlinear Causality Hypothesis

As discussed in Section 3.2, to identify any nonlinear Granger causality relationship from any two series, say $\{x_t\}$ and $\{y_t\}$ in bivariate settings, one has to first apply the linear models in (1a) and (1b) to $\{x_t\}$ and $\{y_t\}$ to identify their linear causal relationships and obtain their corresponding stationary residuals, $\{\hat{\varepsilon}_{1t}\}$ and $\{\hat{\varepsilon}_{2t}\}$. Thereafter, one has to apply a non-linear 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 relations 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 multivariate settings. One has to apply the n-equation VAR model in (3) or (5) to the series to identify their linear causal relationships and obtain their corresponding residuals. Thereafter, one has to apply a non-linear Granger causality test to the residual series instead of the original time series. For simplicity, in this section we will 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 stationary residuals of two vectors of variables being examined.

We first define the lead vector and lag vector of a time series, say $X_{i,t}$, similar to the terms defined in Definition 3.2 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 defined, respectively, as

$$\begin{aligned} X_{i,t}^{m_{x_i}} &\equiv (X_{i,t}, X_{i,t+1}, \cdots, X_{i,\ t+m_{x_i}-1}), m_{x_i} = 1, 2, \cdots, t = 1, 2, \cdots, \quad \text{and} \\ X_{i,\ t-L_{x_i}}^{L_{x_i}} &\equiv (X_{i,\ t-L_{x_i}}, X_{i,\ t-L_{x_i}+1}, \cdots, X_{i,\ t-1}), L_{x_i} = 1, 2, \cdots, t = L_{x_i}+1, L_{x_i}+2, \cdots. \end{aligned}$$

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 $Y_{i,\ t}$, M_y, L_y, m_y, and l_y can be defined similarly.

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

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

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

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

where $\|\cdot\|$ denotes the maximum norm defined in Definition 3.2.

The vector series $\{Y_t\}$ is said not to strictly Granger cause another vector series $\{X_t\}$ if:

$$\begin{aligned} &Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e \,\middle|\, \parallel X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \parallel < e, \parallel Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \parallel < e, \right) \\ = &Pr\left(\|X_t^{M_x} - X_s^{M_x}\| < e, \,\middle|\, \parallel X_{t-L_x}^{L_x} - X_{s-L_x}^{L_x} \parallel < e\right) , \end{aligned}$$

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

4.2.2 Test Statistic and It's Asymptotic Distribution

Similar to the bivariate case, the test statistic for testing non-existence of nonlinear Granger causality can be obtained as follows:

$$\sqrt{n}\left(\frac{C_1\big(M_x + L_x, L_y, e, n\big)}{C_2\big(L_x, L_y, e, n\big)} - \frac{C_3\big(M_x + L_x, e, n\big)}{C_4\big(L_x, e, n\big)}\right) , \tag{6}$$

where

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

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

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

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

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

In this paper, we extend the theory by developing the following theorem:

Theorem 4.1. *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 [7], if the null hypothesis, H_0, is true, the test statistic defined in (6) is distributed as $N\big(0,\sigma^2(M_x,L_x,L_y,e)\big)$. When the test statistic in (6) is too far away from zero, we reject the null hypothesis. A consistent estimator of $\sigma^2(M_x,L_x,L_y,Lz,e)$ follows:*

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

in which each component $\Sigma_{i,j}$ $(i,j=1,\cdots,4)$, of the covariance matrix Σ is given by:

$$\Sigma_{i,j}=4\cdot\sum_{k\geq1}\omega_k E(A_{i,t}\cdot A_{j,t+k-1}),$$

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

$$
\begin{aligned}
A_{1,t} &= h_{11}\left(x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}, e\right) - C_1(M_x+L_x, L_y, e)\,,\\
A_{2,t} &= h_{12}\left(x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}, e\right) - C_2(L_x, L_y, e)\,,\\
A_{3,t} &= h_{13}\left(x_{t-L_x}^{M_x+L_x}, e\right) - C_3(M_x+L_x, e)\,,\ \textit{and}\\
A_{4,t} &= h_{14}\left(x_{t-L_x}^{L_x}, e\right) - C_4(L_x, e)\,,
\end{aligned}
$$

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:*

$$
\begin{aligned}
&h_{11}\left(x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}, e\right) = E\big(h_1 \,\big|\, x_{t-L_x}^{M_x+L_x}, y_{t-L_y}^{L_y}\big)\,,\\
&h_{12}\left(x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}, e\right) = E\big(h_2 \,\big|\, x_{t-L_x}^{L_x}, y_{t-L_y}^{L_y}\big)\,,\\
&h_{13}\left(x_{t-L_x}^{M_x+L_x}, e\right) = E\big(h_3 \,\big|\, x_{t-L_x}^{M_x+L_x}\big)\,, \textit{and}\\
&h_{14}\left(x_{t-L_x}^{L_x}, e\right) = E\big(h_4 \,\big|\, x_{t-L_x}^{L_x}\big)\,.
\end{aligned}
$$

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

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

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

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:*

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

The proof of Theorem 4.1 is shown in the appendix.

5 Conclusion Remarks

In this paper, we first discuss linear causality tests in multivariate settings and thereafter develop a non-linear causality test in multivariate settings.

We note that there could be some limitations for the Hiemstra-Jones test. For example, Diks and Panchenko [10] point out that Hiemstra-Jones test might have an over-rejection bias on the null hypothesis of Granger non-causality. Their simulation results show that rejection probability will go to one as the sample size increases. Diks and Panchenko [11] address this problem by replacing the global test by an average of local conditional dependence measures. Their new test shows weaker evidence for volume causing returns than Hiemstra-Jones test does. Besides Hiemstra-Jones test, other forms of nonlinear causality test have also been developed. For example, Marinazzo, et al. [27] adopt the theory of reproducing kernel Hilbert spaces to develop a nonlinear Granger causality test. On the other hand, Diks and DeGoede [9] develop an information theoretic test statistics for Granger causality. They use bootstrap methods instead of asymptotic distribution to calculate the significance of the test statistics. Thus, further extension of this paper could include development of multivariate settings for the more powerful linear and non-linear causality tests.

Appendix : Proof of Theorem 4.1

Before we prove the theorem, we first introduce the U-statistic (Kowalski and Tu [22]) in the following definition, which is essential in the proof.

Definition 5.1. *Consider an i.i.d. sample of* $p \times 1$ *column vector of response* $\mathbf{y_i}$ *(*$1 \leq i \leq n$*). Let* $\mathbf{h}\big(\mathbf{y}_1, \cdots, \mathbf{y}_m\big)$ *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:*

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

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

We note that in Definition 5.1 "*multivariate*" refers to the dimension of U-statistic. Let $\Theta = E\big(\mathbf{h}(\mathbf{y}_1, \cdots, \mathbf{y}_m)\big)$. Then, we have

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

Now, we proceed on to prove Theorem 4.1. We denote that

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

Then, we have

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

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 also satisfy the mixing conditions of Denker and Keller [7] such that

$$\begin{aligned}
\mathbf{z_t} &= \left(X_{1,t-L_{x_1}}^{M_{x_1}+L_{x_1}}, \cdots, X_{n_1,t-L_{x n_1}}^{M_{x_1}+L_{x_1}}, Y_{1,t-L_{y_1}}^{L_{y_1}}, \cdots, Y_{n_2,t-L_{y n_2}}^{L_{y n_2}}\right)\\
t &= max\big(L_x,L_y\big),\cdots,n; \ n = T - m_x - max\big(L_x,L_y\big)+1
\end{aligned}$$

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 $\big(M_x,L_x,L_y,e\big)$, we denote $\Theta = \big(\theta_1,\ \theta_2,\ \theta_3,\ \theta_4\big)'$ in which $\theta_1 \equiv C_1\big(M_x+L_x,L_y,e\big)$, $\theta_2 \equiv C_2\big(L_x,L_y,e\big)$, $\theta_3 \equiv C_3\big(M_x+L_x,e\big)$, and $\theta_4 \equiv C_4\big(L_x,e\big)$. In addition, we denote $\mathbf{U_n} = \big(U_{1n},U_{2n},U_{3n},U_{4n}\big)'$ in which $U_{1n} \equiv C_1\big(M_x+L_x,L_y,e,n\big)$, $U_{2n} \equiv C_2\big(L_x,L_y,e,n\big)$,

$U_{3n} \equiv C_3\big(M_x + L_x, e, n\big)$, and $U_{4n} \equiv C_4\big(L_x, e, n\big)$. One could easily show that $\mathbf{U_n}$ is a one-sample, 2-argument, and 4-variable U-statistic vector with kernel vector $\mathbf{h}\big(\mathbf{z_t}, \mathbf{z_s}\big) = \big(h_1, h_2, h_3, h_4)'$, where

$$
\begin{aligned}
h_1 &\equiv \prod_{i=1}^{n_1} I\big(X_{i,t-L_{x_i}}^{m_{x_i}+L_{x_i}}, X_{i,s-L_x}^{m_{x_i}+L_{x_i}}, e) \cdot \prod_{i=1}^{n_2} I\big(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_{y_i}}^{L_{y_i}}, e)\,, \\
h_2 &\equiv \prod_{i=1}^{n_1} I\big(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_x}^{L_{x_i}}, e) \cdot \prod_{i=1}^{n_2} I\big(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_y}^{L_{y_i}}, e)\,, \\
h_3 &\equiv \prod_{i=1}^{n_1} I\big(X_{i,t-L_x}^{m_{x_i}+L_{x_i}}, X_{i,s-L_{x_i}}^{m_{x_i}+L_{x_i}}, e)\,, \\
h_4 &\equiv \prod_{i=1}^{n_1} I\big(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_{x_i}}^{L_{x_i}}, e)\,,
\end{aligned}
$$

and

$$
\mathbf{U_n} = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{1 \le t < s \le n} \mathbf{h}\big(\mathbf{z_t}, \mathbf{z_s}\big)\,.
$$

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

$$
\begin{aligned}
E\big(h_1\big) &= Pr\big(\parallel X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \parallel < e) \cdot Pr\big(\parallel Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \parallel < e) \\
&= Pr\big(\parallel X_{t-L_x}^{M_x+L_x} - X_{s-L_x}^{M_x+L_x} \parallel < e, \;\; \parallel Y_{t-L_y}^{L_y} - Y_{s-L_y}^{L_y} \parallel < e) \\
&= \theta_1
\end{aligned}
$$

since residual series $\{X_t\}$ and $\{Y_t\}$ obtained from the VAR model are assumed to be independent with each other. Similarly, we have

$$
E\big(h_i\big) = \theta_i, \quad i = 2, 3, 4\,.
$$

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 modification 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 [7] 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 [8], the sequence $\mathbf{U_n}$ of 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

$$f(\Theta) = \frac{\theta_1}{\theta_2} - \frac{\theta_3}{\theta_4} = 0\,.$$

Thus, using the delta method (Serfling [36], 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

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

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

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

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

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

Applying the results of Denker and Keller [7] and Newey and West [29], we obtain a consistent estimator of $\Sigma_{i,j}$ elements to be:

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

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

$$
\begin{aligned}
\hat{A}_{1,t} = &\frac{1}{n-1}\left(\sum_{s\neq t}\prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{m_{x_i}+L_{x_i}}, X_{i,s-L_x}^{m_{x_i}+L_{x_i}}, e)\cdot\prod_{i=1}^{n_2} I(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_{y_i}}^{L_{y_i}}, e)\right) \\
&- C_1(M_x+L_x, L_y, e, n)\,, \\
\hat{A}_{2,t} = &\frac{1}{n-1}\left(\sum_{s\neq t}\prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_x}^{L_{x_i}}, e)\cdot\prod_{i=1}^{n_2} I(Y_{i,t-L_{y_i}}^{L_{y_i}}, Y_{i,s-L_y}^{L_{y_i}}, e)\right) \\
&- C_2(L_x, L_y, e, n)\,, \\
\hat{A}_{3,t} = &\frac{1}{n-1}\left(\sum_{s\neq t}\prod_{i=1}^{n_1} I(X_{i,t-L_x}^{m_{x_i}+L_{x_i}}, X_{i,s-L_{x_i}}^{m_{x_i}+L_{x_i}}, e)\right) \\
&- C_3(m+L_x, e, n)\,, \\
\hat{A}_{4,t} = &\frac{1}{n-1}\left(\sum_{s\neq t}\prod_{i=1}^{n_1} I(X_{i,t-L_{x_i}}^{L_{x_i}}, X_{i,s-L_{x_i}}^{L_{x_i}}, e)\right) \\
&- C_4(L_x, e, n)\,, \\
t, s = &\ max(L_x, L_y), \cdots, n \text{ and } n = T - m_x - max(L_x+L_y) + 1\,,
\end{aligned}
$$

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

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

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

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

and the assertion of the theorem follows.

□

References

[1] I. Abdalla, V. Murinde, Exchange rate and stock price interactions in emerging financial markets: Evidence on India, Korea, Pakistan and the Philippines, Applied Financial Economics 7 (1997) 25-35.

[2] E.G. Baek, W.A. Brock, A general test for nonlinear Granger causality : bivariate model, working paper, Korea Development Institute, University of Wisconsin-Madison, (1992).

[3] D. Brock, D. Hsieh, B. LeBaron, A Test of Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence, MIT Press, Cambridge, MA (1991).

[4] W. Brock, W. Dechect, J. Scheinkman, A test for independence based on the correlation dimension, Working Paper. Department of Economics, University of Wisconsin, Madison, (1987).

[5] W. Brock, W. Dechect, J. Scheinkman, A test for independence based on the correlation dimension, Econometric Reviews 15 (1996) 197-235.

[6] T.C. Chiang, Z. Qiao, W.K. Wong, New evidence on the relation between return volatility and trading volume, Journal of Forecasting 29(5) (2009) 502 - 515.

[7] M. Denker, G. Keller, On U-statistics and v. Mises' statistics for weakly dependent processes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 64 (1983) 505-522.

[8] M. Denker, G. Keller, Rigorous statistical procedures for data from dynamical system, Journal of Statistical Physics 44 (1986) 67-93.

[9] C. Diks, J. DeGoede, A general nonparametric bootstrap test for Granger causality, in: Global Analysis of Dynamical Systems, eds. H.W. Broer, B. Krauskopf, G. Vegter, (2001) 391-403, IoP Publishing, Bristol.

[10] C. Diks, V. Panchenko, A note on the Hiemstra-Jones test for Granger non-causality, Studies in Nonlinear Dynamics and Econometrics 9(2) (2005) Art. No. 4, 1-7.

[11] C. Diks, V. Panchenko, A new statistic and practical guidelines for nonparametric Granger causality testing, Journal of Economic Dynamics and Control 30(9-10) (2006) 1647-1669.

[12] J.M. Dufour, E. Renault, Short run and long run causality in time series: Theory, Econometrica 66 (1998) 1099-1125.

[13] R.F. Engle, C.W.J. Granger, cointegration and error correction: representation, estimation, and testing, Econometrica 55 (1987) 251-276.

[14] P.H. Franses, M. McAleer, Testing for unit roots and non-linear transformations, Journal of Time Series Analysis 19(2) (2001) 147-164.

[15] J. Geweke, Inference and causality in economic time series, In Z. Griliches, M.D. Intrilligator (eds.), Handbook of Econometrics (1984) 1102-1144. North-Holland.

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

[17] P. Grassberger, I. Procassia, measuring the strangeness attractors, Physica D 9 (1983) 189-208.

[18] J. Hidalgo, Nonparametric test for causality with long-range dependence, Econometrica 68 (2000) 1465-1490.

[19] C. Hiemstra, J.D. Jones, Testing for linear and nonlinear Granger causality in the stock price-volume relation, Journal of Finance 49(5) (1994) 1639-1664.

[20] Y. Hosoya, The decomposition and measurement of the interdependency between second-order stationary processes, Probability Theory and Related Fields 88 (1991) 429-444.

[21] D. Hsieh, Chaos and nonlinearity dynamics: application to financial market, Journal of Finance 46 (1991) 1839-1877.

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

[23] P.L. Leung, W.K. Wong, Three-factor profile analysis with GARCH innovations, Mathematics and Computers in Simulation 77(1) (2008) 1-8.

[24] H. Lütkepohl, Introduction to Multiple Time Series Analysis, (1991) Springer-Verlag.

[25] H. Lütkepohl, Testing for causation between two variables in higher dimensional VAR models, In H. Schneeweis, K. Zimmerman, (eds.), Studies in Applied Econometrics (1993) 75-91.

[26] H. Lütkepohl, D.S. Poskitt, Testing for causation using infinite order vector autoregressive processes, Econometric Theory 12 (1996) 61-87.

[27] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear Granger causality, Physical Review Letters 100(14), Article 144103 (2008) 1-4.

[28] P. Newbold, Causality testing in economics, In O.D. Anderson (ed.), Time Series Analysis: Theory and Practice 1 (1982) 701-716. North-Holland.

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

[30] D.A., Pierce, L.D. Haugh, Causality in temporal systems: Characterization and survey, Journal of Econometrics 5 (1977) 265-293.

[31] Z. Qiao, Y. Li, W.K. Wong, Policy change and lead-lag relations among China's segmented stock markets, Journal of Multinational Financial Management 18 (2008) 276-289.

[32] Z. Qiao, M. McAleer, W.K. Wong, Linear and nonlinear causality of consumption growth and consumer attitudes, Economics Letters 102(3) (2009) 161-164.

[33] J. Scheinkman, B. LeBaron, Nonlinear dynamics and stock returns, Journal of Business 62 (1989) 311-37.

[34] D. Tjϕstheim, Granger causality in multiple time series, Journal of Econometrics 17 (1981) 157-176.

[35] A. Saidi, R. Roy, Robust optimal tests for causality in multivariate time series, Econometric Theory 24 (2008) 948-986.

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

[37] C.A. Sims, Macroeconomics and reality, Econometrica 48 (1980) 1-48.

[38] N. Wiener, The theory of prediction, in E.F. Beckenback (ed.), Modern Mathematics for Engineers (1956) 165-190. McGraw-Hill.