GRANGER TYPE CAUSALITY FOR NONLINEAR DATA

Ohama, Gan
Faculty of Economics, Shiga University

GRANGER TYPE CAUSALITY FOR NONLINEAR DATA

By

Gan Ohama*

Abstract

The causality proposed by Granger (1969) and several tests for it are often used in economic science to assess relationships among more than two factors of interest. However, the Granger causality presumes linearity and has limitation to analyze the relationship of "nonlinear" factors. In this paper we extend the Granger causality to nonlinear relationships and propose methods for analysis of data by the causality; the efficiency of the methods is also investigated.

Key Words and Phrases: Granger causality, Nonlinear Granger causality, cross-validation

1. Introduction

When we have two sets of time series data, for example money supply data and gross national product (GNP) data, it might be a natural question that which one causes another. Granger (1969) defined causality from the viewpoint of prediction. By this definition, if one factor does help to predict the other, the former is recognized as the "causal factor" to the other. This idea has been applied in economic science, and several statistical tests have been proposed for the Granger causality by Sims (1972) and Sargent (1976). Nakano (1995) improved the Granger causality by introducing the idea of local causality. The Granger causality has been generalized as the nonlinear causality taking account of a nonlinear relationship between two time series. Baek and Brock (1992) approached it by means of conditional probabilities, Okabe and Inoue (1994) defined it by assuming that the data at a certain time can be predicted perfectly as a polynomial function of the preceding data. These ideas of nonlinear causality, however, might not be appropriate for the data of small sample size with large noise, typically seen in economic data. The aim of this paper is to define nonlinear causality based on prediction, which is analyzable and similar fashion to Granger (1969), and to propose two methods of nonlinear causality analysis. The approach is not based on a standard statistical testing method, but similar to a method of model selection. In this sense the approach is different from Baek and Brock (1992) and Okabe and Inoue (1994).

2. Preliminary

In this section we briefly introduce the Granger causality. Suppose that X_t and Y_t are real random variables with mean 0 and defined on the common probability space $(\Omega, \mathscr{B}, \Pr)$ for any integer t. Denote by $P_\mathscr{M}$ the projection operator onto $\mathscr{M}$, by $\mathfrak{M}_{s,t}(X)$

* Faculty of Economics, Shiga University, 1-1-1 Banba, Hikone 522-8522, Japan. tel +81-749-27-1114
  ohama@biwako.shiga-u.ac.jp

the closed subspace in $L^2(\Omega, \mathscr{B}, \Pr)$ spanned by $\{X_k | s \leq k \leq t\}$. Furthermore, denote by $\mathscr{M}_{s,t}(X)$ the closed subspace consists of all measurable random variables with respect to the σ-field generated by $\{X_k | s \leq k \leq t\}$, $\mathscr{F}_{s,t}(X)$ say, and $\sigma^2(X_t | \mathscr{M})$ by the variance of $X_t - P_{\mathscr{M}}(X_t)$. We use $\mathfrak{M}_t(X)$ instead of $\mathfrak{M}_{-\infty,t}(X)$ for short, and also we use $\mathfrak{M}_{s,t}(X, Y)$, $\mathscr{M}_{s,t}(X, Y)$ and $\mathscr{F}_{s,t}(X, Y)$ in the similar way.

The definition of the Granger causality between $\{X_t\}$ and $\{Y_t\}$ is as follows. To simplify the argument, we assume that there is no useful information, that is, data, to predict X_t except $\{X_t\}$ and $\{Y_t\}$.

DEFINITION 2.1. For a two dimensional weakly stationary time series $\{(X_t, Y_t)\}$, it is said that $\{Y_t\}$ *causes* $\{X_t\}$ *in the sense of Granger*, written $\{Y_t\} \xrightarrow{GC} \{X_t\}$, if and only if

$$\sigma^2(X_t | \mathfrak{M}_{t-1}(X, Y)) < \sigma^2(X_t | \mathfrak{M}_{t-1}(X)).$$

We also say $\{Y_t\}$ *does not cause* $\{X_t\}$ *in the sense of Granger*, written $\{Y_t\} \xrightarrow{GC} \{X_t\}$, if and only if

$$\sigma^2(X_t | \mathfrak{M}_{t-1}(X, Y)) = \sigma^2(X_t | \mathfrak{M}_{t-1}(X)).$$

Note that the above two variances are not dependent on an integer t because of the weak stationarity. Granger (1969) originally defined the causality by means of $\mathscr{M}_{t-1}(X)$ and $\mathscr{M}_{t-1}(X, Y)$ instead of $\mathfrak{M}_{t-1}(X)$ and $\mathfrak{M}_{t-1}(X, Y)$, but he only considered a linear predictor, thus his causality is essentially "linear" and can be represented as above.

3. Nonlinear causality

The simplest way of extending the Granger causality to nonlinear one might be replacing a linear predictor with a nonlinear predictor in the definition as follows.

DEFINITION 3.1. For a two dimensional strictly stationary time series $\{(X_t, Y_t)\}$ with finite second order moments, it is said that $\{Y_t\}$ *causes* $\{X_t\}$ *nonlinearly in the sense of Granger*, written $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$, if and only if

$$\sigma^2(X_t | \mathscr{M}_{t-1}(X, Y)) < \sigma^2(X_t | \mathscr{M}_{t-1}(X)).$$

We also say that $\{Y_t\}$ *does not causes* $\{X_t\}$ *nonlinearly in the sense of Granger*, written $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$, if and only if

$$\sigma^2(X_t | \mathscr{M}_{t-1}(X, Y)) = \sigma^2(X_t | \mathscr{M}_{t-1}(X)).$$

Note that the above two quantities are not dependent on an integer t because of the strict stationarity. We assume that $\{(X_t, Y_t)\}$ is strictly stationary throughout this paper. It is a well-known fact that

$$P_{\mathscr{M}_{t-1}(X,Y)} X_t = E[X_t | \mathscr{F}_{t-1}(X, Y)],$$

and

$$P_{\mathscr{M}_{t-1}(X)} X_t = E[X_t | \mathscr{F}_{t-1}(X)].$$

Thus if $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ we have

$$Var[X_t - E[X_t|\mathscr{F}_{t-1}(X,Y)]] = Var[X_t - E[X_t|\mathscr{F}_{t-1}(X)]].$$

This equation is equivalent to the following.

$$E[E[X_t|\mathscr{F}_{t-1}(X,Y)]^2] = E[E[X_t|\mathscr{F}_{t-1}(X)]^2].$$

Note that

$$E[E[X_t|\mathscr{F}_{t-1}(X,Y)]|\mathscr{F}_{t-1}(X)] = E[X_t|\mathscr{F}_{t-1}(X)] \quad a.e.$$

since $\mathscr{F}_{t-1}(X) \subset \mathscr{F}_{t-1}(X,Y)$, and so we have

$$E[(E[X_t|\mathscr{F}_{t-1}(X)] - E[X_t|\mathscr{F}_{t-1}(X,Y)])^2] = 0,$$

that is,

$$E[X_t|\mathscr{F}_{t-1}(X)] = E[X_t|\mathscr{F}_{t-1}(X,Y)] \quad a.e. \tag{3.1}$$

Thus (3.1) is the necessary and sufficient condition for $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$.

Let us consider the following model

$$X_t = F(X_{t,d_o}, Y_{t,d_o}) + \varepsilon_t, \tag{3.2}$$

where ε_t is an additive white Gaussian noise, $X_{t,d} = (X_{t-1}, \ldots, X_{t-d})'$ and $Y_{t,d} = (Y_{t-1}, \ldots, Y_{t-d})'$ for any integer d, and F is a function from $\mathbf{R}^{2d_o}$ into $\mathbf{R}$. It should be mentioned that F is usually unknown in a practical case. d_o is the non-negative integer satisfying the condition of the order of nonlinear autoregressive (NLAR) model in Cheng and Tong (1992). Namely, d_o satisfies the following.

(i) $\sigma^2(X_t|\mathscr{M}_{t-d_o,t-1}(X,Y)) < \sigma^2(X_t|\mathscr{M}_{t-d,t-1}(X,Y))$ for any $d < d_o$.

(ii) $\sigma^2(X_t|\mathscr{M}_{t-d_o,t-1}(X,Y)) = \sigma^2(X_t|\mathscr{M}_{t-d,t-1}(X,Y))$ for any $d > d_o$.

We say that $d_o = 0$ if and only if $Var[X_t] = \sigma^2(X_t|\mathscr{M}_{t-1}(X,Y))$, and in this case F is constantly 0.

REMARK. For example, $d_o = 2$ in the following model.

$$X_t = -0.2X_{t-1} + 1.2Y_{t-1} - 1.3X_{t-2} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0,1).$$

Under model (3.2), equation (3.1) is equivalent to

$$E[X_t|\mathscr{F}_{t-1}(X)] = E[X_t|\mathscr{F}_{t-d_o,t-1}(X,Y)] \quad a.e. \tag{3.3}$$

from the property of d_o. We will prove the next theorem in Section 6. The conditions (A1) to (A7) will be given in the same section.

THEOREM 3.2. *If conditions* (A1) *to* (A7) *hold,*

$$\{Y_t\} \xrightarrow{NLGC} \{X_t\} \Longleftrightarrow {}^\exists G : \mathbf{R}^{d_o} \to \mathbf{R}, \text{ Lipschitz continuous,}$$
$$s.t. \quad F(X_{t,d_o}, Y_{t,d_o}) = G(X_{t,d_o}) \quad a.e.$$

It should be noted that $F(X_{t,d_o}, Y_{t,d_o})$ can be written as a function of $X_{t-1}, X_{t-2}, \ldots,$ if $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ by the definition of nonlinear non-causality without any conditions. The important point of this theorem is that if $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$, $F(X_{t,d_o}, Y_{t,d_o})$ can be written as a Lipschitz continuous function of $X_{t-1}, X_{t-2}, \ldots, X_{t-d_o}$ if the conditions hold. The theorem shows that equation (3.3) is equivalent to

$$E[X_t|\mathscr{F}_{t-d_o,t-1}(X)] = E[X_t|\mathscr{F}_{t-d_o,t-1}(X,Y)] \quad a.e. \tag{3.4}$$

under appropriate conditions. Therefore, for the model (3.2) satisfying the conditions, we have

$$\{Y_t\} \xrightarrow{NLGC} \{X_t\}$$
$$\Longleftrightarrow \quad \sigma^2(X_t|\mathscr{M}_{t-d_o,t-1}(X,Y)) = \sigma^2(X_t|\mathscr{M}_{t-d_o,t-1}(X))$$
$$\Longleftrightarrow \quad Var[X_t - E[X_t|\mathscr{F}_{t-d_o,t-1}(X,Y)]]$$
$$= Var[X_t - E[X_t|\mathscr{F}_{t-d_o,t-1}(X)]]. \tag{3.5}$$

As Cheng and Tong (1992, 1993), conditional expectations in (3.5) could be estimated by kernel estimators and variances of residual could be estimated by cross-validations. We conclude whether $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ or not by comparing the estimated variances. The ideas are briefly sketched in the next section.

4. Methods of nonlinear causality analysis

We estimate the conditional expectations

$$F(X_{t,d}, Y_{t,d}) = E[X_t|\mathscr{F}_{t-d,t-1}(X,Y)],$$

and

$$G(X_{t,d}) = E[X_t|\mathscr{F}_{t-d,t-1}(X)],$$

by leave-one-out kernel estimators

$$\hat{F}_{\backslash t}(X_{t,d}, Y_{t,d}) = \frac{\sum X_t K_{2d}(\|(X_{t,d}, Y_{t,d}) - (X_{s,d}, Y_{s,d})\|/B_{2d,N})}{\sum K_{2d}(\|(X_{t,d}, Y_{t,d}) - (X_{s,d}, Y_{s,d})\|/B_{2d,N})},$$

and

$$\hat{G}_{\backslash t}(X_{t,d}) = \frac{\sum X_t K_d(\|(X_{t,d} - X_{s,d})\|/B_{d,N})}{\sum K_d(\|(X_{t,d} - X_{s,d})\|/B_{d,N})},$$

respectively, where summations are taken over s from r to N except t, N is a sample size and r is a positive integer larger than d, K_n is a non-negative even kernel from $\mathbb{R}^n$ into $\mathbb{R}$ such that

$$K_n(u) = \prod_{j=1}^{n} k(u_j), \quad u = (u_1, \ldots, u_n), \quad \int_{\mathbb{R}} k(x)dx = 1,$$

and $B_{n,N}$ is a bandwidth of that kernel. Also we estimate the variances of residual $\sigma^2(X_t|\mathscr{M}_{t-d,t-1}(X,Y))$ and $\sigma^2(X_t|\mathscr{M}_{t-d,t-1}(X))$ by cross-validations

$$CV_1(d) = \frac{1}{N-r+1}\sum_{t=r}^{N} \left(X_t - \hat{F}_{\backslash t}(X_{t,d}, Y_{t,d})\right)^2 W_{1,d}(X_{t,d}, Y_{t,d}),$$

and

$$CV_0(d) = \frac{1}{N-r+1}\sum_{t=r}^{N}\left(X_t - \hat{G}_{\backslash t}(X_{t,d})\right)^2 W_{0,d}(X_{t,d}),$$

respectively. $W_{1,d}$ and $W_{0,d}$ are weight functions. We propose the following methods for the analysis of nonlinear causality. Denote by Min_B a minimization with respect to a bandwidth and Min_d a minimization with respect to d within $0 \leq d \leq L$, where L is a pre-specified positive integer.

(a) Compute $\text{Min}_d\text{Min}_B CV_1(d)$, say CV_1, and let $\hat{d}_1$ be the argument which gives the minimum of $\text{Min}_B CV_1(d)$ with respect to d ($0 \leq d \leq L$).

(b) Compute $\text{Min}_B CV_0(\hat{d}_1)$, say CV_0.

(c) If $CV_0 - CV_1 > 0$ we conclude $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$, otherwise we conclude $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$.

We call this method as the NLCA (nonlinear causality analysis). The method which replaces (b) by (b') given as follows is called the CNLCA (conservative nonlinear causality analysis).

(b') Compute $\text{Min}_d\text{Min}_B CV_0(d)$, say CV_0.

It should be mentioned that our methods are not statistical testing and we can not control error probabilities. Our methods are kinds of model selection. We select either "causal model" or "non-causal model" based on CV_0 and CV_1. The justification of these methods is given by the next theorem. All conditions and the proof of the theorem are given in Section 6.

THEOREM 4.1. *Suppose $d_o \leq L$ and X_t and Y_t are bounded. Let S_{x_d} and S_{x_d,y_d} be the supports of $X_{t,d}$ and $(X_{t,d}, Y_{t,d})$ respectively. Put $W_{0,d}(x) = I_{S_{x_d}}(x)$ and $W_{1,d}(x,y) = I_{S_{x_d,y_d}}(x,y)$, where I_S denotes the indicator function on S. If conditions (A1) to (A17) hold, we have*

$$(P1) \qquad \{Y_t\} \xrightarrow{NLGC} \{X_t\} \implies \lim_{N\to\infty} \Pr(CV_0 - CV_1 > 0) = 0,$$

in both methods, the NLCA and CNLCA. Furthermore in the NLCA, we have

$$(P2) \qquad \{Y_t\} \xrightarrow{NLGC} \{X_t\} \implies \lim_{N\to\infty} \Pr(CV_0 - CV_1 > 0) = 1.$$

Note that the supports S_{x_d} and S_{x_d,y_d} are not dependent on t because of the strict stationarity. Even though our methods are not statistical testing, the probability in (P1) corresponds to the probability of the type I error in testing the statistical null hypothesis $H_0 : \{Y_t\} \xrightarrow{NLGC} \{X_t\}$ versus the alternative $H_1 : \{Y_t\} \xrightarrow{NLGC} \{X_t\}$, and the probability in (P2) corresponds to the power of the test. From this point of view, it might be said that our methods have desirable asymptotic properties. In this theorem, we assumed that X_t and Y_t are bounded. However, as we will see in the next section, this assumption would not be essential to the theorem.

5. Simulation study

The aim of this section is to show the efficiency of our methods and to compare them with the Granger-Sargent test. As for the Granger-Sargent test, see Sargent (1976). We employed weight functions $W_{0,d}(x) = I_{S_{x_d}}(x)$ and $W_{1,d}(x,y) = I_{S_{x_d}}(x,y)$. The probability density function of a standard Gaussian random variable is used as a kernel function $k(x)$ which satisfy the conditions in Theorem 4.1. Note that $W_{0,d}(X_{t,d}) = W_{1,d}(X_{t,d}, Y_{t,d}) = 1$ a.e. for any d. We applied the NLCA method, CNLCA method, and the Granger-Sargent test with level 0.05, which is denoted by GS(0.05), and 0.3, which is denoted by GS(0.3), to the data from the following models.

$$\text{(M1)} \quad \begin{cases} X_t = -(1/12)X_{t-1} + 1.2Y_{t-1} + 0.5X_{t-2} - 0.3Y_{t-2} + \varepsilon_t, \\ Y_t = 1.4Y_{t-1} - 0.8Y_{t-2} + \eta_t, \end{cases}$$

$$\text{(M2)} \quad \begin{cases} X_t = 0.5X_{t-1}Y_{t-1} - 0.5X_{t-2} + \varepsilon_t, \\ Y_t = 0.8Y_{t-1} - 0.2Y_{t-2} + \eta_t, \end{cases}$$

$$\text{(M3)} \quad \begin{cases} X_t = 1.2Y_{t-1}\exp(-Y_{t-1}^2) + \varepsilon_t, \\ Y_t = 0.3Y_{t-1} - 0.95Y_{t-2}\exp(-Y_{t-2}^2) + \eta_t, \end{cases}$$

$$\text{(M4)} \quad \begin{cases} X_t = 0.5\text{Sign}(X_{t-1}Y_{t-1})\sqrt{|X_{t-1}Y_{t-1}|} - 0.5X_{t-2} + \varepsilon_t, \\ Y_t = -0.9Y_{t-1} - 0.95Y_{t-2}\exp(-Y_{t-2}^2) + \eta_t, \end{cases}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are sequences of independent standard Gaussian random variables.

Let us look into each model in detail before giving the results. Only model (M1) is a linear model and others are nonlinear models. In each model, $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ while $\{X_t\} \xrightarrow{NLGC} \{Y_t\}$. It is difficult to know whether there exist the Granger causality in each model except for model (M1). In model (M1), $\mathfrak{M}_{t-1}(X)$, $\mathfrak{M}_{t-1}(Y)$ and $\mathfrak{M}_{t-1}(X,Y)$ are identical to $\mathscr{M}_{t-1}(X)$, $\mathscr{M}_{t-1}(Y)$ and $\mathscr{M}_{t-1}(X,Y)$ respectively. Thus the nonlinear Granger causality is the same as the Granger causality in this model, that is, $\{Y_t\} \xrightarrow{GC} \{X_t\}$ and $\{X_t\} \xrightarrow{GC} \{Y_t\}$. In model (M2), (M3), and (M4), we have to compute $\sigma^2(X_t|\mathfrak{M}_{t-1}(X,Y))$, $\sigma^2(X_t|\mathfrak{M}_{t-1}(X))$ and $\sigma^2(Y_t|\mathfrak{M}_{t-1}(Y))$ in order to know whether there exists the Granger causality theoretically. If the second order moment of X_t and Y_t are finite, that is, $\{(X_t, Y_t)\}$ is weakly stationary, we can compute these quantities from its autocovariance function, $R(h)$ say. (See Section 5.1 and Section 11.4 of Brockwell and Davis (1991).) However, we do not know $R(h)$ in each model, so we use the estimator

$$\frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{t=1}^{N-h}\begin{pmatrix} (X_{t+h}^{(i)} - \overline{X_N^{(i)}})(X_t^{(i)} - \overline{X_N^{(i)}}) & (X_{t+h}^{(i)} - \overline{X_N^{(i)}})(Y_t^{(i)} - \overline{Y_N^{(i)}}) \\ (Y_{t+h}^{(i)} - \overline{Y_N^{(i)}})(X_t^{(i)} - \overline{X_N^{(i)}}) & (Y_{t+h}^{(i)} - \overline{Y_N^{(i)}})(Y_t^{(i)} - \overline{Y_N^{(i)}}) \end{pmatrix},$$

say $\hat{R}(h)$, instead of $R(h)$, and approximate variances $\sigma^2(X_t|\mathfrak{M}_{t-1}(X))$, $\sigma^2(Y_t|\mathfrak{M}_{t-1}(Y))$ and $\sigma^2(X_t|\mathfrak{M}_{t-1}(X,Y))$. Here $\{X_t^{(i)}\}$ and $\{Y_t^{(i)}\}$ $(i = 1, 2, \ldots, M)$ are the multiple independent copies of $\{X_t\}$ and $\{Y_t\}$ respectively, $\overline{X_N^{(i)}}$ and $\overline{Y_N^{(i)}}$ are the sample means of $\{X_t^{(i)}\}$ and $\{Y_t^{(i)}\}$ $(t = 1, 2, \ldots, N)$ respectively. If the diagonal elements of $R(h)$ converge to 0 as $h \to \infty$, we have $\lim_{N,M \to \infty} \hat{R}(h) = R(h)$ by the law of large number and Proposition 11.2.1 of Brockwell and Davis (1991). Table 1, Table 2, Table 3, and

Table 4 list the approximated values of $\sigma^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$ and others for some k which are differentiated with "$\,\hat{}\,$". $M = 1000$ and $N = 1000$ are employed in these approximations.

k	1	2	3	4	5	10
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.879	1.019	1.019	1.019	1.019	1.019
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X))$	5.261	5.254	3.419	3.419	3.363	3.331
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.593	1.018	1.018	1.018	1.018	1.018
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$	2.784	1.018	1.018	1.018	1.018	1.018

Table 1: The approximations of the variances of the best linear prediction errors in model (M1).

k	1	2	3	4	5	10
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	3.105	3.073	3.071	3.067	3.067	3.066
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X))$	3.105	3.073	3.071	3.067	3.067	3.066
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.042	1.000	1.000	1.000	1.000	1.000
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$	1.042	1.000	1.000	1.000	1.000	1.000

Table 2: The approximations of the variances of the best linear prediction errors in model (M2).

k	1	2	3	4	5	10
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.077	1.077	1.077	1.077	1.077	1.077
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X))$	1.122	1.122	1.122	1.122	1.122	1.122
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.067	1.045	1.045	1.045	1.045	1.045
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$	1.075	1.048	1.048	1.048	1.048	1.048

Table 3: The approximations of the variances of the best linear prediction errors in model (M3).

k	1	2	3	4	5	10
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	3.863	1.462	1.462	1.462	1.462	1.462
$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X))$	3.863	1.462	1.462	1.462	1.462	1.462
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.055	1.053	1.053	1.053	1.053	1.053
$\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$	1.055	1.053	1.053	1.053	1.053	1.053

Table 4: The approximations of the variances of the best linear prediction errors in model (M4).

We can compute the exact values of $\sigma^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$ and other variances in model (M1) which are given in Table 5. Table 1 and Table 5 indicate that our approximations are fairly good, so we may assess the existence of the Granger causality by comparing

$$\hat{\sigma}^2(X_t|\,\mathfrak{M}_{t-k,t-1}(X,Y)) \quad \text{and} \quad \hat{\sigma}^2(Y_t|\,\mathfrak{M}_{t-k,t-1}(X,Y)),$$

with

$$\hat{\sigma}^2(X_t|\mathfrak{M}_{t-k,t-1}(X)) \quad \text{and} \quad \hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y)),$$

respectively for sufficiently large k. From Table 2 and Table 4, we may assess that $\{Y_t\} \xrightarrow{GC} \{X_t\}$ and $\{X_t\} \xrightarrow{GC} \{Y_t\}$ in model (M2) and (M4). Indeed, $\{X_t\} \xrightarrow{GC} \{Y_t\}$ in model (M2) theoretically. Table 3 suggests that the Granger causality from $\{Y_t\}$ to $\{X_t\}$ might exist in model (M3). Also the table shows that the difference between $\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$ and $\hat{\sigma}^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$ is very small, thus we avoid assessing $\{X_t\} \xrightarrow{GC} \{Y_t\}$ or $\{X_t\} \xrightarrow{GC} \{Y_t\}$.

k	1	2	3	4	5	10
$\sigma^2(X_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.867	1	1	1	1	1
$\sigma^2(X_t|\mathfrak{M}_{t-k,t-1}(X))$	5.256	5.248	3.405	3.404	3.346	3.312
$\sigma^2(Y_t|\mathfrak{M}_{t-k,t-1}(X,Y))$	1.587	1	1	1	1	1
$\sigma^2(Y_t|\mathfrak{M}_{t-k,t-1}(Y))$	2.778	1	1	1	1	1

Table 5: The exact variances of the best linear prediction errors in model (M1).

We now give the results of the simulation of our methods and the Granger-Sargent test for our models. The sample sizes employed in the simulation are $N = 60, 110, 210$ and 410. $r = 10$ and $L = 5$ are employed. Number of iteration is 3000 for $N = 60, 110, 210$, and 1000 for $N = 410$. Table 6, Table 7, Table 8, and Table 9 show the proportions of the trials that concluded there existed the causality from one to another. Each left half part of the table is for $\{Y_t\} \to \{X_t\}$ and each right half part of the table is for $\{X_t\} \to \{Y_t\}$. As we have mentioned, our methods are not statistical testing. Even so, since $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ and $\{X_t\} \xrightarrow{NLGC} \{Y_t\}$ in all models, the values in the left half parts of the tables might be considered as the "empirical powers," and those in the right half parts as the "empirical probabilities of the type I errors" from the viewpoint of testing statistical hypothesis of nonlinear causality. Therefore, it might be said that the result is better if the values in the left half parts are closer to 1, and those values in the right half parts are closer to 0.

N	$\{Y_t\} \to \{X_t\}$				$\{X_t\} \to \{Y_t\}$			
	60	110	210	410	60	110	210	410
NLCA	1	1	1	1	.622	.582	.355	.078
CNLCA	1	1	1	1	.139	.03	.000	0
GS(0.3)	1	1	1	1	.124	.122	.117	.088
GS(0.05)	1	1	1	1	.013	.012	.008	.007

Table 6: The ratios of trials in model (M1) that concluded there was causality.

N	$\{Y_t\} \to \{X_t\}$				$\{X_t\} \to \{Y_t\}$			
	60	110	210	410	60	110	210	410
NLCA	.898	.988	.996	.999	.283	.194	.091	.057
CNLCA	.865	.983	.995	.999	.199	.134	.062	.036
GS(0.3)	.239	.266	.315	.339	.089	.106	.099	.084
GS(0.05)	.058	.073	.108	.136	.009	.01	.013	.006

Table 7: The ratios of trials in model (M2) that concluded there was causality.

N	$\{Y_t\} \to \{X_t\}$				$\{X_t\} \to \{Y_t\}$			
	60	110	210	410	60	110	210	410
NLCA	.542	.77	.964	1	.383	.426	.349	.193
CNLCA	.513	.746	.955	1	.292	.297	.179	.068
GS(0.3)	.367	.587	.851	.986	.123	.148	.222	.315
GS(0.05)	.101	.232	.528	.894	.015	.021	.036	.078

Table 8: The ratios of trials in model (M3) that concluded there was causality.

N	$\{Y_t\} \to \{X_t\}$				$\{X_t\} \to \{Y_t\}$			
	60	110	210	410	60	110	210	410
NLCA	.871	.957	.997	1	.106	.046	.016	.008
CNLCA	.833	.944	.997	1	.072	.036	.01	.007
GS(0.3)	.044	.027	.029	.021	.098	.103	.107	.104
GS(0.05)	.001	.001	.002	.001	.01	.011	.012	.014

Table 9: The ratios of trials in model (M4) that concluded there was causality.

Table 6 shows that, in model (M1), every methods have detected the causality completely, and the Granger-Sargent test has controlled the type I error less than the level. This is a natural consequent because all assumptions needed for the Granger-Sargent test are satisfied in this model. The table also shows that the proportions of the trials that concluded $\{X_t\} \xrightarrow{NLGC} \{Y_t\}$ incorrectly in the NLCA method are too large when the sample size is small. This might be caused by the poor estimation of d_o.

Table 7, Table 8, and Table 9 indicate that the Granger-Sargent test is unable to detect the nonlinear causality. Table 8 indicates that the Granger-Sargent test detected the causality from $\{Y_t\}$ to $\{X_t\}$ in model (M3). However, as we have discussed in the above, there might be linear causality from $\{Y_t\}$ to $\{X_t\}$ in model (M3), so it might be said that the Granger-Sargent test detected linear causality, not nonlinear causality. We may say that all results of the Granger-Sargent test agree with our previous observation of the existence of linear causality.

On the other hands, the NLCA and CNLCA detected the nonlinear causality sufficiently well with the acceptable ratios of miss-specification, that is, the proportion of the trials that concluded $\{X_t\} \xrightarrow{NLGC} \{Y_t\}$ incorrectly, in model (M2), (M3), and (M4) are relatively small.

It should be noted that the values in the left half parts of Table 6, Table 7, Table 8 and Table 9 are getting closer to 1 and the values in the right half parts are getting to 0 as the increase in the sample size. This indicates that (P1) and (P2) in Theorem 4.1 might be true in all models even though $\{X_t\}$ and $\{Y_t\}$ are not bounded. In conclusion, the simulation study shows that our methods would be applicable to various data. Further discussion is given in Section 7.

6. Proof of the theorems

This section is devoted to prove Theorem 3.2 given in Section 3. and Theorem 4.1 given in Section 4. To begin with, we state all conditions needed to prove them. Let $Z_{t,d}$ denote $(X_{t,d}, Y_{t,d})$ and recall that our model is NLAR model; $X_t = F(X_{t,d_o}, Y_{t,d_o}) + \varepsilon_t$.

(A1) $\cap_{t=-\infty}^{\infty} \mathscr{F}_t(X, Y) = \{\varnothing, \Omega\}$.

(A2) $\{\varepsilon_t\}$ is an independent random sequence with mean 0 and strictly positive variance $\sigma_1{}^2$.

(A3) F is Lipschitz continuous function from $\mathbb{R}^{2d_o}$ in $\mathbb{R}$.

(A4) The support of $(X_{t,d_o}, Y_{t,k})$ is convex for any $k = 1, \ldots, d_o$.

(A5) Put $\eta_t = Y_t - E[Y_t | \mathscr{F}_{t-1}(X, Y)]$ then $\{\eta_t\}$ is an independent sequence with mean 0 and strictly positive variance $\sigma_2{}^2$.

(A6) ε_t and η_s are independent for any t and s.

(A7) The support of η_t is connected.

(A8) $\int_{\mathbb{R}} |u| k(u) du < \infty$.

(A9) For $1 \leq d \leq L$, $X_{t,d}$ and $Z_{t,d}$ have density functions g_d and f_d which are Lipschitz continuous respectively.

(A10) A kernel function k is Lipschitz continuous and has a compact support.

(A11) For every $t, s, \tau, t', s', \tau' \in \mathbb{N}$ and $1 \leq d \leq L$, the joint probability density functions of $(X_{t,d}, X_{s,d}, X_{\tau,d}, X_{t',d}, X_{s',d}, X_{\tau',d})$ and $(Z_{t,d}, Z_{s,d}, Z_{\tau,d}, Z_{t',d}, Z_{s',d}, Z_{\tau',d})$ are bounded. If, for example, $t = t'$, the assumption requires that the joint probability density function of $(X_{t,d}, X_{s,d}, X_{\tau,d}, X_{s',d}, X_{\tau',d})$ exists and is bounded.

(A12) Let $p^{-1} + q^{-1} = 1$. For some $p > 2$ and $\delta > 0$ such that $\delta < (2/q) - 1$,

$$E|\varepsilon_1|^{2p(1+\delta)} < \infty \quad \text{and} \quad E|F(Z_{t,d_o})|^{2p(1+\delta)} < \infty.$$

(A13) For δ in (A11), $\beta_j^{\delta/(1+\delta)} = O(j^2)$, where

$$\beta_j = \sup_{i \in \mathbb{N}} E\left[\sup_{A \in \mathscr{F}_{i+j,\infty}(X,Y)} \{ |\Pr(A | \mathscr{F}_{1,i}(X,Y)) - \Pr(A)| \} \right].$$

(A14) Let $j = j(N)$ be a positive integer and $i = i(N)$ be the largest positive integer such that $2ij \leq N$,

$$\limsup_{N \to \infty} (1 + 6e^{1/2} \beta_j^{1/(1+i)})^i < \infty.$$

(A15) For $i = i(N)$ in (A14) and the bandwidth $B_{k,N}$,

$$\limsup_{N \to \infty} \{ i(N) B_{k,N}{}^k \} < \infty \quad (1 \leq k \leq 2L).$$

(A16) $N B_{k,N}{}^{2k} \to \infty$ as $N \to \infty$ $\quad (1 \leq k \leq 2L)$.

(A17) $N B_{k,N}{}^{2(k+1)} \to 0$ as $N \to \infty$ $\quad (1 \leq k \leq 2L)$.

Note that the above conditions hold for all t if they hold for a certain t_o by the strict stationarity of $\{(X_t, Y_t)\}$.

6.1. Preparation

In this subsection, we prepare some mathematical idea that is used in the proof of the theorems. We first define a support of a random vector and show some lemmas about it. Next we define the projection onto a closed convex set, not closed subspace, in $\mathbf{R}^n$ and show some lemmas. Both of them are very general idea.

6.1.1. Support of a random vector

DEFINITION 6.1. For a random vector $X \in \mathbf{R}^n$, the *support* of X is the set

$$S_X = \left\{ x \in \mathbf{R}^n \middle|^{\forall} \varepsilon > 0,\ \Pr(X \in B(x; \varepsilon)) > 0 \right\},$$

where $B(x; \varepsilon)$ is the ε-neighborhood of x.

It is easy to prove the following three lemmas.

LEMMA 6.2. $\Pr(X \in S_X{}^c) = 0$.

LEMMA 6.3. *A support S_X is closed.*

LEMMA 6.4. *Let X and Y be random vectors in $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively. If S_X, S_Y and $S_{X,Y}$ are the supports of X, Y, and (X, Y) respectively, $S_{X,Y} \subset S_X \times S_Y$. Furthermore if X and Y are independent, $S_{X,Y} = S_X \times S_Y$.*

LEMMA 6.5. *For a continuous function g from $\mathbf{R}^n$ into $\mathbf{R}$,*

$$g(X) = 0 \ a.e. \iff g(x) = 0 \quad \text{for any} \quad x \in S_X.$$

PROOF.

$(\Longrightarrow)$
Suppose there exists $x \in S_X$ such that $|g(x)| > 0$. Since g is continuous, there exists $\varepsilon > 0$ such that

$$\|u - x\| < \varepsilon \Longrightarrow |g(u)| > |g(x)|/2.$$

So we have

$$\begin{aligned}
\Pr(X \in B(x; \varepsilon)) &= \Pr(\|X - x\| < \varepsilon) \\
&\leq \Pr(|g(X)| > |g(x)|/2) \leq \Pr(|g(X)| > 0).
\end{aligned}$$

Since $g(X) = 0$ *a.e.*, $\Pr(|g(X)| > 0) = 0$. So we have $\Pr(X \in B(x; \varepsilon)) = 0$, but it contradicts to $x \in S_X$, therefore $g = 0$ on S_X.

$(\Longleftarrow)$
It is straightforward from Lemma 6.2 to see that

$$\begin{aligned}
&\Pr(|g(X)| > 0) \\
&= \Pr(|g(X)| > 0, X \in S_X) + \Pr(|g(X)| > 0, X \in S_X{}^c) \\
&\leq \Pr(\varnothing) + \Pr(X \in S_X{}^c) = 0.
\end{aligned}$$

LEMMA 6.6. *Let X be a random vector in $\mathbb{R}^n$ and Y, Z be random variables in* $\mathbb{R}$. *Put $T = Y + Z$ and let $S_{X,Y,Z}$ and $S_{X,T}$ be the supports of $\{X, Y, Z\}$ and $\{X, T\}$ respectively. Define a function h from $\mathbb{R}^{n+2}$ into $\mathbb{R}$ as follows.*

$$h(x, y, z) = (x, y + z) \quad (x \in \mathbb{R}^n, y, z \in \mathbb{R}).$$

If Z and (X, Y) are independent, $S_{X,T} = \overline{h(S_{X,Y,Z})}$.

PROOF.

(i) $S_{X,T} \subset \overline{h(S_{X,Y,Z})}$.

Suppose that there exists a point (x_0, t_0) in $S_{X,T} \cap \overline{h(S_{X,Y,Z})}^c$. From Lemma 6.4 we have $S_{X,Y,Z} = S_{X,Y} \times S_Z$, where $S_{X,Y}$ and S_Z are the supports of (X, Y) and Z respectively. (x_0, t_0) is thus in $h(S_{X,Y} \times S_Z)^c$, that is, $(x_0, t_0 - z) \in S_{X,Y}{}^c$ for any $z \in S_Z$. By the definition of a support, there exists $\delta(z) > 0$ such that

$$\Pr((X, Y) \in B((x_0, t_0 - z); \delta(z))) = 0.$$

Put $\overline{\delta(z)} = \sup\{\delta(z) | \Pr((X, Y) \in B((x_0, t_0 - z); \delta(z))) = 0\}$ and $\delta_0 = \inf_{z \in S_Z} \overline{\delta(z)}$. We first show that $\delta_0 > 0$.

Suppose $\delta_0 = 0$ then there exists the sequence $\{z_n\}$ in S_Z such that $\overline{\delta(z_n)} < 1/n$. Therefore we have

$$\Pr[(X, Y) \in B((x_0, t_0 - z); 1/n)] > 0,$$

that is, $B((x_0, t_0 - z); 1/n) \cap S_{X,Y} \neq \varnothing$. Let (x_n, y_n) be an element of this intersection then we have

$$\|(x_0, t_0) - h(x_n, y_n, z_n)\| = \|(x_0 - x_n, t_0 - y_n - z_n)\|$$
$$= \|(x_0, t_0 - z_n) - (x_n, y_n)\| < 1/n \xrightarrow{n \to \infty} 0.$$

Since $(x_n, y_n, z_n) \in S_{X,Y} \times S_Z$, we have

$$(x_0, t_0) = \lim_{n \to \infty} h(x_n, y_n, z_n) \in \overline{h(S_{X,Y,Z})}.$$

It is a contradiction. Thus $\delta_0 > 0$ and by the double expectation theorem and the assumption that Z and (X, Y) are independent, we have

$$0 < \Pr((X, T) \in B((x_0, t_0); \delta_0/2))$$
$$= \int_{S_Z} \Pr((X, Y + z) \in B((x_0, t_0); \delta_0/2)) dF_Z(z)$$
$$= \int_{S_Z} \Pr((X, Y) \in B((x_0, t_0 - z); \delta_0/2)) dF_Z(z), \tag{6.6}$$

where F_Z is the distribution function of Z. We also have that

$$^{\exists}\delta_z \geq \overline{\delta(z)} - \delta_0/2 \text{ such that } \Pr((X, Y) \in B((x_0, t_0 - z); \delta_z)) = 0,$$

for any $z \in S_Z$. Therefore we have, from $\overline{\delta(z)} > \delta_0$,

$$\Pr((X, Y) \in B((x_0, t_0 - z); \delta_0/2))$$
$$\leq \Pr((X, Y) \in B((x_0, t_0 - z); \overline{\delta(z)} - \delta_0/2))$$
$$\leq \Pr((X, Y) \in B((x_0, t_0 - z); \delta_z)) = 0,$$

for any $z \in S_Z$. It contradict to the inequality (6.6). So we have $S_{X,Y} \subset \overline{h(S_{X,Y,Z})}$.

(ii) $\overline{h(S_{X,Y,Z})} \subset S_{X,T}$.

For any $(x,t) \in \overline{h(S_{X,Y,Z})}$, there exist the sequences $\{(x_n, y_n)\}$ in $S_{X,Y}$ and $\{z_n\}$ in S_Z such that $h(x_n, y_n, z_n)$ converges to (x,t). Thus for any $\varepsilon > 0$, there exists $M \in \mathbb{N}$ such that

$$n > M \implies \|(x,t) - (x_n, y_n + z_n)\| < \varepsilon/2.$$

Thus for $n > M$, we have

$$
\begin{aligned}
&\Pr((X,T) \in B((x,t); \varepsilon)) \\
&\geq \Pr((X,Y) \in B((x_n, y_n); \varepsilon/4), Z \in B(z_n; \varepsilon/4)) \\
&= \Pr((X,Y) \in B((x_n, y_n); \varepsilon/4)) \Pr(Z \in B(z_n; \varepsilon/4)) > 0.
\end{aligned}
$$

Namely, (x,t) is in $S_{X,T}$, so we have $\overline{h(S_{X,Y,Z})} \subset S_{X,T}$.

LEMMA 6.7. *Under the same conditions as Lemma* 6.6,

$$S_Z \text{ is compact} \implies S_{X,T} = h(S_{X,Y,Z}).$$

PROOF. It suffices to show that $\overline{h(S_{X,Y,Z})} \subset h(S_{X,Y,Z})$ if S_Z is compact. For any $(x,t) \in \overline{h(S_{X,Y,Z})}$, there exist the sequences $\{(x_n, y_n)\}$ in $S_{X,Y}$ and $\{z_n\}$ in S_Z such that $\lim_{n\to\infty} h(x_n, y_n, z_n) = (x,t)$. Since S_Z is compact, there exist subsequence $\{z_{i_n}\} \subset S_Z$ and $z \in S_Z$ such that $\lim_{n\to\infty} z_{i_n} = z$. Therefore we have

$$
\begin{aligned}
|y_{i_n} - (t - z)| &= |y_{i_n} + z_{i_n} - t + z - z_{i_n}| \\
&\leq |y_{i_n} + z_{i_n} - t| + |z - z_{i_n}| \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

Since $(x_{i_n}, y_{i_n}) \in S_{X,Y}$ and $S_{X,Y}$ is closed, we have $(x, t - z) \in S_{X,Y}$. Therefore

$$(x,t) = h(x, t - z, z) \in h(S_{X,Y} \times S_Z),$$

that is, $\overline{h(S_{X,Y,Z})} \subset h(S_{X,Y,Z})$.

6.1.2. Projection onto a closed convex set in $\mathbb{R}^n$

We define the projection onto a closed convex set in this subsection. The definition is the same as the projection onto a closed space, but the orthogonality of the projection does not work in our case.

DEFINITION 6.8. Let M be a closed convex set in $\mathbb{R}^n$. The *projection* of $x \in \mathbb{R}^n$ onto M is the element $\hat{x} \in M$ such that $\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$.

We have to show that the definition is well-defined, but it is easy to see by mimicking the proof of Theorem 2.3.1 of of Brockwell and Davis (1991).

LEMMA 6.9. *Let* $\hat{x}$ *be the projection of* x *onto a closed convex set* M. *For any* $y \in M$, $\langle x - \hat{x}, y - \hat{x} \rangle \leq 0$.

PROOF. Suppose that $\langle x - \hat{x}, y - \hat{x} \rangle > 0$ and put $d = \min\{\|x - \hat{x}\|, \|y - \hat{x}\|\}$. If $d = 0$, it shows $\langle x - \hat{x}, y - \hat{x} \rangle = 0$. So we can assume $d > 0$. Put

$$z = \frac{d}{\|x - \hat{x}\|} x + \left(1 - \frac{d}{\|x - \hat{x}\|}\right) \hat{x},$$

and

$$y' = \hat{x} + \frac{d}{\|y - \hat{x}\|} \frac{\langle x - \hat{x}, y - \hat{x} \rangle}{\|x - \hat{x}\| \|y - \hat{x}\|} (y - \hat{x}).$$

Since $\hat{x}, y \in M$ and

$$0 < \frac{d}{\|y - \hat{x}\|} \frac{\langle x - \hat{x}, y - \hat{x} \rangle}{\|x - \hat{x}\| \|y - \hat{x}\|} \le 1,$$

we have $y' \in M$. Note that $y' \ne \hat{x}$. Then we have

$$\|z - y'\|^2 = \left\| \frac{d}{\|x - \hat{x}\|} (x - \hat{x}) - \frac{d}{\|y - \hat{x}\|} \frac{\langle x - \hat{x}, y - \hat{x} \rangle}{\|x - \hat{x}\| \|y - \hat{x}\|} (y - \hat{x}) \right\|^2$$

$$= d^2 \left\{ 1 - \left(\frac{|\langle x - \hat{x}, y - \hat{x} \rangle|}{\|x - \hat{x}\| \|y - \hat{x}\|} \right)^2 \right\} < d^2$$

and

$$\|x - z\| = \left(1 - \frac{d}{\|x - \hat{x}\|}\right) \|x - \hat{x}\| = \|x - \hat{x}\| - d.$$

Therefore

$$\|x - y'\| \le \|x - z\| + \|z - y'\|$$
$$< \|x - \hat{x}\| - d + d = \|x - \hat{x}\|.$$

It is a contradiction.

LEMMA 6.10. *Let $\hat{x}$ and $\hat{y}$ be the projections of x and y onto a closed convex set M respectively, then $\|x - y\| \ge \|\hat{x} - \hat{y}\|$.*

PROOF.

$$\|x - y\|^2$$
$$= \|\hat{x} - \hat{y}\|^2 + (\|x - \hat{x}\| - \|y - \hat{y}\|)^2$$
$$- 2\{\langle x - \hat{x}, \hat{y} - \hat{x} \rangle + \langle y - \hat{y}, \hat{x} - \hat{y} \rangle\}.$$

Since $\hat{x}, \hat{y} \in M$ we have

$$\langle x - \hat{x}, \hat{y} - \hat{x} \rangle \le 0, \quad \langle y - \hat{y}, \hat{x} - \hat{y} \rangle \le 0,$$

from Lemma 6.9. So we have

$$\|x - y\|^2 \ge \|\hat{x} - \hat{y}\|^2.$$

6.2. Proof of Theorem 3.2

We need seven lemmas to prove Theorem 3.2. We assume time series $\{X_t\}$ and $\{Y_t\}$ are strictly stationary, come from model (3.2), and $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ throughout this section.

LEMMA 6.11. *Let S_η be the support of η_t. If* (A5) *and* (A7) *hold,*

$$^{\exists}\varepsilon_o > 0 \quad s.t. \quad (-\varepsilon_o, \varepsilon_o) \subset S_\eta.$$

PROOF. Suppose that for any $\varepsilon > 0$, $(-\varepsilon, \varepsilon) \not\subset S_\eta$. From condition (A7) we have $\inf S_\eta \geq 0$ or $\sup S_\eta \leq 0$, and since $E[\eta_t] = 0$ we have $S_\eta = \{0\}$. However it is a contradiction to $Var[\eta_t] > 0$.

LEMMA 6.12. *If* (A1) *to* (A7) *hold,*

$$F(X_{t,d_o}, Y_{t,d_o}) = F(X_{t,d_o}, E[Y_{t-1}|\mathscr{F}_{t-2}(X,Y)], Y_{t-1,d_o-1}) \quad a.e.$$

PROOF. To simplify our notation, put

$$\mathfrak{F} = F(X_{t,d_o}, Y_{t,d_o})$$
$$\text{and} \quad \tilde{\mathfrak{F}} = F(X_{t,d_o}, E[Y_{t-1}|\mathscr{F}_{t-2}(X,Y)], Y_{t-1,d_o-1})$$

in this proof. It is straightforward to see that $\mathscr{F}_t(X,Y) = \mathscr{F}_t(\varepsilon, \eta)$ from our model and condition (A1), so we have, from conditions (A2) and (A5),

$$E[X_t|X_{t,d_o}, Y_{t,d_o}] = F(X_{t,d_o}, Y_{t,d_o}). \tag{6.7}$$

Since $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ and (6.7) we have

$$X_t = E[X_t|X_{t,d_o}, Y_{t,d_o}] + \varepsilon_t$$
$$= E[X_t|\mathscr{F}_{t-1}(X)] + \varepsilon_t. \tag{6.8}$$

By the similar argument to the former one we have $\mathscr{F}_t(X) = \mathscr{F}_t(\varepsilon)$, thus from conditions (A2) and (A5), X_t is independent of η_s for any t and s. Since $\mathfrak{F} = X_t - \varepsilon_t$, $\mathfrak{F}$ is independent of η_{t-1}. Furthermore, as $\tilde{\mathfrak{F}}$ is measurable with respect to $\sigma(X_{t-1}) \vee \mathscr{F}_{t-2}(X,Y)$, $\tilde{\mathfrak{F}}$ is independent of η_{t-1}, too. Therefore for any $\varepsilon > 0$ and $C > 0$,

$$\Pr(|\mathfrak{F} - \tilde{\mathfrak{F}}| > \varepsilon) = \Pr\left(|\mathfrak{F} - \tilde{\mathfrak{F}}| > \varepsilon \mid |\eta_{t-1}| < C\right).$$

From condition (A3), there exists a constant C' such that

$$|F(z) - F(z')| \leq C'\|z - z'\|.$$

So we have

$$|\mathfrak{F} - \tilde{\mathfrak{F}}| \leq C'|Y_{t-1} - E[Y_{t-1}|\mathscr{F}_{t-2}(X,Y)]| = C'|\eta_{t-1}|.$$

Therefore

$$\Pr\left(|\mathfrak{F} - \tilde{\mathfrak{F}}| > \varepsilon \mid |\eta_{t-1}| < C\right) \leq \Pr\left(C'|\eta_{t-1}| > \varepsilon \mid |\eta_{t-1}| < C\right)$$
$$= \frac{\Pr\left(\varepsilon/C' < |\eta_{t-1}| < C\right)}{\Pr\left(|\eta_{t-1}| < C\right)}. \tag{6.9}$$

By Lemma 6.11, S_η contains 0 so the denominator in (6.9) is always strictly positive for any $C > 0$. On the other hand, the numerator in the equation is 0 if $C \leq \varepsilon/C'$. We thus have $\Pr(|\mathfrak{F} - \tilde{\mathfrak{F}}| > \varepsilon) = 0$ for any $\varepsilon > 0$, that is, $\mathfrak{F} = \tilde{\mathfrak{F}}$ a.e.

Let $S_{x,y}$, $S_{x,w,y'}$ and $S_{x,w,y',\eta}$ be the supports of random vectors (X_{t,d_o}, Y_{t,d_o}), $(X_{t,d_o}, W_{t-1}, Y_{t-1,d_o-1})$ and $(X_{t,d_o}, W_{t-1}, Y_{t-1,d_o-1}, \eta_{t-1})$ respectively. Put $W_t = E[Y_t | \mathscr{F}_{t-1}(X, Y)]$. These supports are not dependent on t because of the strict stationarity. From Lemma 6.5 and Lemma 6.12, for any $(x, w, y, e) \in S_{x,w,y',\eta}$ we have

$$F(x, w + e, y) - F(x, w, y) = 0. \tag{6.10}$$

Recall that η_{t-1} is independent of $(X_{t,d_o}, W_{t-1}, Y_{t-1,d_o-1})$, so we have $S_{x,w,y',\eta} = S_{x,w,y'} \times S_\eta$ by Lemma 6.4. Thus we have the next lemma.

LEMMA 6.13. *If* (A1) *to* (A7) *hold, for any* $(x, w, y) \in S_{x,w,y'}$ *and* $e \in S_\eta$,

$$F(x, w + e, y) - F(x, w, y) = 0.$$

From this lemma we have the next lemma.

LEMMA 6.14. *Put* $T_{x,y} = \{w \in \mathbf{R} | (x, w, y) \in S_{x,w,y'}\}$. *If* (A1) *to* (A7) *hold,* $F(x, w_1, y) = F(x, w_2, y)$ *for any* $w_1, w_2 \in T_{x,y'}$.

PROOF. We can assume $w_1 < w_2$ without loss of generality. We will prove this lemma by considering several cases.

(i) When S_η is not bounded.
From Lemma 6.11, we have $w_1 - w_2 \in S_\eta$ or $w_2 - w_1 \in S_\eta$, and in each case we have $F(x, w_1, y) = F(x, w_2, y)$ straightforwardly from Lemma 6.13.

(ii) When S_η is bounded.
Since support is closed, S_η is compact. Therefore there exist the maximum and minimum of S_η, M and m respectively say. From Lemma 6.11 we have $M > 0$ and $m < 0$. Let T_1 be the connected component of $T_{x,y}$ that contains w_1 and $\{T_j\}$ be the set of connected components of $T_{x,y}$ such that

$$\bar{t}_i < \underline{t}_{i+1} \quad \text{and} \quad (\bar{t}_i, \underline{t}_{i+1}) \cap T_{x,y} = \varnothing, \tag{6.11}$$

where $\bar{t}_i = \max T_i$ and $\underline{t}_i = \min T_i$. Suppose that $w_2 \in T_n$.
(ii)-(a) When $n = 1$.
Since T_1 is a connected component, the interval $[w_1, w_2]$ is contained in T_1. So we have $(x, w, y) \in S_{x,w,y'}$ for any $w \in [w_1, w_2]$. Let p be the integer part of $(w_2 - w_1)/M$ and put $r = w_2 - w_1 - pM$. Note that $0 \leq r < M$. For any $k = 0, 1, \ldots, p$, we have $(x, w_1 + kM, y) \in S_{x,w,y'}$. Therefore, since $r, M \in S_\eta$ and Lemma 6.13, we have

$$F(x, w_1, y) = F(x, w_1 + M, y)$$

$$\vdots$$

$$= F(x, w_1 + pM, y)$$
$$= F(x, w_1 + pM + r, y) = F(x, w_2, y).$$

(ii)-(b) When $n > 1$.
We first show that $\bar{t}_i + M \geq \underline{t}_{i+1} + m$. Put $h(x, w, y, e) = (x, w + e, y)$. From Lemma 6.7 we have $S_{x,y} = h(S_{x,w,y'} \times S_\eta)$, and since $0 \in S_\eta$,

$$S_{x,w,y'} = h(S_{x,w,y'} \times \{0\}) \subset h(S_{x,w,y'} \times S_\eta) = S_{x,y}.$$

Suppose $\bar{t}_i + M < \underline{t}_{i+1} + m$. For any $w \in (\bar{t}_i + M, \underline{t}_{i+1} + m)$, we have

$$w_1 \leq \bar{t}_i \leq \bar{t}_i + M < w < \underline{t}_{i+1} + m \leq \underline{t}_{i+1} \leq w_2.$$

Since (x, w_1, y), $(x, w_2, y) \in S_{x,w,y'} \subset S_{x,y}$, and $S_{x,y}$ is convex, we have $(x, w, y) \in S_{x,y}$. Therefore we have that there exists $(x', w', y') \in S_{x,w,y'}$ and $e \in S_\eta$ such that

$$(x, w, y) = h(x', w', y', e) = (x', w' + e, y'),$$

from Lemma 6.7. Namely, there exists $e \in S_\eta$ such that $(x, w - e, y) \in S_{x,w,y'}$. So we have $w - e \in T_{x,y} \subset (\bar{t}_i, \underline{t}_{i+1})^C$, that is, $w - e \leq \bar{t}_i$ or $w - e \geq \underline{t}_{i+1}$. If $w - e \leq \bar{t}_i$, we have

$$e \geq w - \bar{t}_i > \bar{t}_i + M - \bar{t}_i = M, \tag{6.12}$$

and if $w - e \geq \underline{t}_{i+1}$, we have

$$e \leq w - \underline{t}_{i+1} < \underline{t}_{i+1} + m - \underline{t}_{i+1} = m. \tag{6.13}$$

Both (6.12) and (6.13), however, contradict to $M = \max S_\eta$ and $m = \min S_\eta$, so we have

$$\bar{t}_i + M \geq \underline{t}_{i+1} + m. \tag{6.14}$$

We next show that $F(x, \bar{t}_i, y) = F(x, \underline{t}_{i+1}, y)$ for any i. If $\bar{t}_i \geq \underline{t}_{i+1} + m$ we have $m \leq \bar{t}_i - \underline{t}_{i+1} < 0$, that is $\bar{t}_i - \underline{t}_{i+1} \in S_\eta$. From Lemma 6.13, we have

$$F(x, \underline{t}_{i+1}, y) = F(x, \underline{t}_{i+1} + (\bar{t}_i - \underline{t}_{i+1}), y) = F(x, \bar{t}_i, y).$$

Conversely if $\bar{t}_i < \underline{t}_{i+1} + m$ we have

$$0 < \underline{t}_{i+1} + m - \bar{t}_i \leq \bar{t}_i + M - \bar{t}_i = M,$$

from (6.14). So we have $\underline{t}_{i+1} + m - \bar{t}_i \in S_\eta$. Thus from Lemma 6.13 we have

$$\begin{aligned}
F(x, \bar{t}_i, y) &= F(x, \bar{t}_i + (\underline{t}_{i+1} + m - \bar{t}_i), y) \\
&= F(x, \underline{t}_{i+1} + m, y) \\
&= F(x, \underline{t}_{i+1}, y).
\end{aligned}$$

Therefore,

$$F(x, \bar{t}_i, y) = F(x, \underline{t}_{i+1}, y). \tag{6.15}$$

We show $F(x, w_1, y) = F(x, w_2, y)$ at last. From the same argument as in (ii)-(a), we have the following equations.

$$\begin{aligned}
F(x, w_1, y) &= F(x, \bar{t}_1, y), \\
F(x, \underline{t}_n, y) &= F(x, w_2, y), \\
F(x, \underline{t}_i, y) &= F(x, \bar{t}_i, y), \qquad (i = 1, 2, \ldots, n - 1).
\end{aligned}$$

Therefore, from these equations and (6.15), we have

$$F(x, w_1, y) = F(x, \bar{t}_1, y)$$
$$= F(x, \underline{t}_2, y) = F(x, \bar{t}_2, y)$$
$$\vdots$$
$$= F(x, \underline{t}_n, y) = F(x, w_2, y).$$

LEMMA 6.15. *If (A1) to (A7) hold, we have* $F(x, y_1, y) = F(x, y_2, y)$ *for any* $(x, y_1, y), (x, y_2, y) \in S_{x,y}$.

PROOF.
(i) When S_η is not bounded.
From Lemma 6.4 and Lemma 6.6 we have $S_{x,y} = \overline{h(S_{x,w,y'} \times S_\eta)}$. There thus exist $(x_n^{(i)}, w_n^{(i)}, y_n^{(i)}) \in S_{x,w,y'}$ and $e_n^{(i)} \in S_\eta$ $(i = 1, 2)$ such that

$$(x, y_i, y) = \lim_{n \to \infty} (x_n^{(i)}, w_n^{(i)} + e_n^{(i)}, y_n^{(i)}).$$

Since F is continuous and from Lemma 6.13, we have

$$|F(x, y_1, y) - F(x, y_2, y)|$$
$$= \lim_{n \to \infty} \left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right|. \tag{6.16}$$

Since S_η is not bounded, we have $w_n^{(1)} - w_n^{(2)} \in S_\eta$ or $w_n^{(2)} - w_n^{(1)} \in S_\eta$ for each $n \in \mathbf{N}$. Therefore we have

$$\left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right|$$
$$\leq \left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right| I_{S_\eta}(w_n^{(1)} - w_n^{(2)})$$
$$+ \left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right| I_{S_\eta}(w_n^{(2)} - w_n^{(1)}), \tag{6.17}$$

where I_{S_η} is the indicator function on S_η. Let R_n denote the first term of the right-hand side of (6.17) and R_n' denote the second term, then we have

$$R_n \leq \left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(1)}, y_n^{(2)}) \right| I_{S_\eta}(w_n^{(1)} - w_n^{(2)})$$
$$+ \left| F(x_n^{(2)}, w_n^{(1)}, y_n^{(2)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right| I_{S_\eta}(w_n^{(1)} - w_n^{(2)}).$$

Since F is Lipschitz continuous, there exists a constant $C > 0$ such that

$$|F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(1)}, y_n^{(2)})|$$
$$\leq C \|(x_n^{(1)} - x_n^{(2)}, 0, y_n^{(1)} - y_n^{(2)})\|,$$

for any n. From Lemma 6.13, we have

$$\left| F(x_n^{(2)}, w_n^{(1)}, y_n^{(2)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right| I_{S_\eta}(w_n^{(1)} - w_n^{(2)}) = 0.$$

Therefore,

$$R_n \leq C\|(x_n^{(1)} - x_n^{(2)}, 0, y_n^{(1)} - y_n^{(2)})\| I_{S_\eta}(w_n^{(1)} - w_n^{(2)}),$$

and by the same argument as the above we have

$$R_n' \leq C\|(x_n^{(1)} - x_n^{(2)}, 0, y_n^{(1)} - y_n^{(2)})\| I_{S_\eta}(w_n^{(2)} - w_n^{(1)}).$$

Consequently we have

$$\left| F(x_n^{(1)}, w_n^{(1)}, y_n^{(1)}) - F(x_n^{(2)}, w_n^{(2)}, y_n^{(2)}) \right|$$
$$\leq 2C\|(x_n^{(1)} - x_n^{(2)}, 0, y_n^{(1)} - y_n^{(2)})\|.$$

So we have

$$|F(x, y_1, y) - F(x, y_2, y)| \leq 2C \lim_{n \to \infty} \|(x_n^{(1)} - x_n^{(2)}, 0, y_n^{(1)} - y_n^{(2)})\| = 0.$$

(ii) When S_η is bounded.
In this case S_η is compact, so we have $S_{x,y} = h(S_{x,w,y'} \times S_\eta)$ from Lemma 6.7. We thus have,

$$^\exists w_i, \ e_i \quad s.t. \quad (x, w_i, y) \in S_{x,w,y'}, \ e_i \in S_\eta, \ w_i + e_i = y_i \ (i = 1, 2).$$

Therefore from Lemma 6.13 and Lemma 6.14, we have

$$\left| F(x, y_1, y) - F(x, y_2, y) \right| = \left| F(x, w_1 + e_1, y) - F(x, w_2 + e_2, y) \right|$$
$$= \left| F(x, w_1, y) - F(x, w_2, y) \right| = 0.$$

LEMMA 6.16. *If* (A1) *to* (A7) *hold, there exists a constant* C *such that* $|F(x, y, y) - F(x', y', y')| \leq C\|(x, y) - (x', y')\|$ *for any* (x, y, y) *and* (x', y', y') *in* $S_{x,y}$.

PROOF. We prove this lemma for any $(x, y, y), (x', y', y') \in h(S_{x,w,y'} \times S_\eta)$ first. By the definition of $h(S_{x,w,y'} \times S_\eta)$, There exist w, w', e and e' such that

$$(x, w, y), \ (x', w', y') \in S_{x,w,y'}, \quad e, e' \in S_\eta \quad \text{and } w + e = y, w' + e' = y'.$$

From Lemma 6.13 we have

$$|F(x, y, y) - F(x', y', y')| = |F(x, w, y) - F(x', w', y')|.$$

We can assume $w' > w$ without loss of generality. Let M be a positive number in S_η. The existence of M is guaranteed by Lemma 6.11. It should be noted that any m in the interval $[0, M]$ is in S_η. Since $(x, w + m, y)$ and $(x', w' + m, y')$ are in a convex set $S_{x,y}$, we have

$$(x, w, y) + k(x' - x, w' - w, y' - y) + (0, m, 0) \in S_{x,y} \tag{6.18}$$

for $0 \leq {}^\forall k \leq 1$ and $0 \leq {}^\forall m \leq M$. Let q be the integer part of $(w' - w)/M$ and put $r = w' - w - qM$. Note that $0 \leq r < M$. Define (x_j, w_j, y_j) $(j = 0, 1, \ldots, q)$ as follows.

$$(x_j, w_j, y_j) = (x, w, y) + \frac{jM}{qM + r}(x' - x, w' - w, y' - y).$$

Note that $w_{j+1} = w_j + M$ for $j = 0, 1, \ldots, q-1$, and $w_q + r = w + qM + r = w'$. Put $(x_{q+1}, w_{q+1}, y_{q+1}) = (x', w', y')$. We have

$$
\begin{aligned}
&|F(x, w, y) - F(x', w', y')| \\
={}&|F(x_0, w_0, y_0) - F(x_0, w_1, y_0) \\
&+ F(x_0, w_1, y_0) - F(x_1, w_1, y_1) \\
&+ F(x_1, w_1, y_1) - F(x_1, w_2, y_1) \\
&+ F(x_1, w_2, y_1) - F(x_2, w_2, y_2) \\
&\qquad\qquad\vdots \\
&+ F(x_{q-1}, w_{q-1}, y_{q-1}) - F(x_{q-1}, w_q, y_{q-1}) \\
&+ F(x_{q-1}, w_q, y_{q-1}) - F(x_q, w_q, y_q) \\
&+ F(x_q, w_q, y_q) - F(x_q, w_{q+1}, y_q) \\
&+ F(x_q, w_{q+1}, y_q) - F(x_{q+1}, w_{q+1}, y_{q+1})| \\
\leq{}&\textstyle\sum_{j=0}^{q}|F(x_j, w_j, y_j) - F(x_j, w_{j+1}, y_j)| \\
&+ \textstyle\sum_{j=1}^{q+1}|F(x_{j-1}, w_j, y_{j-1}) - F(x_j, w_j, y_j)|. \qquad (6.19)
\end{aligned}
$$

From (6.18) we have

$$(x_j, w_j, y_j),\ (x_j, w_{j+1}, y_j) \in S_{x,y} \quad (j = 0, 1, \ldots, q).$$

Thus from Lemma 6.15, we have

$$F(x_j, w_j, y_j) = F(x_j, w_{j+1}, y_j) \quad (j = 0, 1, \ldots, q).$$

Therefore the first term of the right-hand side of (6.19) is equal to 0. Since F is Lipschitz continuous, we have

$$
\begin{aligned}
&|F(x_{j-1}, w_j, y_{j-1}) - F(x_j, w_j, y_j)| \leq C\|(x_{j-1} - x_j, y_{j-1} - y_j)\| \\
&= \begin{cases} \frac{CM}{qM+r}\|(x'-x, y'-y)\| & (1 \leq j \leq q), \\ \frac{Cr}{qM+r}\|(x'-x, y'-y)\| & (j = q+1), \end{cases}
\end{aligned}
$$

for a constant C. So we have

$$
\begin{aligned}
&|F(x, w, y) - F(x', w', y')| \\
\leq{}& C\textstyle\sum_{j=1}^{q+1}\|(x_{j-1} - x_j, y_{j-1} - y_j)\| \\
={}& C\left(\textstyle\sum_{j=1}^{q}\frac{M}{qM+r} + \frac{r}{qM+r}\right)\|(x'-x, y'-y)\| = C\|(x'-x, y'-y)\|,
\end{aligned}
$$

for (x, y, y) and (x', y', y') in $h(S_{x,w,y'} \times S_\eta)$. Furthermore, for any (x, y, y) and (x', y', y') in $S_{x,y}$,

$$
\begin{aligned}
&^\exists(x_n, y_n, y_n),\ (x_n', y_n', y_n') \in h(S_{x,w,y'} \times S_\eta) \quad s.t. \\
&\lim_{n\to\infty}(x_n, y_n, y_n) = (x, y, y), \quad \lim_{n\to\infty}(x_n', y_n', y_n') = (x', y', y'),
\end{aligned}
$$

since $S_{x,y} = \overline{h(S_{x,w,y'} \times S_\eta)}$. Therefore, from the former part of this proof, we have

$$
\begin{aligned}
&\left| F(x, y, y) - F(x', y', y') \right| \\
&= \left| F(\lim_{n \to \infty} (x_n, y_n, y_n)) - F(\lim_{n \to \infty} (x'_n, y'_n, y'_n)) \right| \\
&= \lim_{n \to \infty} \left| F(x_n, y_n, y_n) - F(x'_n, y'_n, y'_n) \right| \\
&\leq C \lim_{n \to \infty} \left\| (x_n - x'_n, y_n - y'_n) \right\| = C \left\| (x - x', y - y') \right\|.
\end{aligned}
$$

LEMMA 6.17. *Let $S_{x,y'}$ be a support of $(X_{t,d_o}, Y_{t-1,d_o-1})$. If (A1) to (A7) hold, there is a Lipschitz continuous function F_2 from $S_{x,y'}$ into $\mathbb{R}$ such that*

$$
F(X_{t,d_o}, Y_{t-1}, Y_{t-1,d_o-1}) = F_2(X_{t,d_o}, Y_{t-1,d_o-1}) \quad a.e.
$$

PROOF. Define a function P from $\mathbb{R}^{2d_o}$ into $\mathbb{R}^{2d_o-1}$ as $P(x, y, y) = (x, y)$. It is straightforward to show that $\overline{P(S_{x,y})}$ is convex and so we can consider the projection onto $\overline{P(S_{x,y})}$. Let $(\hat{x}, \hat{y})$ be the projection of $(x, y) \in S_{x,y'}$ onto $\overline{P(S_{x,y})}$. Note that $(\hat{x}, \hat{y})$ is uniquely defined for (x, y). Define a function F_2 from $S_{x,y'}$ into $\mathbb{R}$ as follows.

$$
F_2(x, y) = \begin{cases} \lim_{n \to \infty} F(x_n, y_n, y_n) & (\text{if } (x, y) \in \overline{P(S_{x,y})}), \\ F_2(\hat{x}, \hat{y}) & (\text{otherwise}), \end{cases}
$$

where $(x_n, y_n, y_n) \in S_{x,y}$ satisfies $\lim_{n \to \infty} P((x_n, y_n, y_n)) = (x, y)$. We first see the well-definedness of F_2, then its Lipschitz continuity, and that F_2 is the function stated in the lemma at last.

(i) The well-definedness of F_2.
It suffices to show it for $(x, y) \in \overline{P(S_{x,y})}$ since the uniqueness of a projection. Suppose that (x_n, y_n, y_n) and (x'_n, y'_n, y'_n) are in $S_{x,y}$ and satisfy

$$
\begin{aligned}
(x, y) &= \lim_{n \to \infty} P((x_n, y_n, y_n)) \\
&= \lim_{n \to \infty} P((x'_n, y'_n, y'_n)).
\end{aligned}
$$

Note that $\lim_{n \to \infty} x_n = \lim_{n \to \infty} x'_n = x$ and $\lim_{n \to \infty} y_n = \lim_{n \to \infty} y'_n = y$ by the definition of the function P. From Lemma 6.16 we have

$$
\begin{aligned}
&\left| \lim_{n \to \infty} F(x_n, y_n, y_n) - \lim_{n \to \infty} F(x'_n, y'_n, y'_n) \right| \\
&= \lim_{n \to \infty} \left| F(x_n, y_n, y_n) - F(x'_n, y'_n, y'_n) \right| \\
&\leq \lim_{n \to \infty} C \left\| (x_n - x'_n, y_n - y'_n) \right\| = 0.
\end{aligned}
$$

Therefore F_2 is well-defined.

(ii) The Lipschitz continuity of F_2.
Since $(\hat{x}, \hat{y}) = (x, y)$ if (x, y) is in $\overline{P(S_{x,y})}$, $F_2(x, y) = F_2(\hat{x}, \hat{y})$ for any (x, y) in $S_{x,y}$.

So we have, from Lemma 6.16,

$$
\begin{aligned}
\left|F_2(x,y) - F_2(x',y')\right| &= \left|F_2(\hat{x},\hat{y}) - F_2(\hat{x}',\hat{y}')\right| \\
&= \lim_{n\to\infty}\left|F(\hat{x}_n,\hat{y}_n,\hat{y}_n) - F(\hat{x}'_n,\hat{y}'_n,\hat{y}'_n)\right| \\
&\le \lim_{n\to\infty} C\left\|(\hat{x}_n - \hat{x}'_n, \hat{y}_n - \hat{y}'_n)\right\| \\
&= C\left\|(\hat{x} - \hat{x}', \hat{y} - \hat{y}')\right\| = C\left\|(\hat{x},\hat{y}) - (\hat{x}',\hat{y}')\right\|.
\end{aligned}
$$

From Lemma 6.10, $\|(\hat{x},\hat{y}) - (\hat{x}',\hat{y}')\| \le \|(x,y) - (x',y')\|$. Therefore F_2 is Lipschitz continuous.

(iii) $F(X_{t,d_o}, Y_{t-1}, Y_{t-1,d_o-1}) = F_2(X_{t,d_o}, Y_{t-1,d_o-1})$ a.e.

From Lemma 6.5, it suffices to show that $F(x,y,y) = F_2(x,y)$ for any $(x,y,y) \in S_{x,y}$. It is clear that $(x,y) \in P(S_{x,y})$ so we have

$$
|F(x,y,y) - F_2(x,y)| = |F(x,y,y) - F(x,y,y)| = 0.
$$

Therefore we have

$$
F(X_{t,d_o}, Y_{t-1}, Y_{t-1,d_o-1}) = F_2(X_{t,d_o}, Y_{t-1,d_o-1})   a.e.
$$

Proof of Theorem 3.2.
We have already shown that

$$
F(Z_{t,d_o}) = F_2(X_{t,d_o}, X_{t-1,d_o-1})   a.e.
$$

with a Lipschitz continuous function F_2 in Lemma 6.17. Furthermore F_2 and its argument $(X_{t,d_o}, Y_{t-1,d_o-1})$ satisfy all conditions needed in the proofs of Lemma 6.11 to Lemma 6.17, so we have

$$
F_2(X_{t,d_o}, X_{t-1,d_o-1}) = F_3(X_{t,d_o}, Y_{t-2,d_o-2})   a.e.
$$

with a Lipschitz continuous function F_3. By the same argument we have

$$
\begin{aligned}
F(Z_{t,d_o}) &= F_2(X_{t,d_o}, Y_{t-1,d_o-1})   a.e. \\
&= F_3(X_{t,d_o}, Y_{t-2,d_o-2})   a.e. \\
&\quad\vdots \\
&= F_{d_o}(X_{t,d_o}, Y_{t-d_o})   a.e. \\
&= F_{d_o+1}(X_{t,d_o})   a.e.
\end{aligned}
$$

with Lipschitz continuous functions $F_2, F_3, \ldots, F_{d_o+1}$. Thus put $G = F_{d_o+1}$ then we have the theorem.

6.3.   Proof of Theorem 4.1

To prove Theorem 4.1 we refer to Cheng and Tong (1993). Next two lemmas are simple extension of the theorems in Cheng and Tong (1993). We can prove them

similarly. In these lemmas, weight functions are not specified. Put $n = N - r + 1$ and

$$\varepsilon_t^{(d)} = X_t - E[X_t | Z_{t,d}], \qquad \sigma_N^2(d) = n^{-1} \sum_{t=r}^{N} (\varepsilon_t^{(d)})^2 W_{1,d}(Z_{t,d}),$$

$$\bar{\varepsilon}_t^{(d)} = X_t - E[X_t | X_{t,d}], \qquad \bar{\sigma}_N^2(d) = n^{-1} \sum_{t=r}^{N} (\bar{\varepsilon}_t^{(d)})^2 W_{0,d}(X_{t,d}).$$

Let $\sigma^2(X_t | Z_{t,d})$ and $\sigma^2(X_t | X_{t,d})$ denote the variances of $\varepsilon_t^{(d)}$ and $\bar{\varepsilon}_t^{(d)}$ respectively.

LEMMA 6.18. *Suppose* (A1), (A2), (A3) *and* (A8) *to* (A17) *hold. If*

$$(A18-1) \quad 0 \le W_{1,d}(x) \le 1 \quad \text{and} \quad \int_{\mathbb{R}^{2d}} W_{1,d}(x) dx < \infty,$$

hold for some $d \in \mathbb{N}$, *then we have*

$$CV_1(d) = \sigma_N^2(d)\{1 + \beta_1(d)\gamma_1(d)\rho_{1,d}^{2d}/N + o_p(\rho_{1,d}^{2d}/N)\}, \tag{6.20}$$

where $\rho_{1,d} = B_{2d,N}^{-1}$, $\beta_1(d) = \int \{K_{2d}(u)\}^2 du$, *and*

$$\gamma_1(d) = \int W_{1,d}(x) dx / \int W_{1,d}(x) f_d(x) dx,$$

A function f_d *is the density function of* $Z_{t,d}$ *and integrals are taken over* $\mathbb{R}^{2d}$. *If*

$$(A18-2) \quad 0 \le W_{0,d_o}(x) \le 1 \quad \text{and} \quad \int_{\mathbb{R}^{d_o}} W_{0,d_o}(x) dx < \infty$$

hold, then we have

$$CV_0(d_o) = \bar{\sigma}_N^2(d_o)\{1 + \beta_0(d_o)\gamma_0(d_o)\rho_{0,d_o}^{d_o}/N + o_p(\rho_{0,d_o}^{d_o}/N)\} \tag{6.21}$$

where $\rho_{0,d_o} = B_{d_o,N}^{-1}$, $\beta_0(d_o) = \int \{K_d(u)\}^2 du$, *and*

$$\gamma_1(d) = \int W_{1,d}(x) dx / \int W_{1,d}(x) f_d(x) dx.$$

Integrals are taken over $\mathbb{R}^{d_o}$.

LEMMA 6.19. *Let* A_n *be a compact set in* $\mathbb{R}^{2n}$ *for each* $n = 1, 2, \ldots, L$, *and put*

$$\bar{A}_n = \{(x, y) \in \mathbb{R}^{2n} \mid {}^{\exists}x, y \in \mathbb{R}; \ (x, x, y, y) \in A_{n+1}\}$$

for $n = 1, 2, \ldots, L - 1$. *Put also* $W_{1,d}(x) = I_{A_d}(x)$, *where* I_A *denotes the indicator function on* A. *If* (A1), (A2), (A3) *and* (A8) *to* (A17) *hold, and if*

$$(A19) \quad \bar{A}_n \subset A_n \quad \text{for any } n = 1, 2, \ldots, L - 1,$$

then we have the following.
(a) *For* $0 \le d < d_o$, $\lim_{N \to \infty} \Pr(\hat{d}_1 = d) = 0$.
(b) *For* $d_o < d \le L$,

$$\limsup_{N \to \infty} \Pr(\hat{d}_1 = d) \le 1 - \liminf_{N \to \infty} \Pr(\cap_{t=r}^{N}\{Z_{t,d} \in A_{2d}\}).$$

Proof of Theorem 4.1.

Since X_t and Y_t are bounded, both supports of X_t and Y_t are compact. From Lemma 6.4, we have $S_{x_d,y_d} \subset S_{x_{d-1},y_{d-1}} \times S_x \times S_y$, where $S_{x_{d-1},y_{d-1}}$, S_x, and S_y are the supports of $(X_{t,d-1}, Y_{t,d-1})$, X_t, and Y_t respectively. We thus have $S_{x_d,y_d} \subset S_x \times \cdots \times S_x \times S_y \times \cdots \times S_y$ and that S_{x_d,y_d} is compact. Therefore our weight function $W_{1,d}(x)$ satisfies (A18-1) for any $d = 1, 2, \ldots, L$. Furthermore $S_{x_{d_o}}$ is also compact, $W_{0,d}(x)$ satisfies (A18-2). We also have $\tilde{S}_{x_d,y_d} \subset S_{x_d,y_d}$ from Lemma 6.4, where $\tilde{S}_{x_d,y_d}$ is defined similarly to $\tilde{A}_n$ in Lemma 6.19. Therefore the supports of our weight functions satisfy (A19). Consequently, we can use Lemma 6.18 and Lemma 6.19 in this proof. We prove this theorem separating into 3 parts.

(i) Proof of (P1) for the NLCA method.

We first prove (P1) for the NLCA method. Put $S_{z_d} = S_{x_d,y_d}$. It is straightforward to see that

$$
\begin{aligned}
&\Pr(CV_0 - CV_1 > 0) \\
&\leq \Pr(CV_0 - CV_1 > 0, \hat{d}_1 = d_0) + \Pr(\hat{d}_1 < d_0) + \Pr(\hat{d}_1 > d_0) \\
&\leq \Pr(CV_0 - CV_1 > 0, \hat{d}_1 = d_0) \\
&\qquad + \sum_{d=0}^{d_o-1} \Pr(\hat{d}_1 = d) + \sum_{d=d_o+1}^{L} \Pr(\hat{d}_1 = d).
\end{aligned}
\tag{6.22}
$$

The second term of the right-hand side of (6.22) converges to 0 as $N \to \infty$ from the former part of Lemma 6.19. From the latter part of Lemma 6.19 and Lemma 6.2, we have

$$
\begin{aligned}
\exists \lim_{N \to \infty} \Pr(\hat{d}_1 = d) &\leq 1 - \liminf_{N \to \infty} \Pr(\cap_{t=r}^{N} Z_{t,d} \in S_{z_d}) \\
&= \liminf_{N \to \infty} \Pr(\cup_{t=r}^{N} Z_{t,d} \notin S_{z_d}) \\
&\leq \liminf_{N \to \infty} \sum_{t=r}^{N} \Pr(Z_{t,d} \notin S_{z_d}) = 0,
\end{aligned}
$$

for $d_o < d \leq L$. Therefore the third term of (6.22) converges to 0 as $N \to \infty$. To show that the first term of (6.22) converges to 0, we mimic the proof of Theorem 2 in Cheng and Tong (1992). Let B_1 be the bandwidth that gives $\text{Min}_B CV_1(d_o)$ and B_0 be one that gives $\text{Min}_B CV_0(d_o)$, then for the first term of (6.22),

$$
\Pr(CV_0 - CV_1 > 0, \hat{d}_1 = d_o) = \Pr(CV_0(d_o, B_0) - CV_1(d_o, B_1) > 0).
$$

To simplify our notation, we omit B_0 and B_1. From Lemma 6.2 we have

$$
\begin{aligned}
&\Pr(CV_0(d_o) - CV_1(d_o) > 0) \\
&= \Pr(\{CV_0(d_o) - CV_1(d_o) > 0\} \cap \cap_{t=r}^{N} \{Z_{t,d_o} \in S_{z_{d_o}}\}) \\
&\qquad + \Pr(\{CV_0(d_o) - CV_1(d_o) > 0\} \cap \cap_{t=r}^{N} \{Z_{t,d_o} \notin S_{z_{d_o}}\}) \\
&= \Pr(\{CV_0(d_o) - CV_1(d_o) > 0\} \cap \cap_{t=r}^{N} \{Z_{t,d_o} \in S_{z_{d_o}}\}).
\end{aligned}
\tag{6.23}
$$

From Lemma 6.4 we have

$$
Z_{t,d_o} \in S_{z_{d_o}} \Rightarrow X_{t,d_o} \in S_{x_{d_o}},
$$

for each t. Furthermore $\varepsilon_t^{(d_o)} = \tilde{\varepsilon}_t^{(d_o)} = \varepsilon_t$ since $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ and Theorem 3.2. From which we have that if $Z_{t,d_o} \in S_{z_{d_o}}$ for any $t = r, r+1, \ldots, N$,

$$\sigma_N^2(d_o) = \tilde{\sigma}_N^2(d_o) = n^{-1}\sum_{t=r}^{N}\varepsilon_t^2.$$

Therefore on the set $\cap_{t=r}^{N}\{Z_{t,d_o} \in S_{z_{d_o}}\}$, we have

$$\begin{aligned}
CV_0(d_o) &- CV_1(d_o)\\
=&\tilde{\sigma}_N^2(d_o)\{1 + \beta_0(d_o)\gamma_0(d_o)\rho_{0,d_o}{}^{d_o}/N + o_p(\rho_{0,d_o}{}^{d_o}/N)\}\\
&\quad - \sigma_N^2(d_o)\{1 + \beta_1(d_o)\gamma_1(d_o)\rho_{1,d_o}{}^{2d_o}/N + o_p(\rho_{1,d_o}{}^{2d_o}/N)\}\\
=&n^{-1}\sum_{t=r}^{N}\varepsilon_t^2\{\beta_0(d_o)\gamma_0(d_o)\rho_{0,d_o}{}^{d_o}/N - \beta_1(d_o)\gamma_1(d_o)\rho_{1,d_o}{}^{2d_o}/N\\
&\quad\quad\quad\quad\quad\quad + o_p(\rho_{0,d_o}{}^{d_o}/N) + o_p(\rho_{1,d_o}{}^{2d_o}/N)\},
\end{aligned}$$

from Lemma 6.18. Note that

$$\rho_{0,d_o}{}^{d_o}/N = 1/(NB_{d_o,N}d_o) = o\left(1/(NB_{2d_o,N}{}^{2d_o})\right) = o\left(\rho_{1,d_o}{}^{2d_o}/N\right),$$

and $n^{-1}\sum_{t=r}^{N}\varepsilon_t^2 = \sigma_1^2 + o_p(1)$. Therefore we have

$$NB_{2d_o,N}{}^{2d_o}\{CV_0(d_o) - CV_1(d_o)\} = -\sigma_1^2\beta_1(d_o)\gamma_1(d_o) + o_p(1)$$

on the set $\cap_{t=r}^{N}\{Z_{t,d_o} \in S_{z_{d_o}}\}$. Since $\sigma_1^2\beta_1(d_o)\gamma_1(d_o)$ is strictly positive, we have

$$\limsup_{N\to\infty} \Pr(\{CV_0(d_o) - CV_1(d_o) > 0\} \cap \cap_{t=r}^{N}\{Z_{t,d_o} \in S_{z_{d_o}}\}) = 0.$$

Thus from (6.23), if $\{Y_t\} \xrightarrow{NLGC} \{X_t\}$ we have

$$\lim_{N\to\infty} \Pr(CV_o - CV_1 > 0) = 0.$$

(ii) Proof of (P1) for the CNLCA method.
For the CNLCA method we can prove similarly. Put

$$\hat{d}_o = \arg\min_{0\le d\le L}\min_{B} CV_0(d).$$

From the former part of this proof we have

$$\begin{aligned}
\lim_{N\to\infty} &\Pr(CV_0 - CV_1 > 0)\\
\le \lim_{N\to\infty} &\Pr(CV_0 - CV_1 > 0, \hat{d}_1 = d_o)\\
\le \lim_{N\to\infty} &\{\Pr(CV_0 - CV_1 > 0, \hat{d}_o = d_o, \hat{d}_1 = d_o)\\
&\quad\quad\quad + \Pr(\hat{d}_o < d_o) + \Pr(\hat{d}_o > d_o)\}\\
= \lim_{N\to\infty} &\{\Pr(CV_0 - CV_1 > 0, \hat{d}_o = d_o, \hat{d}_1 = d_o)\\
&\quad\quad\quad + \sum_{d=0}^{d_o-1}\Pr(\hat{d}_o = d) + \sum_{d=d_o+1}^{L}\Pr(\hat{d}_o = d)\}. \quad\quad (6.24)
\end{aligned}$$

Since $\{Y_t\} \xrightarrow{\,NLGC\,} \{X_t\}$ and Theorem 3.2, conditions needed in Theorem 2 in Cheng and Tong (1992) are all satisfied, so the second term and the third term of (6.24) converge to 0. Also we have shown that the first term of (6.24) converges to 0, so we have proved (P1) for the CNLCA method.

(iii) Proof for (P2).

We prove (P2) for the NLCA method at last. Suppose $\{Y_t\} \xrightarrow{\,NLGC\,} \{X_t\}$. From the same argument as the former part of this proof,

$$\lim_{N\to\infty} \Pr(CV_0 - CV_1 > 0) \leq \lim_{N\to\infty} \Pr(CV_0 - CV_1 > 0, \hat{d}_1 = d_o)$$
$$= \lim_{N\to\infty} \{1 - \Pr(CV_0(d_o) \leq CV_1(d_o))\}. \qquad (6.25)$$

From Lemma 6.18 we have

$$CV_0(d_o) \leq CV_1(d_o)$$
$$\Longleftrightarrow \tilde{\sigma}_N^2(d_o)\{1 + \beta_0(d_o)\gamma_0(d_o)\rho_{0,d_o}{}^{d_o}/N + o_p(\rho_{0,d_o}{}^{d_o}/N)\}$$
$$\leq \sigma_N^2(d_o)\{1 + \beta_1(d_o)\gamma_1(d_o)\rho_{1,d_o}{}^{2d_o}/N + o_p(\rho_{1,d_o}{}^{2d_o}/N)\}$$
$$\Longleftrightarrow \frac{\tilde{\sigma}_N^2(d_o)}{\sigma_N^2(d_o)} \leq \frac{1 + \beta_1(d_o)\gamma_1(d_o)\rho_{1,d_o}{}^{2d_o}/N + o_p(\rho_{1,d_o}{}^{2d_o}/N)}{1 + \beta_0(d_o)\gamma_0(d_o)\rho_{0,d_o}{}^{d_o}/N + o_p(\rho_{0,d_o}{}^{d_o}/N)}. \qquad (6.26)$$

Let ξ_N denote the right hand side of (6.26). Note that

$$\xi_N \xrightarrow{P} 1. \qquad (6.27)$$

From an ergodic theorem we have

$$\sigma_N^2(d_o) \xrightarrow{N\to\infty} \sigma^2(X_t|Z_{t,d_o})\Pr(Z_{t,d_o} \in S_{z_{d_o}}) \quad a.s.$$
$$\tilde{\sigma}_N^2(d_o) \xrightarrow{N\to\infty} \sigma^2(X_t|X_{t,d_o})\Pr(X_{t,d_o} \in S_{x_{d_o}}) \quad a.s.$$

$\Pr(Z_{t,d_o} \in S_{z_{d_o}}) = \Pr(X_{t,d_o} \in S_{x_{d_o}}) = 1$ from Lemma 6.2 so that

$$\tilde{\sigma}_N^2(d_o)/\sigma_N^2(d_o) \xrightarrow{N\to\infty} \sigma^2(X_t|X_{t,d_o})/\sigma^2(X_t|Z_{t,d_o}) \quad a.s. \qquad (6.28)$$

Let s_0 be the right hand side of (6.28). Note that $\sigma^2(X_t|Z_{t,d_o}) < \sigma^2(X_t|X_{t,d_o})$ because of $\{Y_t\} \xrightarrow{\,NLGC\,} \{X_t\}$. We thus have $s_0 > 1$ and so, from (6.26), (6.27) and (6.28),

$$\Pr(CV_0(d_o) \leq CV_1(d_o))$$
$$= \Pr\left(\tilde{\sigma}_N^2(d_o)/\sigma_N^2(d_o) \leq \xi_N\right)$$
$$= \Pr\left(\xi_N - 1 + s_0 - \tilde{\sigma}_N^2(d_o)/\sigma_N^2(d_o) \geq s_0 - 1\right)$$
$$\leq \Pr\left(|\xi_N - 1| + |\tilde{\sigma}_N^2(d_o)/\sigma_N^2(d_o) - s_0| \geq s_0 - 1\right)$$
$$\leq \Pr\left(|\xi_N - 1| > (s_0 - 1)/3\right)$$
$$\qquad\qquad + \Pr\left(|\tilde{\sigma}_N^2(d_o)/\sigma_N^2(d_o) - s_0| > (s_0 - 1)/3\right)$$
$$\to 0 \quad \text{as} \quad N \to \infty.$$

From (6.25), we thus have $\lim_{N\to\infty} \Pr(CV_0 - CV_1 > 0) = 1$.

7. Discussion

We defined Granger type causality for nonlinear data, and proposed two methods for the analysis of nonlinearity. Simulation results indicated that these two methods are effective for analyzing nonlinear causality, and that the Granger-Sargent test is not appropriate for nonlinear causality. If we are interested in only linear causality, the Granger-Sargent test is sufficient, but there would be many models in economic science which show nonlinear causality that our methods could be applied.

It should be mentioned that X_t and Y_t are not necessarily bounded in our simulation models, though they are required to be in Theorem 4.1, our results of the simulation were fairly good. This fact suggests that the bounded condition for X_t and Y_t might not be crucial for the result.

In this paper we assumed the strict stationarity. It would be difficult to testing the strict stationarity. As Nakano (1995) pointed out, we should test weak stationarity at least. We omit this test to simplify the argument. The methods for testing weak stationarity has been proposed by Okabe and Nakano (1991) and Ohama and Yanagawa (1997).

Acknowledgement

References

Baek, E. and Brock, W. (1992). A general test for nonlinear Granger causality: Bivariate model, *Working Paper, Iowa State University and University of Wisconsin, Madison*.

Billingsley, P. (1995). *Probability and Measure*, 3rd ed., John Wiley & Sons, Inc.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series :Theory and Methods*, 2nd ed., Springer-Verlag.

Cheng, B. and Tong, H. (1992). On consistent nonparametric order determination and chaos, *Journal of Royal Statistical Society* B, **54**, No. 2, 427-449.

Cheng, B. and Tong, H. (1993). On residual sums of squares in non-parametric autoregression, *Stochastic Processes and their Applications*, **48**, 157-174.

Cheng, B. and Tong, H. (1995). Orthogonal projection, embedding dimension and sample size in chaotic time series from a statistical perspective, in *Chaos and Forecasting: Proceedings of the Royal Society Discussion Meeting, London, 2-3 March 1994* (H. Tong ed.), 1-29.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and Ccoss-spectral methods, *Econometrica,* **37**, No. 3, 424-438.

Nakano, Y. (1995). On a causal analysis of economic time series, *Hokkaido Mathematical Journal,* **24**, 179-213.

Ohama, G. and T. Yanagawa (1997). Testing stationarity using residual, *Bulletin of Informatics and Cybernetics*, **29**, No. 1, 15-39.

Okabe, Y. and Inoue, A. (1994). The theory of KM_2O-Langevin equations and its applications to data analysis (II): Causal analysis (I), *Nagoya Mathematical Journal*, **134**, 1-28.

Okabe, Y. and Nakano, Y. (1991). The theory of KM_2O-Langevin equations and its applications to data analysis (I): Stationary analysis, *Hokkaido Mathematical Journal*, **20**, 45-90.

Sargent, T. J. (1976). A classical macroeconometric model for the United States, *Journal of Political Economy*, **84**, No. 2, 207-237.

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

Tong, H. (1990). *Non-linear Time Series*, Oxford University Press.