Consistent Nonparametric Tests for Granger Causality [1]

Yoshihiko Nishiyama*
(*Institute of Economic Research, Kyoto University*),
Kohtaro Hitomi (*Kyoto Institute of Technology*),
Yoshinori Kawasaki (*The Institute of Statistical Mathematics*),
Kiho Jeong (*Kyungpook National University*)

Abstract

Since the pioneering work by Granger (1969), many authors have proposed tests of causality between economic time series. Most of them are concerned only with "linear causality in mean", or if a series linearly affects the (conditional) mean of the other series. It is no doubt of a primary interest, but dependence between series may be nonlinear, and/or not only through the conditional mean. Indeed conditional heteroskedastic models are widely studied recently. The purpose of this paper is to propose nonparametric tests for possibly nonlinear causality up to K-th conditional moment. A desirable property of the tests is that they have nontrivial power against $T^{1/2}$-local alternatives, where T is sample size. Their null asymptotic distributions are not normal, but we can easily calculate the critical regions by simulation. Monte Carlo experiments showed that the proposed tests have good power properties, and much better than the competitors.

Keywords: Causality up to K-th moment; nonparametric test; local alternatives

JEL classifications: C12, C14, C32

[1] This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for 21st Century COE Program, and Japan Society for the Promotion of Science Grant 15330040.

* Corresponding author. E-mail address: nishiyama@kier.kyoto-u.ac.jp

1 Introduction

Causality between variables has been one of the main interest in time series econometrics since the pioneering work by Granger (1969). We propose a nonparametric test for Granger-type causality. A conceptually similar work will be Robinson (1989) and Hidalgo (2000). The former proposed a nonparametric test on certain density weighted conditional moment restrictions, while the latter introduced a nonparametric Granger causality test in the frequency domain for weakly stationary linear processes. He is mainly concerned with the test under long range dependent observations, and it does not have a power against some alternatives of series with nonlinear dynamics. We construct a test statistic based on moment conditions allowing for nonlinear dependence. It has a nontrivial power against $T^{1/2}$-local alternatives where T is sample size. Its null asymptotic distribution is non-Gaussian, but we can easily calculate the critical region by simulation. When applied to regression analysis for cross section data, this test reduces to a nonparametric omitted variable test, or significance test of regressors, which was considered in Okui and Hitomi (2002).

Causality is, philosophically, not an easy concept to capture, but Granger (1969) gave a practical definition to deal with it in the context of time series analysis. Suppose we have a two dimensional time series $(x_t, y_t), t = 1, \cdots, T$. We are concerned if there exists any causality between x and y. Granger defined that y_t is said to cause x_t in mean if

$$\mathrm{E}[x_t - P(x_t|x_{t-1}, \ldots, x_1)]^2 > \mathrm{E}[x_t - P(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1)]^2, \tag{1.1}$$

where $P(A_t|B_t)$ is the optimum linear (or least squares) predictor of A_t given B_t (see Granger (1969, p.429)), and denoted it as $y_t \to x_t$. Otherwise $y_t \nrightarrow x_t$. An interpretation of this definition is that we say y_t causes x_t when we can improve the linear prediction of x_t using the information carried by $y_{t-1}, \ldots, y_1$. Granger remarks this definition of causality means "linear causality in mean". Under the linearity assumption that the process has a representation $y_t = \sum_{j=-\infty}^{\infty} \alpha_j x_{t-j} + u_t$, we can test the null hypothesis $H_0 : y_t \nrightarrow x_t$ against $H_1 : y_t \to x_t$ as in Sims (1972) or Hosoya (1977) using the property that H_0 is equivalent to $\alpha_j = 0$ for all $j < 0$. This approach has been most commonly used since Sims (1972). See e.g. Geweke (1982), Sims, Stock and Watson (1990), Toda and Phillips (1993), Hosoya (1991) and Lutkepohl and Poskitt (1996). To the best of our knowledge, Hidalgo (2000) is the newest result following this line allowing for long range dependence without a specification on the distribution. However, this approach may fail to detect some nonlinear causal relationships. The reason is that they construct test statistics based on linear projections of the series of interest, but we can only say that the error terms are uncorrelated with the series of interest, not independent. In many of the aforementioned research, the author(s) apply frequency domain analysis where causality is captured through cross-spectrums, or covariances of y and x. But covariances can easily be zero under nonlinear relationships even if the two variables are dependent. Then, it is unlikely that tests based on the covariances possess a good power property against certain alternatives.

We propose a nonparametric test which has power even when the observations are nonlinearly dependent. In order for this purpose, we replace the linear projections by the optimum predictor, or conditional expectations, namely we rewrite (1.1) to define the possibly nonlinear causality as

$$\mathrm{E}[x_t - \mathrm{E}(x_t|x_{t-1}, \ldots, x_1)]^2 > \mathrm{E}[x_t - \mathrm{E}(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1)]^2.$$

Straightforward calculation gives

$$\mathrm{E}[x_t - P(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1)]^2 = \mathrm{E}[x_t - P(x_t|x_{t-1}, \ldots, x_1)]^2$$
$$-\mathrm{E}[\mathrm{E}(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) - \mathrm{E}(x_t|x_{t-1}, \ldots, x_1)]^2.$$

Thus we define "y_t (possibly nonlinearly) causes x_t in mean" if

$$\mathrm{E}[\mathrm{E}(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) - \mathrm{E}(x_t|x_{t-1}, \ldots, x_1)]^2 > 0, \tag{1.2}$$

and we call this simply the causality in mean throughout this paper. We test the null hypothesis,

$$H_0 : \mathrm{E}[\mathrm{E}(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) - \mathrm{E}(x_t|x_{t-1}, \ldots, x_1)]^2 = 0 \tag{1.3}$$

or

$$H_0 : \mathrm{E}(x_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) = \mathrm{E}(x_t|x_{t-1}, \ldots, x_1) \text{ w.p. } 1 \tag{1.4}$$

against the alternative hypothesis (1.2).

We also propose a test which can detect causality in higher order moments. A motivation for this consideration is that we may also would like to detect, for example, the following nonlinear dependence between series.

$$x_t = g(x_{t-1}, \ldots, x_{t-p}) + \sigma(y_{t-1})\epsilon_t,$$

when y_{t-1} has information in predicting x_t^2. This type of modeling is getting more and more popular recently in analyzing, for instance, financial data. In general, we would like to know if y_{t-1} is useful in predicting x_t^K for a given positive integer K. It straightforwardly gives a definition of "causality in K-th moment" similarly to the above, or we say "y_t causes x_t in K-th moment" if

$$\mathrm{E}[x_t^K - \mathrm{E}(x_t^K|x_{t-1}, \ldots, x_1)]^2 > \mathrm{E}[x_t^K - \mathrm{E}(x_t^K|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1)]^2. \tag{1.5}$$

This definition also employs the MSE criterion for the prediction error. By a similar manipulation to the derivation of (1.4), the null hypothesis corresponding to the alternative of (1.5) can be written as

$$H_0 : \mathrm{E}(x_t^K | x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) = \mathrm{E}(x_t^K | x_{t-1}, \ldots, x_1)] \text{ w.p. } 1 \tag{1.6}$$

We construct test statistics based on the moment conditions (1.4) and (1.6). The next section provides the test statistic for causality in mean and its null distribution as well as the regularity conditions. Section 3 gives results on the properties of local alternatives. Section 4 presents testing procedure for causality up to K-th moment. We report Monte Carlo results in Section 5. Section 6 concludes this paper. The proofs are in the Appendix.

2 Test Statistic for Causality in Mean

2.1 Hypotheses and the corresponding moment conditions

This section provides heuristic arguments of how to test (1.4) and then provides the test statistics. We restrict ourselves to the case when x_t follows a nonlinear AR model of the form

$$\mathrm{E}[x_t | x_{t-1}, \ldots, x_{t-p}, y_{t-1}, \ldots, y_{t-q}] = m(x_{t-1}, \ldots, x_{t-p}, y_{t-1}, \ldots, y_{t-q}),$$

and (x_t, y_t) is a strictly stationary process. We assume p and q are fixed and known integers for the moment, and $m(\cdot)$ is an unknown function satisfying certain smoothness conditions. Denote

$$X_{t-1} = (x_{t-1}, \ldots, x_{t-p}), Y_{t-1} = (y_{t-1}, \ldots, y_{t-q}), Z_{t-1} = (X_{t-1}, Y_{t-1}),$$

and put

$$g(X_{t-1}) = \mathrm{E}[x_t | X_{t-1}]$$

then the event

$$m(Z_{t-1}) = g(X_{t-1})$$

is equivalent to the event

$$\mathrm{E}(u_t | Z_{t-1}) = 0$$

where

$$u_t = x_t - g(X_{t-1})$$

Therefore, we can represent the null and alternative hypotheses, respectively, as

$$H_0 \;: \mathrm{P}[\mathrm{E}(u_t | Z_{t-1}) = 0] = 1 \tag{2.1}$$

and

$$H_1: \ \mathrm{P}[\mathrm{E}(u_t|Z_{t-1}) = 0] < 1. \tag{2.2}$$

Let $s_X = \{s(\cdot)|E[s(X_{t-1})^2] < \infty\}$ and $s_Z = \{s(\cdot)|E[s(Z_{t-1})^2] < \infty\}$ be the Hilbert L_2 spaces. We can decompose s_Z into s_X and $s_X^{\perp}$, where $s_X^{\perp}$ is a Hilbert space orthogonal to s_X. That is, for any function $p(z) \in s_Z$, we can represent $p(z) = p_X(x) + p_{X^{\perp}}(z)$ such that $p_X(x) \in s_X$ and $p_{X^{\perp}}(z) \in s_X^{\perp}$. Noting u_t is orthogonal to s_X by construction, we have

$$\mathrm{E}(u_t|Z_{t-1}) = 0 \Longleftrightarrow \mathrm{E}(u_t p(Z_{t-1})) = 0, \quad \text{for } \forall p(z) \in s_X^{\perp} \tag{2.3}$$

Thus we can rewrite the null and alternative hyptohesis to

$$H_0: \ \mathrm{E}[u_t p(Z_{t-1})] = 0, \quad \text{for } \forall p(z) \in s_X^{\perp} \tag{2.4}$$

and

$$H_1: \ \mathrm{E}[u_t p(Z_{t-1})] \neq 0, \quad \text{for some } \ p(z) \in s_X^{\perp}. \tag{2.5}$$

Let $H = \{h_i(z)\}_{i=1}^{\infty}$ be a complete orthonormal basis over $s_X^{\perp}$ and assume the innovation u_t is an i.i.d. sequence with mean zero and $\mathrm{E}(u_t^2|Z_{t-1}) = \sigma^2$ under the null hypothesis. We consider tests without the restriction of $\mathrm{E}(u_t^2|Z_{t-1}) = \sigma^2$ in Section 4. Define, for $r = \max(p, q)$,

$$a_i = \frac{1}{\sqrt{T}\sigma} \sum_{t=r+1}^{T} u_t h_i(Z_{t-1}), \tag{2.6}$$

then due to a central limit theorem for martingale difference sequences, we have $a_i \xrightarrow{d} N(0,1)$ under the null. We construct a test statistic combining these moment conditions as follows. Let $\{w_i\}_{i=1}^{\infty}$ be a user determined summable positive sequence, such as $w_i = 0.9^i$, then

$$S_T = \sum_{i=1}^{T} w_i a_i^2 \xrightarrow{d} \sum_{i=1}^{\infty} w_i \epsilon_i^2 \tag{2.7}$$

where ϵ_i are i.i.d. $N(0,1)$ random variables. It is obvious $S_T \xrightarrow{d} \infty$ under the alternative because the expectation of a_i is $O(\sqrt{T})$. We will show later that a feasible version of S_T has a nontrivial power under $\sqrt{T}$-local alternatives.

2.2 Test statistics

The previous subsection has given the ideas on how to test the null hypothesis (1.4), however (2.7) is infeasible. We propose a feasible test statistic in this section. The unknown components in (2.6) are u_t, σ^2 and $h_i(z)$ which are replaced by their "estimates". We obtain the "estimates" of u_t straightforwardly by

$$\hat{u}_t = x_t - \hat{g}(X_{t-1}) \tag{2.8}$$

where $\hat{g}(X_{t-1})$ is a nonparametric estimate of $g(X_{t-1})$. A consistent estimate for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t^2. \tag{2.9}$$

We construct estimates of $h_i(z)$ as follows. Given $\{q_i(z)\}_{i=1}^{\infty}$, a basis of s_Z, $\{p_i(z) = q_i(z) - E[q_i(Z_{t-1})|X_{t-1} = x]\}_{i=1}^{\infty}$ forms a basis of $s_X^{\perp}$. We obtain an orthonormal basis on $s_X^{\perp}$ by making $\{h_i(z)\}_{i=1}^{\infty}$ orthonormal, applying, for example, the Gram-Schmitt procedure. Since $g_{q_i}(x) = \mathrm{E}[q_i(Z_{t-1})|X_{t-1} = x]$ is unknown, we estimate it nonparametrically. Denote it $\hat{g}_{q_i}(x)$. Plugging these estimates, we produce $\{\hat{h}_i(z)\}_{i=1}^{\infty}$. Formally, we can construct an orthonormal basis recursively from

$$\hat{h}_1(z) = \left\{\frac{1}{T-p}\sum_{t=r+1}^{T} q_1(Z_{t-1})^2\right\}^{-1/2} q_1(z),$$

$$\hat{h}_i(z) = \left\{\frac{1}{T-p}\sum_{t=r+1}^{T} q_i(Z_{t-1})^2 - \sum_{j=1}^{i-1}\varsigma_{ji}\right\}^{-1/2} \left\{q_i(z) - \sum_{j=1}^{i-1}\varsigma_{ji}h_j(z)\right\}, \tag{2.10}$$

where

$$\varsigma_{ji} = \frac{1}{T-p}\sum_{t=r+1}^{T}\hat{h}_j(Z_{t-1})q_i(Z_{t-1}). \tag{2.11}$$

Plugging (2.9)-(2.10) into (2.6), we make a sample analogue of (2.7),

$$\hat{S}_T = \sum_{i=1}^{T} w_i\hat{a}_i^2, \tag{2.12}$$

where

$$\hat{a}_i = \frac{1}{\sqrt{T}\hat{\sigma}}\sum_{t=p+1}^{T}\hat{u}_t\hat{h}_i(Z_{t-1}). \tag{2.13}$$

2.3 The null distribution

We give a set of regularity conditions and the null distribution of the test statistic proposed in the previous section.

Definition 1 Let $\{z_t\}, t = 1, \dots, T$ be a strictly stationary time series defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_s^t$ be the σ-algebra generated by $\{z_s, \dots, z_t\}$. Then we say $\{z_t\}, t = 1, \dots, T$ is absolutely regular when

$$\beta(k) = \mathrm{E}\left\{\sup_{A\in\mathcal{F}_k^{\infty}}|P(A|\mathcal{F}_{-\infty}^0) - P(A)|\right\} \to 0 \quad \text{as} \quad k\to\infty. \tag{2.14}$$

We call the coefficient $\beta(k)$ the coefficient of absolute regularity.

We note that the ϕ-mixing condition implies absolute regularity and absolute regularity implies the strong-mixing. The following lemma by Yoshihara (1976) is essentially useful to prove Theorem 1.

Lemma 1 *Let $\{z_t\}, t = 1, \ldots, T$ be an absolutely regular sequence of random variables with coefficient $\beta(k)$. Let $t_1 < \ldots < t_n$ be integers. Let $F(a, b)$ be the distribution function of $z_{t_a}, \ldots, z_{t_b} (a \le b)$. Let $s(\xi) = s(\xi_1, \ldots, \xi_n)$ be a Borel-measurable function. Then for $\delta > 0$,*

$$\left| \int s(\xi) dF(1, n) - \int s(\xi) dF(1, j) dF(j + 1, n) \right| \le 3M^{1/(1+\delta)} \{\beta(t_{j+1} - t_j)\}^{\delta/(1+\delta)} \tag{2.15}$$

providing $M \equiv \max\left(\int |s(\xi)|^{1+\delta} dF(1, n), \int |s(\xi)|^{1+\delta} dF(1, j) dF(j + 1, n)\right)$ exists.

We use a kernel method to estimate the above nonparametric functions, namely,

$$\hat{g}(x) = \frac{\hat{r}(x)}{\hat{f}(x)}, \hat{r}(x) = \frac{1}{Th^p} \sum_{s=p+1}^{T} K\left(\frac{X_{s-1} - x}{h}\right) x_s \tag{2.16}$$

$$\hat{f}(x) = \frac{1}{Th^p} \sum_{s=p+1}^{T} K\left(\frac{X_{s-1} - x}{h}\right) \tag{2.17}$$

and

$$\hat{g}_{q_i}(x) = \frac{\hat{r}_{q_i}(x)}{\hat{f}(x)}, \quad \hat{r}_{q_i}(x) = \frac{1}{Th^p} \sum_{s=p+1}^{T} K\left(\frac{X_{s-1} - x}{h}\right) q_i(x_s), \tag{2.18}$$

where $K(\cdot)$ is a symmetric kernel function, h is the bandwidth satisfying some conditions described below.

Assumption A1. $z_t = (x_t, y_t), t = 1, \ldots, T$ is a strict stationary and absolutely regular sequence of stochastic vector with absolue regularity coefficient $\beta(k) = O(k^{-(2+\eta)/\eta})$ for some $\eta > 0$.

Assumption A2. $V(u_t|Z_{t-1}) = \sigma^2 < \infty$ under the null hypothesis.

Assumption A3. Let $K : R^p \to R$ be a kernel function satisfying $K(-x) = K(x)$,

$$\int x_1^{l_1} \cdots x_p^{l_p} K(x) dx \begin{cases} = 1 & \text{if } \ l_1 + \cdots + l_p = 0 \\ = 0 & \text{if } \ l_1 + \cdots + l_p < L \\ \neq 0 & \text{for some } l_1 + \cdots + l_p = L, \end{cases} \tag{2.19}$$

$\int K(x)^2 dx + \int \|x\|^L |K(x)| dx < \infty$, and $\|x\|^L |K(x)| \to 0$ as $\|x\| \to \infty$. h is a positive constant decaying to zero satisfying $T^{-1}h^{-p} + \sqrt{T} h^L = o(1)$ as $T \to \infty$.

Assumption A4. Let $f_{t_1,t_2,\ldots,t_j}(\xi_1, \xi_2, \ldots, \xi_j)$ be joint density function of $x_{t_1}, x_{t_2}, \ldots, x_{t_j}$. $f_{t,t-1,\ldots,t-p}(\xi_1, \xi_2, \ldots, \xi_p)$ is L times differentiable, and the first L derivatives are uniformly continuous and bounded for $L > p/2$, then

$$\sup_{\xi_1,\xi_2} |f_{t,s}(\xi_1, \xi_2)| + \sup_{\xi_1} |f_t(\xi_1)| < \infty.$$

Let $f^{(l_1,\ldots,l_p)}(x) = \partial^{l_1+\ldots+l_p} f(x)/\partial x_1^{l_1} \cdots \partial x_p^{l_p}$ and $w_i(Z_t) = f(X_t)^{-1} g(X_t) h_i(Z_t)$, then $\mathrm{E}|w_i(Z_t) f^{(l_1,\ldots,l_p)}(X_t)|^{2+\epsilon} < \infty$ for all $l_1, \ldots, l_p$ satisfying $0 \le l_1, \ldots, l_p \le L$, $l_1 + \ldots + l_p = L$ and some $\epsilon > 0$.

Theorem 2 *Under assumptions A1-A4,*

$$\hat{a}_i = a_i + o_p(1). \tag{2.20}$$

Theorem 3 *Under assumptions A1-A4,*

$$\hat{S}_T = \sum_{i=1}^{T} w_i \hat{a}_i^2 \xrightarrow{d} \sum_{i=1}^{\infty} w_i \epsilon_i^2. \tag{2.21}$$

3 Statistical properties of the test statistic under the $\sqrt{T}$-local alternatives

The above section proposes a test statistic for Granger-type causality and its null distribution. We shall show that $\hat{S}_T$ has a nontrivial power against $\sqrt{T}$-local alternatives. When we have a prior knowledge on the direction of alternatives, (2.44) has a great advantage in the sense that we can include this information to construct $\hat{a}_i$ to induce a good power property. The reason is that $\hat{S}_T$ possesses more power for the discrepancy from the null toward $h_i(z)$ than $h_j(z)$ for $w_i > w_j$ because it gives more weight on the former than the latter. So, suppose, for instance, we know that the alternative is likely to be the null function plus $\sin(Z_{t-1})$, then, by choosing $h_1(Z_{t-1}) = \sin(Z_{t-1}) - \mathrm{E}[\sin(Z_{t-1})|X_{t-1}]$, we can expect a great power in this testing.

Theorem 4 *Let $k : R^{p+q} \to R$ be a measurable function in the space $s_X^{\perp}$. We consider the following local alternatives.*

$$H_{la}: \ E(u_t|Z_{t-1}) = \frac{1}{\sqrt{T}} k(Z_{t-1})$$

Putting $\kappa_i = E[k(Z_t)h_i(Z_t)]$, when assumptions A1-A4 hold,

$$\hat{S}_T \xrightarrow{d} \sum_{i=1}^{\infty} w_i(\epsilon_i + \kappa_i)^2 \tag{3.1}$$

under H_{la}.

4 Test Statistics for Causality up to K-th Moment

This section provides the test statistic for causality up to K-th moment. The main idea is the same as in the previous section. Here also we restrict ourselves to the case when the series of interest follows a nonlinear AR process under the null. We first give a heuristic explanation on how we test the null of (1.6). Similarly to the case of causality in mean, we rewrite (1.6) in terms of the regression residual. It is easy to show that

$$\mathrm{E}(x_t^k|x_{t-1},\dots,x_{t-p},y_{t-1},\dots,y_{t-q}) = \mathrm{E}(x_t^k|x_{t-1},\dots,x_{t-p}) \text{ for } k = 1,2,\cdots,K \tag{4.1}$$

is equivalent to

$$\mathrm{E}(u_t^k|x_{t-1},\dots,x_{t-p},y_{t-1},\dots,y_{t-q}) = \mathrm{E}(u_t^k|x_{t-1},\dots,x_{t-p}) \text{ for } k = 1,2,\cdots,K. \tag{4.2}$$

Thus the null hypothesis reduces to

$$H_0 : \mathrm{E}[U_t - \mathrm{E}(U_t|X_{t-1})|Z_{t-1}] = 0 \tag{4.3}$$

where

$$U_t = [u_t, u_t^2, \cdots, u_t^K]'. \tag{4.4}$$

This is the moment condition we would like to test now. Let

$$A_i = \Sigma_i^{-1/2} B_i, \tag{4.5}$$

where

$$B_i = \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \{U_t - E(U_t|X_{t-1})\} h_i(Z_{t-1}), \tag{4.6}$$

and

$$\Sigma_i = \mathrm{Var}\left[\{U_t - \mathrm{E}(U_t|Z_{t-1})\}\, h_i(Z_{t-1})\right]. \tag{4.7}$$

We introduce additional assumptions.

Assumption A5. $E|u_t|^K < \infty$.

Assumption A6. $E|u_t|^{2K} < \infty$.

Then the following result holds.

Theorem 5 *Under Assumptions A1-A5,*

$$A_i \xrightarrow{d} N(0, I_K). \tag{4.8}$$

under H_0, where I_K denotes the $K \times K$ identity matrix, and parallel to (2.7), we have

$$S_{KT} = \sum_{i=1}^{T} A_i' A_i w_i \xrightarrow{d} \sum_{i=1}^{\infty} w_i \sum_{k=1}^{K} \epsilon_{ik}^2. \tag{4.9}$$

where $\{\epsilon_{i1,\cdots,}\epsilon_{iK}\}_{i=1,2,\cdots}$ is a random sample from a K dimensional standard normal distribution.

We construct a feasible version of A_i and thus S_{KT} to implement the test as follows. Denote

$$\hat{U}_t = [\hat{u}_t, \hat{u}_t^2, \cdots, \hat{u}_t^K]',$$

$$\hat{E}(U_t|X_{t-1}) = \frac{1}{Th^p \hat{f}(X_{t-1})} \sum_{s=p+1}^{T} K\left(\frac{X_{s-1} - X_{t-1}}{h}\right) \hat{U}_s, \tag{4.10}$$

$$\hat{E}(U_t U_t'|Z_{t-1}) = \frac{1}{Th^{p+q} \hat{f}(Z_{t-1})} \sum_{s=r+1}^{T} K\left(\frac{Z_{s-1} - Z_{t-1}}{h}\right) \hat{U}_s \hat{U}_s', \tag{4.11}$$

$$\hat{E}(U_t|Z_{t-1}) = \frac{1}{Th^{p+q} \hat{f}(Z_{t-1})} \sum_{s=r+1}^{T} K\left(\frac{Z_{s-1} - Z_{t-1}}{h}\right) \hat{U}_s, \tag{4.12}$$

and construct

$$\hat{S}_{KT} = \sum_{i=1}^{T} \hat{A}_i' \hat{A}_i w_i \tag{4.13}$$

where for $i = 1, 2, 3, \cdots, T$,

$$\hat{A}_i = \hat{\Sigma}_i^{-\frac{1}{2}} \hat{B}_i,$$

$$\hat{B}_i = \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \left\{\hat{U}_t - \hat{E}(U_t|X_{t-1})\right\} \hat{h}_i(Z_{t-1}),$$

$$\hat{\Sigma}_i = \frac{1}{T} \sum_{t=r+1}^{T} \left\{\hat{E}(U_t U_t'|Z_{t-1}) - \hat{E}(U_t|Z_{t-1}) \hat{E}(U_t'|Z_{t-1})\right\} \hat{h}_i(Z_{t-1})^2.$$

Then we have

Theorem 6 *If Assumptions A1-A4, and A6 hold,*

$$\hat{S}_{KT} = \sum_{i=1}^{T} \hat{A}_i' \hat{A}_i w_i \xrightarrow{d} \sum_{i=1}^{\infty} w_i \sum_{k=1}^{K} \epsilon_{ik}^2. \tag{4.14}$$

under H_0.

5 A Monte Carlo Study

In this section we report the results of a Monte Carlo study to investigate the performance of our proposed test statistic. Throughout this section, x_t is the time series of our primary interest, and suppose that we want to know whether or not another time series y_t (in lags) accounts for the variation of x_t. Let $\{\eta_t\}$ and $\{\epsilon_t\}$ be the innovation processes of x_t and y_t respectively, and we assume both follow standard Normal distribution identically and independently with innovation variance $\sigma_\eta^2 = \sigma_\epsilon^2 = 1$.

5.1 Testing Causality in Mean

We consider four simulation settings as below. These can be considered as time series analogue of the experiments carried out in Okui and Hitomi (2002).

DGP 0: $x_t = 0.65x_{t-1} + \eta_t, y_t = -0.3y_{t-1} + \epsilon_t$
DGP A1: $x_t = 0.65x_{t-1} + 0.2y_{t-1} + \eta_t, y_t = -0.3y_{t-1} + \epsilon_t$
DGP A2: $x_t = 0.65x_{t-1} + 0.2y_{t-1} + 0.4\sin(-2y_{t-1}) + \eta_t, y_t = -0.3y_{t-1} + \epsilon_t$
DGP A3: $x_t = 0.65x_{t-1} + 0.2y_{t-1}^2 + \eta_t, y_t = -0.3y_{t-1} + \epsilon_t$

Because there is no causal relationship between x_t and y_t in DGP 0, the null hypothesis of non-Granger-causality should be maintained. This experiment is carried out to illuminate the size property of our test. DGP 1 covers the case of linear vector autoregressive models. In DGP 2 and 3, the target time series x_t depends on the lagged covariant time series y_{t-1} in some nonlinear fashions. In other words, we observe x_t and y_t and want to know if y_t causes x_t in Granger's sense, but we do not know in what functional form x_t depends on y_t.

As a competitor to our test, we employ a nonparametric Granger causality test from Hidalgo (2000). The basic idea of Hidalgo's test is to use a nonparametric estimate of cross-spectrum by which the causality from one time series to another is determined. It should be noted that the main concern of Hidalgo (2000) is to allow long memory time series in causality testing, and he also showed that his test has power against $\sqrt{T}$-local alternatives.

The results are summarized in Table 1. T stands for the sample size of time series, and the number of iteration is fixed to 1000. Each entry shows empirical rejection rate of the null hypothesis of non-Granger causality, hence the two columns under DGP 0 are the empirical size, and others show empirical power. Our test is abbreviated as WSC after weighted squared coefficient test named in Okui and Hitomi (2000) while Hidalgo's nonparametric Granger causality test is shortened as HNC.

As is seen clearly, the size of WSC test is quite decent while HNC test looks sloppy. Hidalgo's test depends on how many lags in cross-covariance are used to construct the test statistic (the choice of M in his notation). As he points out, it is possible to make use of some model selection procedures to choose a plausible M, but our choice here is $M = T^{1/4}$, simply following the usual assumption that facilitates the asymptotic theory. What is remarkable is the power of WSC test in DGP A2 and A3 cases, especially when the sample sizes are moderate ($T = 200$) and huge ($T = 500$). On the other hand, Hidalgo's test generally fails to detect the causality arising from nonlinear relationships. However, it is worth mentioning that Hidalgo's test shows high power for DGP A1 experiment even in small sample case ($T = 100$). This is because DGP A1 is exactly the situation that Hidalgo's test expects, and we will visit this issue again in the experiments of $\sqrt{T}$-local alternatives.

Table 1: Empirical size and power of the proposed test (WSC) and Hidalgo's non-parametric causality test (HNC).

	DGP 0		DGP A1		DGP A2		DGP A3	
T	WSC	HNC	WSC	HNC	WSC	HNC	WSC	HNC
100	0.047	0.101	0.309	0.532	0.395	0.225	0.523	0.123
200	0.041	0.112	0.616	0.775	0.738	0.358	0.868	0.124
500	0.057	0.138	0.980	0.992	0.999	0.680	1.000	0.174

Generally speaking, the magnitude of innovation variance of covariant time series is one of the factors that determine the difficulty of causality test. In the previous experiment (Table 1), it is assumed that $\sigma_\eta^2 = \sigma_\epsilon^2 = 1$. Table 2 shows the results of the second run of simulation experiment under the choice of $\sigma_\epsilon^2 = 0.8$ while σ_η^2 is kept to be unity. From Table 2, it is observed that all the entries except for empirical sizes are less than the corresponding figures in Table 1. This gives a reasonable interpretation that the testing problem has become more difficult in this second experiment.

Table 2: Empirical size and power of the proposed testv(WSC) and Hidalgo's non-parametric causality test (HNC): exogenous variable has a smaller innovation variance.

	DGP 0		DGP A1		DGP A2		DGP A3	
T	WSC	HNC	WSC	HNC	WSC	HNC	WSC	HNC
100	0.047	0.101	0.253	0.471	0.288	0.144	0.332	0.117
200	0.041	0.112	0.532	0.714	0.622	0.192	0.667	0.120
500	0.057	0.138	0.941	0.972	0.982	0.340	0.996	0.166

5.2 Power under $\sqrt{T}$-Local Alternatives

Whether or not a statistical test has power against $\sqrt{T}$-local alternative is more or less a theoretical issue rather than practical one. Just to confirm the theory, we carry out a bunch of simulations under local alternatives. For example, in conjunction with DGP A2 setting, we put the model in the alternative hypothesis as

$$x_t = 0.65x_{t-1} + \frac{2}{\sqrt{T}} y_{t-1} + \frac{4}{\sqrt{T}} \sin(-2y_{t-1}) + \eta_t$$

while DGP of y_t is unchanged. Note that choosing $T = 100$ leads to the specification in previous simulations, so the numbers in the first row in Table 3 are more or less similar to those in the first row of Table 1. When T varies from 100 to 500, the coefficients varies from 0.2 to 0.089.

As Table 3 shows, WSC test does have power against $\sqrt{T}$-local alternatives. It is striking that WSC exhibits quite a good performance for DGP A3 case while the empirical rejection rate of HNC test is as low as its empirical size, which suggests HNC test has no power effectively. On the other hand, HNC test is superior to WSC in DGP A1 case. This is a case of linear vector autoregressive model, and for such a case HNC test is specifically designed for. WSC test captures the effect of own lag (namely y_{t-1} here) by the kernel method. Thoughout this section, we used normal kernel $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$, and our bandwidth choice is $T^{-1/5}$. Because we have not yet determined the optimal bandwidth in a data dependent way, it is possible that nonparametric regression with respect to x_{t-1} might have captured the effect of cross lag variable (y_{t-1}) due to the superfluous flexibility or the excessive locality of the kernel. To construct the orthonormal basis, we employed a basis of $\{\exp(\xi_i' Z_{t-1})\}$ where ξ_i are two dimensional iid uniform random variables.

Table 3: Empirical size and power of the proposed test (WSC) and Hidalgo's nonparametric causality test (HNC): the case of local alternatives.

	DGP 0		DGP A1		DGP A2		DGP A3	
T	WSC	HNC	WSC	HNC	WSC	HNC	WSC	HNC
100	0.047	0.101	0.459	0.532	0.397	0.225	0.523	0.123
200	0.041	0.112	0.363	0.525	0.421	0.241	0.578	0.123
300	0.041	0.137	0.375	0.609	0.455	0.320	0.607	0.157
400	0.052	0.138	0.375	0.617	0.441	0.306	0.574	0.122
500	0.057	0.138	0.360	0.601	0.472	0.313	0.626	0.164

5.3 Testing Causality up to 2nd Moment

To demonstrate the usefulness of the methodology proposed in Section 4, we consider the following nonlinear time series model as a DGP.

DGP A4: $x_t = \exp(-y_{t-1}^2)\eta_t, y_t = -0.3y_{t-1} + \epsilon_t$

In this case, we have $\mathrm{E}(x_t|x_{t-1}) = 0$ and $\mathrm{E}(x_t|x_{t-1}, y_{t-1}) = 0$, but the variance of x_t depends on y_{t-1}. Moreover, η_t (innovation of x_t) multiplicatively acts on the exponential function of y_{t-1}^2. Therefore, neither our WSC test in mean nor Hidalgo's nonparametric Granger causality test seem to be able to detect such kind of dependency. (Later it will be shown that they actually do not have any power against DGP A4.) Hereafter, we refer 'WSC test in mean' to WSC1 test, and similarly use the term WSC2 test to stand for 'WSC test up to 2nd moment'.

Under the null of Granger non-Causality up to 2nd moment, y_{t-1} vanishes in the right hand side of the model of x_t, which results in a simple i.i.d. model $x_t = \eta_t$. Some simulation studies find that the size of WSC2 test depends on the choice of the bandwidth $h = CT^p$ where p is fixed to -0.2 throughout this paper. We determine C by simulation so that the 0.95 percentile coincides with the 95% critical value of WSC2 test. Though the suggested values of C can be different by sample size, it only varies between 2.2 and 2.5 in our experiments. Also note that the critical value used in WSC1 test is no longer valid for WSC2 test. Hence new critical values are generated by additional Monte Carlo experiments. We adopt 14.17 as upper 5% critical value.

We generate DGP A4 for $T = 100$, 150, 200 and 300, and the iteration number is fixed to 1,000. The results are summarized in Table 4. At first sight, HNC test seems to have some power against DGP A4 but it is not true. As can be seen in Table 1 and 2, the 'power' in Table 4 just reflects the sloppy size of HNC test, and these rejection rates could be almost equal to the size of HNC test. Compare DGP 0 and DGP A3 columns in Table 1 and 2.

Table 4: Empirical size and power of WSC test up to 2nd moment (WSC2) for DGP A4, in comparison with Hidalgo's nonparametric causality test (HNC) and WSC test in mean (WSC1).

T	HNC	WSC1	WSC2	Size(WSC2)
100	0.160	0.004	0.148	0.086
150	0.174	0.002	0.359	0.072
200	0.173	0.002	0.738	0.059
300	0.164	0.002	1.000	0.050

Apparently, WSC1 test has virtually no power and its rejection rates are far below the nominal size. This is due to the fact that the 5% critical value of WSC1 test is greater than that of WSC2 test almost by 2. On the other hand, our causality test up to 2nd moment (WSC2) shows considerably good performance when we have enough sample size (over 200), while it suffers from slight size distortion and low power in case we have only limited number of observations. Theoretical results on the choice of bandwidth parameter C under the complete lack of knowledge on the DGP process should be called for in a future research.

6 Concluding Remarks

Granger causality tests previously proposed do not detect some nonlinear causal relationships. The reason is that they construct test statistics based on a linear representation of the time series appealing to, say, the Wold decomposition, but we only know that the error terms are uncorrelated with the series of interest, not independent. Observing this fact, we proposed a nonparametric testing procedure for Granger-type causality in the case of any form of nonlinear dependence. We also show that the test has nontrivial power against $\sqrt{T}$-local alternatives. The Monte Carlo study shows that the WSC1 and WSC2 tests perform very well with quite decent empirical size in general and good power property for causality in mean and causality in second moment respectively. Compared with HNC test, WSC1 does slightly worse when the series exhibit linear dependence in mean in fact, but in the case of nonlinear dependence, it obviously outperforms HNC.

A APPENDIX: Proofs of Theorems

(*Sketch of the proof of Theorem 2*) Due to the results in Singh and Ullah (1985), we have

$$T^{1/4}\sup_{\xi}|\hat{h}_i(\xi) - h_i(\xi)| + T^{1/4}\sup_{\xi}|\hat{g}(\xi) - g(\xi)| + T^{1/4}\sup_{\xi}|\hat{f}(\xi) - f(\xi)| = o_p(1) \quad \text{(A.1)}$$

under A1-A4. We drop the subscript i for notational simplicity. Writing

$$\hat{\sigma}\hat{a} = \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} u_t h(Z_{t-1}) + \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} u_t\{\hat{h}(Z_{t-1}) - h(Z_{t-1})\}$$
$$+ \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \{g(X_{t-1}) - \hat{g}(X_{t-1})\}h(Z_{t-1})$$
$$+ \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \{g(X_{t-1}) - \hat{g}(X_{t-1})\}\{\hat{h}(Z_{t-1}) - h(Z_{t-1})\}, \tag{A.2}$$

the first term on the right of (A.2) converges in distribution to a normal random variate with mean zero and variance σ^2 appealing to a central limit theorem for martingale difference sequences. The last term of (A.2) is $o_p(1)$ because of (A.1). We evaluate the other two terms rewriting them into the U statistic form. Since we can handle them similarly, we deal only with the latter. Expanding $\{\hat{f}(X_{t-1})\}^{-1}$ around $\{f(X_{t-1})\}^{-1}$, and applying A5, we have

$$\frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \{g(X_{t-1}) - \hat{g}(X_{t-1})\}h(Z_{t-1})$$
$$= \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \left\{g(X_{t-1}) - \frac{\hat{r}(X_{t-1})}{f(X_{t-1})}\right\} h(Z_{t-1})$$
$$+ \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \frac{h(Z_{t-1})g(X_{t-1})}{f(X_{t-1})}\{\hat{f}(X_{t-1}) - f(X_{t-1})\} + o_p(1). \tag{A.3}$$

Noting $w(Z_{t-1}) = f(X_{t-1})^{-1}h(Z_{t-1})g(X_{t-1})$, the second term on the right of (A.3) has an expression

$$\frac{1}{\sqrt{T}}\sum_{t=r+1}^{T} \frac{h(Z_{t-1})g(X_{t-1})}{f(X_{t-1})}\{\hat{f}(X_{t-1}) - f(X_{t-1})\}$$
$$\approx \sqrt{T}\binom{T}{2}^{-1}\sum_{t=r+1}^{T-1}\sum_{s=t+1}^{T} U(Z_{s-1}, Z_{t-1}) \tag{A.4}$$

where

$$U(Z_{s-1}, Z_{t-1}) = \frac{1}{2}\left[\left\{\frac{1}{h^p}K\left(\frac{X_{t-1} - X_{s-1}}{h}\right) - f(X_{t-1})\right\} w(Z_{t-1})\right.$$
$$\left. + \left\{\frac{1}{h^p}K\left(\frac{X_{s-1} - X_{t-1}}{h}\right) - f(X_{s-1})\right\} w(Z_{s-1})\right]. \tag{A.5}$$

Due to assumptions A1 and A4,

$$u_1(z) = \mathrm{E}\{U(Z_{s-1}, z)\} = \frac{1}{2}\left\{\int \frac{1}{h^p} K\left(\frac{x-\xi_x}{h}\right) f(\xi_x, \xi_y) d\xi - f(x)\right\} w(z)$$

$$= \frac{1}{2}\left\{\frac{h^L}{L!} f^{(L)}(x) \sum_{0\le l_1,\cdots,l_p\le L,} \cdots \sum_{l_1+\cdots+l_p=L} \left\{\int \prod_{i=1}^{p} u_i^{l_i} K(u) du\right\} f^{(l_1,\cdots,l_p)}(x) + o(h^L)\right\} w(z), \tag{A.6}$$

where ξ_x is a p-vector, $\xi = (\xi_x, \xi_y)$ is a $(p+q)$-vector. The third equality uses variable transformation and A3. Define $\theta = \mathrm{E}[u_1(Z_{t-1})]$, $U_1(Z_{t-1}) = u_1(Z_{t-1}) - \theta$ and $U_2(Z_{s-1}, Z_{t-1}) = U(Z_{s-1}, Z_{t-1}) - u_1(Z_{s-1}) - u_1(Z_{t-1}) + \theta$, then as is standard in U statistic theory, we have the H-decomposition

$$\sqrt{T}\binom{T}{2}^{-1} \sum_{t=r+1}^{T-1} \sum_{s=t+1}^{T} U(Z_{s-1}, Z_{t-1}) = \frac{2}{\sqrt{T}} \sum_{t=r+1}^{T} U_1(Z_{t-1})$$

$$+ \sqrt{T}\binom{T}{2}^{-1} \sum_{t=r+1}^{T-1} \sum_{s=t+1}^{T} U_2(Z_{s-1}, Z_{t-1}) + \sqrt{T}\theta. \tag{A.7}$$

Note

$$\mathrm{E}\{U_1(Z_{t-1})\} = \mathrm{E}\{U_2(Z_{s-1}, z)\} = \mathrm{E}\{U_2(z, Z_{t-1})\} = 0 \tag{A.8}$$

for any z. Because of A4, we have

$$\theta = \mathrm{E}\{u_1(Z_{s-1})\} = O(h^L) \tag{A.9}$$

Putting the first term on the right of (A.7) be $U^{(1)}$ and the second be $U^{(2)}$, we have

$$V(U^{(1)}) = \frac{1}{T} \sum_{t=r+1}^{T} V(U_1(Z_{t-1})) + \frac{2}{T} \sum_{t=r+1}^{T-1} \sum_{s=t+1}^{T} \mathrm{Cov}(U_1(Z_{s-1}), U_1(Z_{t-1})). \tag{A.10}$$

The first term on the right is $O(h^{2L})$ due to (A.6) and (A.9), while the summand of the second term equals to

$$\int U_1(\xi_1) U_1(\xi_2) dF_{t,s}(\xi_1, \xi_2) + O(h^{2L}). \tag{A.11}$$

Because of the Lemma, we have

$$\left|\int U_1(\xi_1) U_1(\xi_2) dF_{t,s}(\xi_1, \xi_2) - \int U_1(\xi_1) U_1(\xi_2) dF(\xi_1) dF(\xi_2)\right| \le C\{\beta(s-t)\}^{\delta/(2+\delta)} \tag{A.12}$$

but the second integral equals to zero by (A.8), which yields

$$\int U_1(\xi_1)U_1(\xi_2)dF_{t,s}(\xi_1,\xi_2) \leq Ch^4(s-t)^{-\frac{(2+\eta)\delta}{(2+\delta)\eta}} = Ch^4(s-t)^{-1-\gamma}, \tag{A.13}$$

where $\gamma = 2(\delta-\eta)/\eta(2+\delta)$, and thus taking $\delta > \eta$,

$$\frac{1}{T}\sum_{t=r+1}^{T-1}\sum_{s=t+1}^{T}\text{Cov}(U_1(Z_{s-1}),U_1(Z_{t-1})) \leq \frac{Ch^4}{T}\sum_{t=r+1}^{T-1}\sum_{s=t+1}^{T}(s-t)^{-1-\gamma} = O(h^4). \tag{A.14}$$

Therefore $U^{(1)} = o_p(1)$.

We next show $U^{(2)} = o_p(1)$. Because of (A.8) and the Lemma,

$$\begin{aligned} h^p\text{E}\{U_2(Z_{s-1},Z_{t-1})\} &= h^pE\{U(Z_{s-1},Z_{t-1})\} - h^p\theta \\ &\leq h^p\left|\int U(\xi_1,\xi_2)dF_{t,s}(\xi_1,\xi_2) - \int U(\xi_1,\xi_2)dF(\xi_1)dF(\xi_2)\right| \\ &\leq Ch^{p(2+\delta)}\{\beta(s-t)\}^{\delta/(2+\delta)} \end{aligned} \tag{A.15}$$

so that

$$\text{E}\{U^{(2)}\} \leq \frac{Ch^{p(1+\delta)}}{T^{3/2}}\sum_{t=r+1}^{T-1}\sum_{s=t+1}^{T}(s-t)^{-1-\gamma} = O\left(\frac{h^{p(1+\delta)}}{T^{1/2}}\right). \tag{A.16}$$

Using this, write

$$\text{Var}\{U^{(2)}\} \leq \frac{C}{T^3h^{2p}}\text{E}\left\{\sum_{t=r+1}^{T-1}\sum_{s=t+1}^{T}h^pU_2(Z_{s-1},Z_{t-1})\right\}^2 + O\left(\frac{h^{2p(1+\delta)}}{T}\right) \tag{A.17}$$

For $t < s < u < v$, we have

$$\begin{aligned} &h^{2p}\text{E}\{U_2(Z_{t-1},Z_{s-1})U_2(Z_{u-1},Z_{v-1})\} \\ &\leq h^{2p}\left|\int U_2(\xi_1,\xi_2)U_2(\xi_3,\xi_4)dF_{t,s,u,v}(\xi_1,\xi_2,\xi_3,\xi_4)\right. \\ &\left.-\int U_2(\xi_1,\xi_2)U_2(\xi_3,\xi_4)dF(\xi_1)dF_{s,u,v}(\xi_2,\xi_3,\xi_4)\right| \\ &\leq Ch^{p(3+\delta)}(s-t)^{-1-\gamma}, \end{aligned} \tag{A.18}$$

yielding

$$\left|\sum_{t=r+1}^{T-3}\sum_{s=t+1}^{T-2}\sum_{u=s+1}^{T-1}\sum_{v=u+1}^{T}h^{2p}\text{E}\{U_2(Z_{s-1},Z_{t-1})U_2(Z_{u-1},Z_{v-1})\}\right| = O\left(T^3h^{p(3+\delta)}\right). \tag{2.37}$$

For other cases such as $t < u < s < v, u < t < s < v$ and so on can be treated similarly For $t < s < u$, we have

$$\begin{aligned} & h^{2p}\mathrm{E}\{U_2(Z_{t-1}, Z_{s-1})U_2(Z_{s-1}, Z_{u-1})\} \\ & \leq h^{2p} \left| \int U_2(\xi_1, \xi_2)U_2(\xi_2, \xi_3)dF_{t,s,u}(\xi_1, \xi_2, \xi_3) \right. \\ & \left. - \int U_2(\xi_1, \xi_2)U_2(\xi_2, \xi_3)dF(\xi_1)dF_{s,u}(\xi_2, \xi_3) \right| \\ & \leq Ch^{p(3+\delta)}(s-t)^{-1-\gamma}, \end{aligned} \tag{A.19}$$

yielding

$$\left| \sum_{t=r+1}^{T-2} \sum_{s=t+1}^{T-1} \sum_{u=s+1}^{T} h^{2p}\mathrm{E}\{U_2(Z_{s-1}, Z_{t-1})U_2(Z_{t-1}, Z_{u-1})\} \right| = O\left(T^2 h^{p(3+\delta)}\right). \tag{A.20}$$

Other cases with only one tie are similar. For $t < s$, we have due to the Lemma and $M = O(h^{-p(1+\delta)})$,

$$h^{2p} \left| \int U_2(\xi_1, \xi_2)^2 dF_{t,s}(\xi_1, \xi_2) - \int U_2(\xi_1, \xi_2)^2 dF(\xi_1)dF(\xi_2) \right| \leq Ch^{2p}(s-t)^{-1-\gamma}, \tag{A.21}$$

where the second integral is

$$\int U_2(\xi_1, \xi_2)^2 dF(\xi_1)dF(\xi_2) \leq \frac{C}{h^p} \int f(\xi)w(\xi)^2 d\xi = O(h^{-p}). \tag{A.22}$$

Thus, we have

$$\left| \sum_{t=r+1}^{T-1} \sum_{s=t+1}^{T} h^{2p}\mathrm{E}\{U_2(Z_{s-1}, Z_{t-1})\}^2\} \right| = O\left(T^2 h^{p}\right). \tag{A.23}$$

Combining the above results, we have

$$\mathrm{Var}\{U^{(2)}\} = O\left(h^{p(3+\delta)}\right). \tag{A.24}$$

Therefore we have (A.3)=$o_p(1)$. The second term on the right of (A.2) is handled similarly to show it is $o_p(1)$ rewriting it into U statistic form. □

(*Sketch of the proof of Theorem 6*)

In view of Theorem 5, it suffices to show that

$$\hat{B}_i - B_i \xrightarrow{p} 0, \tag{A.25}$$

and

$$\hat{\Sigma}_i - \Sigma_i \xrightarrow{p} 0, \tag{A.26}$$

for all i. We drop the subscript i for notational simplicity in the sequel.

To prove (A.25), write

$$\hat{B} = \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \{U_t - E(U_t|X_{t-1})\} h(Z_{t-1})$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} (\hat{U}_t - U_t) h(Z_{t-1}) \tag{A.27}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \{\hat{E}(U_t|X_{t-1}) - E(U_t|X_{t-1})\} h(Z_{t-1}) \tag{A.28}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \{U_t - E(U_t|X_{t-1})\}\{\hat{h}(Z_{t-1}) - h(Z_{t-1})\} \tag{A.29}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} (\hat{U}_t - U_t)\{\hat{h}(Z_{t-1}) - h(Z_{t-1})\} \tag{A.30}$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T} \{\hat{E}(U_t|X_{t-1}) - E(U_t|X_{t-1})\}\{\hat{h}(Z_{t-1}) - h(Z_{t-1})\} \tag{A.31}$$

$$= (I) + (II) + (III) + (IV) + (V) + (VI), \tag{A.32}$$

then

$$(I) \xrightarrow{d} N(0, \Sigma) \tag{A.33}$$

by Theorem 5. We show in the following that $(II), (III), (IV), (V), (VI) \xrightarrow{p} 0$. Because

$$\hat{U}_t - U_t = M(u_t) \begin{bmatrix} g(Z_{t-1}) - \hat{g}(Z_{t-1}) \\ \{g(Z_{t-1}) - \hat{g}(Z_{t-1})\}^2 \\ \\ \{g(Z_{t-1}) - \hat{g}(Z_{t-1})\}^K \end{bmatrix} \tag{A.34}$$

(V) is $o_p(1)$ due to (A.1) and Assumption A6, where

$$M(u_t) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 2u_t & 1 & 0 & \cdots & 0 \\ 3u_t^2 & 3u_t & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{K}{K-1} u_t^{K-1} & \binom{K}{K-2} u_t^{K-2} & \binom{K}{K-3} u_t^{K-3} & \cdots & 1 \end{bmatrix}. \tag{A.35}$$

Since $\sup_x|\hat{E}(U_t|X_{t-1}=x)-E(U_t|X_{t-1}=x)| = o_p(T^{-\frac{1}{4}})$, we also know $(VI) \xrightarrow{p} 0$ similarly.

We handle the other terms employing the same technique used in the proof of Theorem 2, or U statistic asymptotic theory for the absolutely regular proccesses. We prove $(II) \xrightarrow{p} 0$. Since

$$(II) = \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T}\left\{M(u_t)\begin{bmatrix} g(Z_{t-1})-\hat{g}(Z_{t-1}) \\ \{g(Z_{t-1})-\hat{g}(Z_{t-1})\}^K \end{bmatrix}\right\}h(Z_{t-1}) \tag{A.36}$$

and

$$\sup|g(z)-\hat{g}(z)|^k = o_p(T^{-\frac{1}{2}}) for k = 2,\cdots,K, \tag{A.37}$$

it suffices to show

$$\frac{1}{\sqrt{T}}\sum_{t=r+1}^{T}u_t^{K-1}\{g(Z_{t-1})-\hat{g}(Z_{t-1})\}h(Z_{t-1}) = o_p(1). \tag{A.38}$$

The left hand side has a U-statistic form of

$$\frac{2}{\sqrt{T}}\sum_{t=r+1}^{T}\psi_t + \frac{2}{T\sqrt{T}}\sum_{t=r+1}^{T}\sum_{s=r+1}^{T}\psi_{ts} \tag{A.39}$$

where

$$\psi_t = C(X_{t-1})u_t^{K-1}h^L h(Z_{t-1}) + o(h^L). \tag{A.40}$$

and

$$\psi_{ts} = \frac{1}{2}\left[u_t^{K-1}\left\{\frac{1}{h^p f(X_{t-1})}K\left(\frac{X_{t-1}-X_{s-1}}{h}\right)x_s - g(X_{t-1})\right\}h(Z_{t-1})\right. \tag{A.41}$$

$$\left. + u_s^{K-1}\left\{\frac{1}{h^p f(X_{s-1})}K\left(\frac{X_{s-1}-X_{t-1}}{h}\right)x_t - g(X_{s-1})\right\}h(Z_{s-1})\right] - \psi_t - \psi_s + E(\psi_t) \tag{A.42}$$

for a function $C: R^p \to R$ satisfying $E|C(X_{t-1})| < \infty$. Applying the same technique used in the proof of Theorem 2 and Assumptions A3 and A6, (A.38) is verified. We can prove (IV) is $o_p(1)$ similarly to (II). To prove $(III) = o_p(1)$, noting

$$\hat{E}(U_t|X_{t-1}) - E(U_t|X_{t-1}) \tag{A.43}$$

$$= \frac{1}{T}\sum_{s=r+1}^{T}\left\{\frac{1}{h^p\hat{f}(X_{t-1})}K\left(\frac{X_{t-1}-X_{s-1}}{h}\right)U_s - E(U_t|X_{t-1})\right\} \tag{A.44}$$

$$+ \frac{1}{T}\sum_{s=r+1}^{T}\left\{\frac{1}{h^p\hat{f}(X_{t-1})}K\left(\frac{X_{t-1}-X_{s-1}}{h}\right)M(u_t)\begin{bmatrix} g(Z_{t-1})-\hat{g}(Z_{t-1}) \\ \{g(Z_{t-1})-\hat{g}(Z_{t-1})\}^K \end{bmatrix}\right\}, \tag{A.45}$$

write

$$(III) = \frac{1}{T\sqrt{T}} \sum_{t=r+1}^{T} \sum_{s=r+1}^{T} \left\{ \frac{1}{h^p \hat{f}(X_{t-1})} K\left(\frac{X_{t-1} - X_{s-1}}{h}\right) U_s - E(U_t|X_{t-1}) \right\} h(Z_{t-1}) \tag{A.46}$$

$$+ \frac{1}{T\sqrt{T}} \sum_{t=r+1}^{T} \sum_{s=r+1}^{T} \left\{ \frac{1}{h^p \hat{f}(X_{t-1})} K\left(\frac{X_{t-1} - X_{s-1}}{h}\right) M(u_t) \right. \tag{A.47}$$

$$\left. \begin{bmatrix} g(Z_{t-1}) - \hat{g}(Z_{t-1}) \\ \{g(Z_{t-1}) - \hat{g}(Z_{t-1})\}^K \end{bmatrix} \right\} h(Z_{t-1}). \tag{A.48}$$

The first term on the right is shown to be $o_p(1)$ similarly to (II), while for the second we only need to consider

$$\frac{1}{T\sqrt{T}} \sum_{t=r+1}^{T} \sum_{s=r+1}^{T} \left[\frac{1}{h^p \hat{f}(X_{t-1})} K\left(\frac{X_{t-1} - X_{s-1}}{h}\right) u_t^{K-1} \{g(Z_{t-1}) - \hat{g}(Z_{t-1})\} \right] h(Z_{t-1}) \tag{A.49}$$

$$= \frac{1}{T^2\sqrt{T}} \sum_{t=r+1}^{T} \sum_{s=r+1}^{T} \sum_{v=r+1}^{T} \left[\frac{1}{h^p \hat{f}(X_{t-1})} K\left(\frac{X_{t-1} - X_{s-1}}{h}\right) u_t^{K-1} \right. \tag{A.50}$$

$$\left. \left\{ g(Z_{t-1}) - \frac{1}{h^p \hat{f}(X_{t-1})} K\left(\frac{X_{v-1} - X_{s-1}}{h}\right) x_v \right\} \right] h(Z_{t-1}) \tag{A.51}$$

which has a U statistic form of degree three, and its projection (in terms of U statistic theory) is shown to be of order $O_p(h^L)$ and the other terms are $o_p(1)$ by assumption A3.

We next prove (A.26). Writing

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=r+1}^{T} \left\{ \hat{E}(U_t U_t'|Z_{t-1}) - \hat{E}(U_t|Z_{t-1}) \hat{E}(U_t'|Z_{t-1}) \right\} h(Z_{t-1})^2$$
$$+ \frac{1}{T} \sum_{t=r+1}^{T} \left\{ \hat{E}(U_t U_t'|Z_{t-1}) - \hat{E}(U_t|Z_{t-1}) \hat{E}(U_t'|Z_{t-1}) \right\} \{\hat{h}(Z_{t-1})^2 - h(Z_{t-1})^2\},$$

the second term on the right is bounded by

$$\sup |\hat{h}(z) - h(z)| \frac{1}{T} \sum_{t=r+1}^{T} \left| \left\{ \hat{E}(U_t U_t'|Z_{t-1}) - \hat{E}(U_t|Z_{t-1}) \hat{E}(U_t'|Z_{t-1}) \right\} \{\hat{h}(Z_{t-1}) + h(Z_{t-1})\} \right|$$
$$= o_p(T^{-\frac{1}{4}}),$$

due to (A.1). The first term equals to

$$\frac{1}{T}\sum_{t=r+1}^{T}\left\{E(U_tU_t'|Z_{t-1})-E(U_t|Z_{t-1})E(U_t'|Z_{t-1})\right\}h(Z_{t-1})^2 \tag{A.52}$$

$$+\frac{1}{T}\sum_{t=r+1}^{T}\left\{\hat{E}(U_tU_t'|Z_{t-1})-E(U_tU_t'|Z_{t-1})\right\}h(Z_{t-1})^2 \tag{A.53}$$

$$-\frac{1}{T}\sum_{t=r+1}^{T}\left\{\hat{E}(U_t|Z_{t-1})\hat{E}(U_t'|Z_{t-1})-E(U_t|Z_{t-1})E(U_t'|Z_{t-1})\right\}h(Z_{t-1})^2. \tag{A.54}$$

(A.52) converges to Σ in probability by law of large numbers for martingale difference sequences under assumption A6. The other two terms are shown to be $o_p(1)$ similarly to the proof that $(II),(III)=o_p(1)$. $\square$

References

[1] Denker, M. and G. Keller (1983), "On U-Statistics and v. Mises' Statistics for Weakly Dependent Processes," Zeitschrift fur Wahrscheinlichkeitstheorie und verwande Gebiete, 64, 505-522.

[2] Geweke, J. (1982), "Measurement for Linear Dependence and Feedback between Multiple Time Series," JASA, 77, 303-323.

[3] Granger, C.W.J., (1969) "Investigating Causal Relations by Econometrics Models And Cross-Spectral Methods," Econometrica, 37, 3, 424-438.

[4] Hidalgo, J. (2000), "Nonparametric Test for Causality with Long-Range Dependence," Econometrica, 68, 6, 1465-1491.

[5] Hosoya, Y. (1977) "On the Granger Condition for Non-Causality," Econometrica, 45, 7, 1735-1736.

[6] Hosoya, Y. (1991) "The Decomposition and Measurement of the Interdependency between Second-Order Stationary Processes," Probability Theory and Related Fields, 88, 429-444.

[7] Lee, A.J. (1990), U-Statistics: Theory and Practice, Marcel Dekker.

[8] Lutkepohl, H. and D.S. Poskitt (1996) "Testing for Causation using Infinite Order Vector Autoregressive Processes," Econometric Theory, 12, 61-87.

[9] Okui, R. and K. Hitomi (2002), "A Consistent Test for Omitted Variables in Nonparametric Regression," mimeo.

[10] Robinson, P.M., (1983) "Nonparametric Estimators for Time Series," Journal of Time Series Analysis, 4, 3, 185-207.

[11] Robinson, P.M., (1989) "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," Review of Economic Studies, 56, 511-534.

[12] Sims, C.A., (1972) "Money, Income, And Causality, " American Economic Review, 62, 540-552.

[13] Sims, C.A., J.H. Stock and M.W. Watson (1990) "Inference in Linear Time Series Models with some Unit Roots," Econometrics, 58, 113-144.

[14] Singh, R.S. and A. Ullah (1985) "Nonparametric Time-Series Estimation of Joint DGP, Conditional DGP, and Vector Autoregression," Econometric Theory, 1, 27-52.

[15] Stinchcombe, M.B. and H. White (1998) "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," Econometric Theory, 14, 295-325.

[16] Toda, H.Y. and P.C.B. Phillips (1993) "Vector Autocorrelation and Causality," Econometrica, 61, 1367-1393.

[17] Yoshihara, K. (1976) "Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes," Zeitschrift fur Wahrscheinlichkeitstheorie und verwande Gebiete, 35, 237-252.