A Consistent Nonparametric Test for Nonlinear Causality[1.]

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

September 27, 2010

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 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 a nonparametric test for possibly nonlinear causality. Taking into account that dependence in higher order moments are becoming an important issue especially in financial time series, we also consider a test for causality up to the K-th conditional moment. Statistically, we can view this test as a nonparametric omitted variable test in time series regression. A desirable property of the test is that it has nontrivial power against $T^{1/2}$-local alternatives, where T is the sample size. Also, we can form a test statistic accordingly if we have some knowledge on the alternative hypothesis. Furthermore, we show that the test statistic includes most of the nonparametric omitted variables test statistics as special cases asymptotically. The null asymptotic distribution is not normal, but we can easily calculate the critical regions by simulation. Monte Carlo experiments show that the proposed test has good size and power properties.

Keywords: Nonlinear causality; Nonparametric test; Omitted variables test; Causality up to K-th moment; Local alternatives; Consistent test

JEL classifications: C12, C14, C32

[1]We thank two anonymous referees for helpful comments and suggestions. We are also grateful for seminar/meeting participants at Kyoto university and Waseda University for helpful comments and discussions.

This research is 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-in-Aid, 15330040, 22330067, The second author gratefully acknowledge the financial support by Grant-in-Aid for Scientific research 19530182. The third author is supported by Grant-in-Aid for Scientific research 21243019, 21500287 and 22530257.

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

1 Introduction

Causality between variables has been one of the main interests in time series econometrics since the pioneering work by Granger (1969). We propose a nonparametric test for Granger-type causality. A conceptually similar work is Bierens and Ploberger (1997), Chen and Fan (1999), Robinson (1989), and Hidalgo (2000). The first three papers proposed nonparametric tests on certain conditional moment restrictions, while the last paper introduced a nonparametric Granger causality test in the frequency domain for weakly stationary linear processes. Hidalgo is mainly concerned with the test under long range dependent observations, but it does not have power against some alternatives of series with nonlinear dynamics. We construct a test statistic based on moment conditions allowing for nonlinear dependence. It has nontrivial power against $T^{1/2}$-local alternatives, where T is the sample size. The 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 not an easy concept to capture philosophically, 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's defintion is that y_t is said to cause x_t in mean if

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

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 that 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, and not independent. In many of the aforementioned research, the author(s) apply frequency domain analysis

where causality is captured through cross-spectra, 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 good power property against certain alternatives. Péguin-Feissolle and Teräsvirta (1999) also propose a causality test for nonlinear AR model, by expanding the nonlinear conditional expectation by a Taylor series and testing if the coefficients are zero or not. They study the properties of the test by simulation without theoretical justification and conclude that it is useful. Chao,Corradi and Swanson (2001) discuss nonlinear causality testing in the context of out-of-sample tests. The test statistic can be considered as a simpler version of Bierens (1990), and it is not consistent in general. We also refer to a series of papers which treats long-run causality, namely predictability multiple periods ahead such as Breitung and Candelon (2006), Dufour, Pelletier and Renault (2006) and Dufour and Renault (1998). McCrorie and Chamebers (2006) point out the problem of spurious causality from aggregation of a continuous time series to a discrete series.

We propose a nonparametric test which has power even when the observations are nonlinearly dependent. 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

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

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

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 "causality in mean" throughout this paper. Our definition of causality is different from the usage of the word in economics or daily life, but it is the matter of predictability. We would like 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). Here "with probability one" is abbreviated to w.p.1.

We construct test statistics based on the moment conditions (1.4) for causality in mean. This is, in statistical terms, a test for omitted variables in time series

regression. Many such tests have been proposed in the literature, for example in Bierens (1982, 1990), Bierens and Ploberger (1997), Chen and Fan (1999), Fan and Li (1996) and Robinson (1989) for i.i.d. and time series observations, among others. All these papers except the last two have nontrivial power against $\sqrt{T}-$local alternative. The test proposed in this paper also has nontrivial power against $\sqrt{T}-$local alternatives, but it has an advantage over the previous ones in that the proposed test can control the power properties easily and directly. Furthermore, we can show that these previously proposed tests can be rewritten, in fact, as special cases of the test statistic proposed below, by selecting user-determined components suitably. We will see this in detail in Section 5. Hiemstra and Jones (1994) also consider testing for nonlinear Granger causality, but it differs from our approach because they look at the conditional distribution, not conditional moments. Their definition of causality is slightly stronger than the present one because they define the non-causality as $F(x_t|x_{t-1},\ldots,x_1,y_{t-1},\ldots,y_1) = F(x_t|x_{t-1},\ldots,x_1)$, where $F(\cdot|\cdot)$ denotes a conditional distribution. Therefore, non-causality in their sense implies the present non-causality (1.4). In our impression, Hiemstra-Jones definition may sometimes be too strong in practice. Furthermore, Dicks and Panchenko (2005) point out an inconsistency problem of the test. We also refer to Qiao, McAleer and Wong (2009) who applied Hiemstra and Jones test to investigate consumer attitudes. The next section provides the test statistic for causality in mean and its null distribution as well as the regularity conditions. Section 3 explains the power properties of the test. Section 4 provides causality in higher order moments. Section 5 discusses about the power properties as well as special cases of the test. We report Monte Carlo results in Section 6. Section 7 concludes this paper. The proof of the theorem and Lemmas 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, 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 null hypothesis $m(Z_{t-1}) = g(X_{t-1})$ 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

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

and

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

We further rewrite the hypotheses in terms of unconditional moment restrictions. 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

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

In order for a technical reason, we slightly modify this null hypothesis to so-called "density-weighted" version;

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

It is obvious that all four representations above are equivalent because $f(X_{t-1})$ is measurable conditionally on Z_{t-1}. Thus, we can rewrite the null and alternative hyptoheses as,

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

and

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

2.2 Test statistics

We first heuristically describe the idea of constructing the test statistic. Let $H(z) = \{h_i(z)\}_{i=1}^{\infty}$ be a complete basis over $s_X^{\perp}$, satisfying

$$Var[u_t f(X_{t-1}) h_i(Z_{t-1})] = 1 \tag{2.1}$$

and

$$Cov[u_t f(X_{t-1}) h_i(Z_{t-1}), u_t f(X_{t-1}) h_j(Z_{t-1})] = 0 \tag{2.2}$$

for all $i \neq j$. Given a sample $\{(x_t, y_t)\}_{t=1}^{T}$, define, for $r = \max(p, q) + 1$ and $i = 1, \cdots, k_T$,

$$a_i = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) h_i(Z_{t-1}), \tag{2.3}$$

fo some $k_T < T$ which is determined later. Appealing to a central limit theorem for martingale difference sequences, we have $a_i \xrightarrow{d} N(0,1)$ under the null, while $|a_i|$ explodes to ∞ under the alternative for some i as $T \to \infty$, because u_t has a non-zero mean conditionally on Z_{t-1} and $\{h_i(Z)\}$ spans a basis of $s_X^{\perp}$. Therefore we can consider a test combining these quantities as follows. Let $\{w_i\}_{i=1}^{\infty}$ be a user determined summable positive sequence, such as $w_i = 0.9^i$, then

$$S_T = \sum_{i=1}^{k_T} w_i a_i^2 \xrightarrow{d} \sum_{i=1}^{\infty} w_i \epsilon_i^2 \tag{2.4}$$

where ϵ_i are i.i.d.$N(0,1)$ random variables. It is obvious $S_T \xrightarrow{d} \infty$ under the alternative because some of a_i must explode.

We can construct $H(Z)$ as follows. Given $\{q_i(Z)\}_{i=1}^{\infty}$, a user-determined basis of s_Z, $[\{q_i(Z) - r_i(X)\}f(X)]_{i=1}^{\infty}$ forms a basis of $s_X^{\perp}$, where $r_i(X) = E[q_i(Z_{t-1})|X_{t-1} = X]$. Let

$$Q(Z)' = (q_1(Z) - r_1(X), \cdots, q_{k_T}(Z) - r_{k_T}(X))f(X) = (Q_1(Z), \cdots, Q_{k_T}(Z))',$$

and

$$M = E[u_t^2 f(X_{t-1})^2 Q(Z_{t-1})Q(Z_{t-1})'].$$

Then, supposing M is positive definite,

$$H(Z) = (h_1(Z), \cdots, h_{k_T}(Z))' = M^{-1/2}Q(Z) \tag{2.5}$$

satisfies the required conditions (2.1) and (2.2).

S_t provides a population quantity to test the null hypothesis (1.4), however (2.4) is infeasible. We propose a feasible plug-in test statistic. The unknown components in (2.3) are u_t, $f(X)$ and $h_i(Z)$ which are replaced by the regression residual and their estimates. We estimate the density by a kernel method as,

$$\hat{f}(X) = \frac{1}{Th^p} \sum_{t=p+1}^{T} K(\frac{X - X_{t-1}}{h}) \tag{2.6}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth decaying to zero as $T \to \infty$. We obtain the residuals straightforwardly by

$$\hat{u}_t = x_t - \hat{g}(X_{t-1}) \tag{2.7}$$

where

$$\hat{g}(X) = \frac{1}{Th^p \hat{f}(X)} \sum_{t=p+1}^{T} K(\frac{X - X_{t-1}}{h})x_t \tag{2.8}$$

is a nonparametric kernel estimate of $g(X)$. Using (2.6) and (2.7), we estimate $u_t f(X_{t-1})$ by

$$\widehat{u_t f}(X_{t-1}) = x_t \hat{f}(X_{t-1}) - \frac{1}{Th^p} \sum_{s=p+1}^{T} K(\frac{X_{t-1} - X_{s-1}}{h}) x_s. \tag{2.9}$$

We construct estimates of $h_i(Z)$ by replacing unknown quantities in (2.5) as follows. Since $r_i(X)$=E$[q_i(Z_{t-1})|X_{t-1} = X]$ is unknown, we estimate it by Nadaraya-Watson kernel estimator

$$\hat{r}_i(X) = \hat{\mathrm{E}}[q_i(Z_{t-1})|X_{t-1} = X] = \frac{1}{Th^p \hat{f}(X)} \sum_{t=p+1}^{T} K(\frac{X - X_{t-1}}{h}) q_i(Z_{t-1}),$$

and $\hat{f}(X)$ is used again. Then we have an estimate for $Q(Z)$ as,

$$\begin{aligned} \hat{Q}(Z)' &= (q_1(Z) - \hat{r}_1(X), \cdots, q_{k_T}(Z) - \hat{r}_{k_T}(X)))\hat{f}(X). \\ &= (\hat{Q}_1(Z), \cdots, \hat{Q}_{k_T}(Z))' \end{aligned}$$

M is estimated by its feasible sample analogue,

$$\hat{M} = \frac{1}{T} \sum_{t=r}^{T} \widehat{u_t f}(X_{t-1})^2 \hat{Q}(Z_{t-1}) \hat{Q}(Z_{t-1})'.$$

Using these estimates, we finally produce $\{\hat{h}_i(Z)\}_{i=1}^{k_T}$ by

$$\hat{H}(Z) = (\hat{h}_1(Z), \cdots, \hat{h}_{k_T}(Z))' = \hat{M}^{-1/2} \hat{Q}(Z). \tag{2.10}$$

Substituting (2.9) and (2.10) into (2.3), we make a sample analogue of (2.4),

$$\hat{S}_T = \sum_{i=1}^{k_T} w_i \hat{a}_i^2, \tag{2.11}$$

where

$$\hat{a}_i = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \widehat{u_t f}(X_{t-1}) \hat{h}_i(Z_{t-1}).$$

We give several remarks. Firstly, we use the density weight because $u_t = x_t - \mathrm{E}[x_t|X_{t-1}]$ and $q(Z_{t-1}) - \mathrm{E}[q(Z_{t-1})|X_{t-1}]$ involves a denominator of $f(X_{t-1})$. If we replace it with its estimate in constructing a feasible statistic, the random denominator may cause a numerical problem by taking a very small number in practice, which results in a very large value of $|\hat{\mathrm{E}}[x_t|X_{t-1}]|$ or $|\hat{\mathrm{E}}[q(Z_{t-1})|X_{t-1}]|$. We would like to avoid such unstable feature of the test statistic. Many papers

in nonparametric and semiparametric statistics treat this problem similarly. Secondly, this procedure allows for conditional heteroscedasticity in u_t because $H(Z)$ absorbs it due to the construction of (2.1) and (2.2). It will be convenient in practice that we need not explicitly estimate the conditional variance $V(u_t|Z_{t-1})$ which will entail a multi-step procedure. Thirdly, we de-mean in making $\hat{Q}(z)$ in order for $\frac{1}{\sqrt{T}}\sum_{t=p}^{T}\{\hat{m}(X_{t-1})-m(X_{t-1})\}\hat{h}_i(Z_{t-1})$ to be degenerated for any function $m(X)$ and its kernel estimate $\hat{m}(X)$ owing to this construction. This is convenient to prove the asymptotic equivalence of a_i and $\hat{a}_i$. Fourthly, it is also possible to consider a test statistic with equal weights, or $k_T^{-1}\sum_{i=1}^{k_T} a_i^2$, however it does not have power against $\sqrt{T}$-local alternatives. In practice, it is not easy to say which performs better in small samples, but the present procedure performs better in large samples. We dicuss about it in Section 3. Finally, it is also possible to test the same hypothesis by integrated conditional moment tests (ICM) by Bierens (1984, 1990) and others. They also have power against $\sqrt{T}$-local alternatives, but it is not clear how much power it has toward which direction of departure from the null unlike ours. Also, if one has an idea on which direction the alternative may depart from the null, we can include this information in the present test, by choosing w_i and $q_i(Z)$ suitably. We will discuss about this issue in detail in Section 3.

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 $\{z_t\}, t = 1, \ldots, T$ 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 $\{z_s, \ldots, z_t\}$. Then we say $\{z_t\}, t = 1, \ldots, T$ is absolutely regular when

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

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 strong-mixing.

We assume the following assumptions.

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

Assumption A2. $E\{|x_t|^{2+\delta} f(X_{t-1})^{1+\delta} + x_t^{2(2+\delta)} f(X_{t-1})^{4(2+\delta)}\} < \infty$ where $\delta > \frac{\eta}{1-\eta}$.

Assumption A3. Put $u = (u_1, \cdots, u_p)$ and let $K : R^p \to R$ be an $L-$th order

kernel function satisfying $K(-u) = K(u)$,

$$\int u_1^{l_1} \cdots u_p^{l_p} K(u) du \begin{cases} = 1 & \text{if } l_1 + \cdots + l_p = 0 \\ = 0 & \text{if } l_1 + \cdots + l_p < L \\ \neq 0 & \text{for some } l_1 + \cdots + l_p = L, \end{cases}$$

$\int K(u)^2 du + \int \|u\|^L |K(u)| du < \infty$, and $\|u\|^L |K(u)| \to 0$ as $\|u\| \to \infty$. 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. $\{q_i(Z)\}_{i=1}^{\infty}$ forms a basis for s_Z, satisfying $\sup_Z |q_i(Z)| \leq C_q < \infty$ for all i and $\sup_Z |\sum_{i=1}^{k_T} \{q_i(Z) - r_i(X)\}^2| < \zeta_0^2(k_T)$ for $\zeta_0^2(k_T) \to \infty$ as $T \to \infty$.

Assumption A5. $\frac{\zeta_0(k_T)^2 k_T^3}{T} \to 0$ as $T \to \infty$.

Assumption A6. For $X = (x_1, \cdots x_p)$, suppose $f(X)$, $g(X)f(X)$, $r_i(X)f(X)$,

$$m_{1ij}(X) = E\left[u_t^2 Q_i(Z_{t-1}) Q_j(Z_{t-1}) | X_{t-1} = X\right] f(X)^2$$

and

$$m_{2j}(X) = E\left[u_t^2 \{q_j(Z_{t-1}) - r_j(X)\} | X_{t-1} = X\right] f(X)^3$$

are L times differentiable. Furthermore, Writing

$$h^{(l_1,\cdots,l_p)}(X) = \partial^{l_1+\cdots+l_p} h(X) / \partial \xi_1^{l_1} \cdots \partial \xi_p^{l_p},$$

$$(hk)^{(l_1,\cdots,l_p)}(X) = \partial^{l_1+\cdots+l_p} h(X)k(X) / \partial \xi_1^{l_1} \cdots \partial \xi_p^{l_p},$$

for generic functions $h(X)$ and $k(X)$, assume

$$E\{(g(X_{t-1})^2 + u_t^2 + u_t^2 f(X_{t-1})) f^{(l_1,\cdots,l_p)}(X_{t-1})^2\} < \infty,$$

$$E[\{1 + u_t^2 f(X_{t-1})^4\} f(X_{t-1})^2 (gf)^{(l_1,\cdots,l_p)}(X_{t-1})^2] < \infty,$$

$$E\{(f(X_{t-1})^2 + u_t^2)(r_i f)^{(l_1,\cdots,l_p)}(X_{t-1})^2\} < \infty,$$

$$E\{m_{1jl}^{(l_1,\cdots,l_p)}(X_{t-1})^2 + m_{2j}^{(l_1,\cdots,l_p)}(X_{t-1})^2\} < \infty,$$

for all i, j and all $l_1, \cdots, l_p$ satisfying $0 \leq l_1, \cdots, l_p \leq L$, $l_1 + \cdots + l_p = L$.

Assumption A7. Eigenvalues of M are bounded and bounded away from zero uniformly in k_T.

A1 is a standard assumption in nonparametric time series analysis, but we believe it can be relaxed to strong mixing condition in view of recent results such as Gao and King (2004) and Hansen (2008). A2 is a condition used in proving that the second order terms of U-statistics, which appear in the proof, are asymptotically negligible. A3 is standard in asymptotic theory of nonparametric regression. In A4, $\zeta_0^2(k_T) = T$

or T^2 typically depending on the choice of $\{q_i(Z)\}$ as in Newey (1997). We believe A4 and A5 must be stronger than necessary, but these assumptions make the proof of the Theorem much easier. Also we point out that similar assumptions to A4 and A5 were made in Newey (1997) in obtaining asymptotic theory for series regression estimators. A6 guarantees that the projection terms of U-statistics in the proof are asymptotically negligible. Newey (1997) also assumes a similar assumption to A7.

Theorem 1 *If assumptions A1-A7 hold,*

$$\hat{S}_T = \sum_{i=1}^{k_T} w_i \hat{a}_i^2 \xrightarrow{d} \sum_{i=1}^{\infty} w_i \epsilon_i^2 \ \text{as } T \to \infty$$

under H_0, *where* ϵ_i *are i.i.d.* $N(0,1)$.

3 Power of the test under non-local and the $\sqrt{T}$-local alternatives

The above section proposes a test statistic for causality in mean and its null distribution. We discuss about its power properties. Firstly, under non-local alternatives, the DGP can be written as,

$$x_t = g(X_{t-1}) + \kappa(Z_{t-1}) + v_t,$$

for some non-constant function $\kappa(Z_{t-1})$, where $v_t = x_t - E(x_t|z_{t-1})$. Because

$$u_t = x_t - E(x_t|X_{t-1}) = x_t - g(X_{t-1}) = \kappa(Z_{t-1}) + v_t,$$

we have

$$\begin{aligned} E\{u_t f(X_{t-1}) h_i(Z_{t-1})\} &= E[\{\kappa(Z_{t-1}) + v_t\} f(X_{t-1}) h_i(Z_{t-1})] \\ &= E[\kappa(Z_{t-1}) f(X_{t-1}) h_i(Z_{t-1})]. \end{aligned}$$

This value is non-zero at least for some i. It is because if $E[\kappa(Z_{t-1}) f(X_{t-1}) h_i(Z_{t-1})]$ are zero for all i,$\kappa(Z)$ must be identically zero since $\{h_i(Z)\}$ is a complete basis. This is the source of the power of this test in the sense that

$$\frac{1}{\sqrt{T}} a_i \xrightarrow{p} E[\kappa(Z_{t-1}) f(X_{t-1}) h_i(Z_{t-1})] = \kappa_i, \tag{3.1}$$

which results in

$$\hat{S}_T \approx T \sum_{i=1}^{k_T} w_i \kappa_i^2, \tag{3.2}$$

and at least one for κ_i is non-zero. To clarify the power structure of the test, suppose we set w_i in the decreasing order without loss of generality. Suppose further only one of the κ's is non-zero, say 1, for simplicity. If $\kappa_1 = 1$, $\hat{S}_T \approx w_1$, while $\hat{S}_T \approx w_{k_T}$ if $\kappa_{k_T} = 1$. Obviously, the test has a larger power in the former case because we reject the null when $\hat{S}_T$ is large. Therefore we see that if the direction of the departure from the null is closer to the basis functions with smaller indice i, the test will be more likely to reject the null, namely $\hat{S}_T$ possesses more power for the discrepancy from the null toward $h_i(z)$ than $h_j(z)$ for $w_i > w_j$. Becasuse there is no ordering in the basis functions, we can interchange them in fact. Therefore, if we have a prior knowledge on the direction of alternatives, we can make use of this information to order $\hat{a}_i$ to induce a better power property. This is, we believe, a great advantage of this test against ICM tests, that cannot control for which directions they have power. To illustrate this, suppose we know that the alternative is likely to be the null regression 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 expect a large power out of this test.

The following theorem states that $\hat{S}_T$ has a nontrivial power against $\sqrt{T}$-local alternatives.

Theorem 2 *Let $\kappa : R^{p+q} \to R$ be a measurable function in the space $s_X^{\perp}$. Supppose the following local alternatives is correct,*

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

Then, if assumptions A1-A6 hold,

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

under H_{la}, where $\tau_i = E[\kappa(Z_t)f(X_t)h_i(Z_t)]$.

We omit the proof because it is obvious noting that $a_i \xrightarrow{d} N(\tau_i, 1)$ under H_{la}. We now show that flat weight $w_i = 1/\sqrt{k_T}$ for all i, for example, is not permitted in order for the test to have non-trivial power against $\sqrt{T}-$ local alternatives, by providing a couter example. In order for the statistic to be $O_p(1)$ under the null, we need to modify the statistic to

$$S_T' = \frac{1}{\sqrt{k_T}} \sum_{i=1}^{k_T} a_i^2 - \sqrt{k_T} \xrightarrow{d} N(0, 2) \text{ under } H_0.$$

Suppose a $\sqrt{T}-$local alternative of, $\kappa(Z) = h_1(Z)f(X)$, namely,

$$H_{la} : \ x_t = g(X_{t-1}) + \frac{h_1(Z_{t-1})f(X_{t-1})}{\sqrt{T}} + v_t$$

with, for simplicity, conditional homoscedasticity $E(u_t|Z_{t-1}) = const$ is correct. Then, $a_i \xrightarrow{d} N(\tau_i, 1)$, and thus $E(a_i^2) \approx 1 + \tau_i^2$, where $\tau_1 \neq 0$, and $\tau_2 = \tau_3 = \cdots = 0$. In this case, we have,

$$\begin{aligned} S_T' &= \frac{1}{\sqrt{k_T}} \sum_{i=1}^{k_T} E(a_i^2) - \sqrt{k_T} + \frac{1}{\sqrt{k_T}} \sum_{i=1}^{k_T} \{a_i^2 - E(a_i^2)\} \\ &= \frac{1}{\sqrt{k_T}} \sum_{i=1}^{k_T} \tau_i^2 + N(0,2) \\ &= \frac{\tau_1^2}{\sqrt{k_T}} + N(0,2). \end{aligned}$$

Then it obviously does not have power against this simple $\sqrt{T}-$local alternative because the first term in the last expression disappear asymptotically. Therefore, flat weight does not provide a test which has power against $\sqrt{T}-$local alternatives in general.

4 Nonparametric Test for Causality up to the K-th Moment

In view of recent empirical studies in financical econometrics, it appears dependence in second, third or fourth moments looks like an issue of interest. We can provide a tool to examine if there does not exist causal relationship in such a higher order sense using the test proposed above. The main idea is the same as in the previous section, but we need to be careful in understanding the results as explained below. Here we also restrict ourselves to the case when the series of interest follows a stationary nonlinear AR process under the null.

To illustrate the motivation for such high-order causality, let's consider, for example, the following nonlinear dependence between series,

$$x_t = g(x_{t-1}) + \sigma(y_{t-1})\epsilon_t, \tag{4.1}$$

where $\{y_t\}$ is a stationary time series, $g(\cdot)$ and $\sigma(\cdot)$ are unknown functions which satisfy certain conditions for stationarity. Then y_{t-1} obviously does not have information in predicting x_t in the sense that $E(x_t|x_{t-1}, y_{t-1}) = E(x_t|x_{t-1})$, but has information in predicting x_t^2. This type of modeling is becoming increasingly popular in analyzing, for instance, financial data.

In general, we may like to know if y_{t-1} is useful in predicting x_t^K for a given positive integer K. It looks possible to give a definition of "causality in the K-th

moment" similarly to the causality in mean, or we say "y_t causes x_t in the K-th moment" if

$$\mathrm{E}[x_t^K - \mathrm{E}(x_t^K|x_{t-1},\ldots,x_1)]^2 > \mathrm{E}[x_t^K - \mathrm{E}(x_t^K|x_{t-1},\ldots,x_1,y_{t-1},\ldots,y_1)]^2. \quad (4.2)$$

Then, by a similar manipulation to the derivation of (1.4), the null hypothesis of non-causality in K-th moment corresponding to the alternative of (4.2) can be written as

$$H_0 : \mathrm{E}(x_t^K|x_{t-1},\ldots,x_1,y_{t-1},\ldots,y_1) = \mathrm{E}(x_t^K|x_{t-1},\ldots,x_1) \text{ w.p. } 1. \quad (4.3)$$

We should point out, however, that we need to be careful in understanding this definition, because we will almost always conclude that there exists causality in the K-th moment if there exists causality in the M-th moment ($M < K$). We illustrate this when $K = 2$ and $M = 1$. Suppose the DGP is

$$x_t = g(x_{t-1}, y_{t-1}) + \epsilon_t, \quad (4.4)$$

where ϵ_t is independent of x_{t-1}, y_{t-1}. Obviously y_t causes x_t in mean. In terms of the present definition of causality in the K-th moment, we conclude y_t causes x_t in the 2nd moment because

$$\begin{aligned} E(x_t^2|x_{t-1},y_{t-1}) &= E(\{g(x_{t-1},y_{t-1}) + \epsilon_t\}^2|x_{t-1},y_{t-1}) \\ &= g(x_{t-1},y_{t-1})^2 + E(\epsilon_t^2) \\ &\neq E(x_t^2|x_{t-1}) \end{aligned}$$

in general. One way of understanding this definition is that (4.2) states that y_t causes x_t up to the K-th moment rather than in the K-th moment. Formally, we define causaltiy up to K-th moment in the following way.

Definition. We say y_t does not cause x_t up to the K-th moment if

$$\mathrm{E}(x_t^k|x_{t-1},\ldots,x_1,y_{t-1},\ldots,y_1) = \mathrm{E}(x_t^k|x_{t-1},\ldots,x_1)] \text{ w.p. } 1 \text{ for all } k = 1,\cdots,K. \quad (4.5)$$

With this definition, there exists causality up to second moment for both (4.1) and (4.4). When $K = 1$, this reduces to non-causality in mean. We note that it is also possible to define higher-order causality in other ways (see, e.g., Comte and Lieberman (2000) and references therein), where they check if the conditional variance of x_t given lagged x, y depends on lagged y or not in defining "second-order causality", namely,

$$E[\{x_t - E(x_t|Z_{t-1})\}^2|X_{t-1}] = E[\{x_t - E(x_t|Z_{t-1})\}^2|Z_{t-1}], \quad (4.6)$$

holds or not (see definition 3 (i) of Comte and Lieberman (2000)). Their interest is in testing for second-order causality in multivariate GARCH models. Our method

is different from theirs in that we work on a nonparametric framework. We can incorporate their setting of multivariate GARCH to test causality of "second-order" nonparametrically in our framework. Also, it is not clear if the asymptotic null distribution of their test statitics is verified.

In view of the above definition, it is easy to construct test statistic for each k. We simply replace x_t in (2.7), (2.8), (2.9) by x_t^k, and construct statistics $\hat{a}_i^{(k)}$, then $\hat{S}_T^{(k)}$ for each of $k = 1, \cdots K$. Specifically,

$$\hat{S}_T^{(k)} = \sum_{i=1}^{k_T} w_i \hat{a}_i^{(k)2}$$

where

$$\hat{a}_i^{(k)} = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \widehat{u_t^{(k)} f}(X_{t-1}) \hat{h}_i^{(k)}(Z_{t-1}),$$

for

$$\widehat{u_t^{(k)} f}(X_{t-1}) = x_t^k \hat{f}(X_{t-1}) - \frac{1}{Th^p} \sum_{s=p+1}^{T} K(\frac{X_{t-1} - X_{s-1}}{h}) x_s^k,$$

$$\hat{M}^{(k)} = \frac{1}{T} \sum_{t=p+1}^{T} \widehat{u_t^{(k)} f}(X_{t-1})^2 \hat{Q}(Z_{t-1}) \hat{Q}(Z_{t-1})',$$

$$\hat{H}^{(k)}(Z) = (\hat{h}_1^{(k)}(Z), \cdots, \hat{h}_{k_T}^{(k)}(Z))' = \hat{M}^{(k)-1/2} \hat{Q}(Z).$$

It is, however, not easy to combine these statistics into one statistic because they are mutually correlated. One way of testing for the joint null of (4.5) is to estimate the correlation, say by boostrap or other simulation method, and obtain the critical value for a suitably combined statistic. This will be computationally expensive. In order for practical use, we recommend to test causality successively especially when K is small. We implement the test for $k = 1$ first and if the null of non-casality in mean is rejected we stop there as we already know that there is causality in the first moment. If not rejected, we go ahead for the causality in second moment, or $k = 2$. We continue this procedure until the null is rejected at certain stage or k reaches K. Obviously there is a statistical problem that we do not know the total size from this successive method, but at least when K is small, say 2 or 3, we may not need to be too careful about it practically because the empirical size must be conservative, because

$$\begin{aligned} P(\hat{S}_T^{(1)} > c_1 \text{ or } \hat{S}_T^{(2)} > c_2 | H_0) &= P(\hat{S}_T^{(1)} > c_1 | H_0) + P(\hat{S}_T^{(2)} > c_2 | H_0) - P(\hat{S}_T^{(1)} > c_1 \text{ and } \hat{S}_T^{(2)} > c_2 | H_0) \\ &\leq P(\hat{S}_T^{(1)} > c_1 | H_0) + P(\hat{S}_T^{(2)} > c_2 | H_0) \end{aligned}$$

where c_1 and c_2 are critical values, say, of 5% significance level, for $\hat{S}_T^{(1)}$ and $\hat{S}_T^{(2)}$ respectively.

Lastly, we point out that it is possible to skip lower moment causality tests when one knows that there is no causality in lower orders. For instance, it is known that we can hardly improve predition of stock returns in financial application. In such a case, if one may wish to test for causality of second moment, it is possible to go straight to $\hat{S}_T^{(2)}$ test without checking $\hat{S}_T^{(1)}$. If $\hat{S}_T^{(2)}$ test rejects the null, we may conclude there exists causality only in the second moment in the sense that (4.6) does not hold.

For the asymptotic theory of $\hat{S}_T^{(k)}$, we strengthen the moment condition to

Assumption A2'. $E\{|x_t|^{k(2+\delta)} f(X_{t-1})^{1+\delta} + x_t^{2k(2+\delta)} f(X_{t-1})^{4(2+\delta)}\} < \infty$ where $\delta > \frac{\eta}{1-\eta}$.

Theorem 3 *Under assumptions A1, A2',A3-A7,*

$$\hat{S}_T^{(k)} = \sum_{i=1}^{k_T} w_i \hat{a}_i^2 \xrightarrow{d} \sum_{i=1}^{\infty} w_i \epsilon_i^2$$

for $k = 1, \cdots, K$.

The proof is omitted because it is straightforward.

5 Special cases - representation of other omitted variable tests

A typical alternative test is a series of integrated moment condition (ICM) type tests of Bierens (1984, 1990), Bierens and Ploberger (1997) and Chen and Fan (1999) and others. This type of test is constructed as follows. Letting

$$z_T(\xi) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{u}_t \exp(\xi' X_{t-1}),$$

we know under the null

$$T_{ICM} = \int z_T(\xi)^2 d\mu \xrightarrow{d} \sum_{i=1}^{\infty} \tau_i \epsilon_i^2,$$

and it explodes under the alternatives. Here,$\{\tau_i\}$ are eigenvalues of the function $\Gamma(\xi_1, \xi_2) = E[\exp(\xi_1' X_t) \exp(\xi_2' X_t)]$, and ϵ_i are i.i.d. Gaussian random variables.

This also has nontrivial power against $\sqrt{T}-$local alternatives like the propose test. However, it is not clear which direction of alternatives T_{ICM} has more/less power. We believe our test is more convenient than the others because it is much clearer for users how much power the test has to each specific direction of alternatives. More specifically, in the present notation, supposing $w_1 > w_2 > \cdots$, the test (2.11) has the largest power against the alternative of $x_t = g(X_{t-1}) + h_1(Z_{t-1}) + u_t$, while it has the smallest power against the alternative $x_t = g(X_{t-1}) + h_{k_T}(Z_{t-1}) + u_t$.

The test statistics proposed above may look like quite primitive statistics, however, it is a quite general class, in fact, in the sense that this includes important nonparametric test statistics for omitted variable tests as special cases. We firstly show that ICM is a special case. If we set

$$\begin{aligned} w_i &= \xi_i, \\ h_i(Z) &= \xi_i^{-1/2} f(X)^{-1} \int \exp(\xi' Z)\psi_i(\xi)d\xi, \end{aligned}$$

where $\psi_i(\xi)$ is the eigenfunction of $\Gamma(\xi_1, \xi_2)$ corresponding to τ_i, then we can show that

$$S_T = \sum_{i=1}^{k_T} w_i a_i^2 = T_{ICM} + o_p(1).$$

Therefore, T_{ICM} is a special case of the present test for $\{w_i\}$ and $\{h_i(Z)\}$ chosen as above asymptotically. Secondly, the test of omitted variables in nonparametric regression by Hong and White (1995), T_{HW}, can be written as

$$T_{HW} = \sum_{i=1}^{k_T} w_i a_i^2 - \sum_{i=1}^{k_T} w_i + o_p(1) = S_T - \frac{\sqrt{k_T}}{\sigma^2} + o_p(1)$$

where

$$\begin{aligned} w_i &= \frac{1}{\sigma^2\sqrt{k_T}} \\ h(Z) &= (h_1(Z), \cdots, h_{k_T}(Z))' = E\{Q(Z_t)Q(Z_t)\}^{-1} f(X)^{-1} Q(Z), \end{aligned}$$

and $\sigma^2 = E(u_t|X_{t-1})$. The omitted variables test by Fan and Li (1996) also has essentially the same structure as T_{HW}, and thus can be represented similarly. The test by Von Neuman (1941) is also written as

$$T_{VN} = \sum_{i=1}^{T} w_i a_i^2 - \sum_{i=1}^{T} w_i + o_p(1)$$

with

$$\begin{aligned} w_i &= \frac{1}{\sigma^2\sqrt{T}} \\ h(Z) &= (h_1(Z), \cdots, h_{k_T}(Z))' = E\{Q(Z_t)Q(Z_t)\}^{-1} f(X)^{-1} Q(Z), \end{aligned}$$

We can also show that omitted variables tests by Stute (1997) and Whang (2000) are also represented in terms of S_T as above. We also note that T_{VN} and T_{HW} tests do not have power against $\sqrt{T}-$local alternatives because of the flat weight as discussed in Section 3.

We may point out that there is an advantage in rewriting these special case test statistics as the weighted squared sum form in terms of w_i and $h_i(Z)$ in order to see in which direction(s) of alternatives these tests have power. In some cases, we could show that ICM tests have power only against a very limited direction of alternatives, because only first (the largest) weight w_1 is extremely large, while the others, $w_2, w_3, \cdots$ are relatively very small (see Hitomi (2000)). We cannot change this structure of ICM and other tests, though the proposed test can directly change them.

6 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, while 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.

6.1 Testing Causality in Mean

We start with the four simulation settings as below. These can be considered as time series analogue of the experiments carried out in Okui and Hitomi (2002). The results in this section are generated using Ox version 6.1 (see Doornik, 2007).

DGP 0: $x_t = 0.65x_{t-1} + \eta_t, \quad y_t = -0.3y_{t-1} + \epsilon_t$
DGP 1: $x_t = 0.65x_{t-1} + 0.2y_{t-1} + \eta_t, \quad y_t = -0.3y_{t-1} + \epsilon_t$
DGP 2: $x_t = 0.65x_{t-1} + 0.2y_{t-1} + 0.4\sin(-2y_{t-1}) + \eta_t, \quad y_t = -0.3y_{t-1} + \epsilon_t$
DGP 3: $x_t = 0.65x_{t-1} + 0.2y_{t-1}^2 + \eta_t, \quad y_t = -0.3y_{t-1} + \epsilon_t$

Because there is no causal relationship between x_t and y_t in DGP 0, the null hypothesis of Granger non-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 covariate time series y_{t-1} in some nonlinear fashions. Observing x_t and y_t, we 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.

The following is the specification of our test. We used normal kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ for nonparametric regression. Our bandwidth choice is $C \times T^{-0.3}$ where T is the length of time series. C is about 7 except for $T = 100$ case where $C = 9$. For basis $\{q_i(Z)\}$, we choose the following 8 functions (actually $T \times 8$ numbers); $\sin y_{t-1}$, $\cos y_{t-1}$, $\sin y_{t-1}\sin x_{t-1}$, $\sin y_{t-1}\cos x_{t-1}$, $\cos y_{t-1}\sin x_{t-1}$, $\cos y_{t-1}\cos x_{t-1}$, $\sin 2y_{t-1}$, $\cos 2y_{t-1}$. So the number k_T is set to 8 in all the experiments. The weight function is chosen as $w_i = 0.9^i$ to construct the test statistic $\hat{S}_T^{(1)}$. Critical value of asymptotic distribution is calculated by a Monte Carlo simulation with this choice of w_i. The upper 5% critical value is estimated as 14.38.

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

The results are summarized in Table 6.1. In each experiment the number of iteration is fixed to 1000. Each entry shows empirical rejection rate of the null hypothesis of Granger non-causality, hence the two columns under DGP 0 are the empirical size, and others show empirical power. Our test for mean causality is signified as $\hat{S}_T^{(1)}$ following the notation in section 4, and Hidalgo's nonparametric Granger causality test is shortened as HNC. As is seen clearly, the size of $\hat{S}_T^{(1)}$ is quite decent while HNC test seems to over-reject he null in this simulation. Hidalgo's test depends on how many lagged cross-covariance terms, M in his notation, are used to construct the test statistic. As he points out, it is possible to make use of some model selection procedures to choose a plausible M, but our choice of M here is the largest integer which does not exceed $T^{1/4}$. This is the slowest rate within admissible range. Because we would, basically, like to compare the power, not size, we did not try to tune M such that the size distortion of HNC becomes minimum.

What is remarkable is the power of $\hat{S}_T^{(1)}$ in DGP 2 and 3 cases, especially when the sample sizes are fairly large. It performs much better than Hidalgo's test in general as expected, when there exists a nonlinear causality. However, it is worth mentioning that Hidalgo's test shows high power for DGP 1 experiment even in small sample case ($T = 100$). This is because DGP 1 is exactly the situation that Hidalgo's test expects, or linear causality, and we will visit this issue again in the experiments of $\sqrt{T}$-local alternatives.

6.2 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. But in order to confirm the

	DGP 0		DGP 1		DGP 2		DGP 3	
T	$\hat{S}_T^{(1)}$	HNC	$\hat{S}_T^{(1)}$	HNC	$\hat{S}_T^{(1)}$	HNC	$\hat{S}_T^{(1)}$	HNC
100	0.049	0.104	0.092	0.427	0.115	0.221	0.109	0.118
200	0.047	0.108	0.236	0.675	0.335	0.343	0.286	0.135
300	0.049	0.129	0.455	0.859	0.614	0.466	0.565	0.150
400	0.053	0.127	0.614	0.913	0.778	0.539	0.781	0.149
500	0.054	0.130	0.760	0.964	0.911	0.627	0.899	0.158
1000	0.049	0.139	0.989	0.999	1.000	0.871	0.998	0.188

Table 1: Empirical size and power of the proposed test ($\hat{S}_T^{(1)}$) and Hidalgo's nonparametric causality test (HNC).

theory, we carry out a bunch of simulations under local alternatives. For example, in conjunction with DGP 2 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. We refer this specification to **DGP 2L**. **DGP 1L** and **DGP 3L** are defined by dividing the deviation from the null by $\sqrt{T}$ in DGP 1 and DGP 3, though we do not display the equation explicitly. Note that choosing $T = 100$ leads to the specification in previous simulations for non-local alternatives. The figures in the first row in Table 6.2 are, naturally, more or less same as those in the first row of Table 6.1. When T grows from 100 to 1000, the coefficient of y_{t-1} shrinks from 0.2 to 0.02.

As Table 6.2 shows, $\hat{S}_T^{(1)}$ does have power against $\sqrt{T}$-local alternatives. It is striking that $\hat{S}_T^{(1)}$ exhibits quite a good performance for DGP 3L case while the empirical rejection rate of HNC test is as low as its empirical size, which suggests HNC test virtually has no power for DGP 3L case. However, HNC test is superior to our test $\hat{S}_T^{(1)}$ in DGP 1L and DGP 2L. DGP 1L is a case of linear vector autoregressive model, and HNC test is specifically designed for such a linear case. DGP 2L even includes linear causal variable (y_{t-1}) on the RHS of the DGP of x_t. $\hat{S}_T^{(1)}$ captures the causal effect of linear term by the kernel method, which might be a roundabout way.

6.3 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 **DGP 4**.

	DGP 1L		DGP 2L		DGP 3L	
T	$\hat{S}_T^{(1)}$	HNC	$\hat{S}_T^{(1)}$	HNC	$\hat{S}_T^{(1)}$	HNC
100	0.092	0.427	0.115	0.221	0.109	0.118
200	0.133	0.457	0.166	0.238	0.156	0.128
300	0.164	0.500	0.228	0.267	0.224	0.137
400	0.151	0.498	0.213	0.265	0.235	0.138
500	0.166	0.529	0.213	0.289	0.224	0.144
1000	0.157	0.569	0.227	0.326	0.279	0.162

Table 2: Empirical power of the proposed test ($\hat{S}_T^{(1)}$) and Hidalgo's nonparametric causality test (HNC): case of local alternatives.

$$\begin{aligned} \textbf{DGP 4:} \quad \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \\ y_t &= \sigma_t \eta_t, \quad \eta_t \sim \text{NID}(0,1) \\ x_t &= 0.65 x_{t-1} + \sqrt{1 + y_{t-1}^2}\epsilon_t, \quad \epsilon_t \sim \text{NID}(0,1), \end{aligned}$$

where $\omega = 0.05$, $\alpha = 0.1$, $\beta = 0.8$. In other words, y_t is generated by GARCH(1,1) model and it enters into the variance of the innovation of x_t. So x_t does not depend on y_{t-1} in mean but there exists causality in the 2nd moment. Therefore at least theoretically, neither $\hat{S}_T^{(1)}$ (our test in mean) nor Hidalgo's nonparametric Granger causality test will not be able to detect such type of dependency. But our test $\hat{S}_T^{(2)}$, based on the kernel regression of x_t^2 onto y_{t-1}, is expected to detect the causality. The bandwidth in this second stage regression is set to $5.6 \times T^{-0.3}$.

Before we observe the empirical power, we must investigate the size property of the two-stage test employed here. The leftmost column in Table 6.3 shows the empirical size of this successive application of $\hat{S}_T^{(2)}$ after $\hat{S}_T^{(1)}$ analyzed in DGP 0 setting. If both tests take 5% as their significance levels, the overall size should be slightly less than 10% . The figures under DGP 0 in Table 6.3 show that the successive test has good size property for T around 300 or more. The results for $T =$ 100 and 200 turned to be conservative because not $\hat{S}_T^{(1)}$ but $\hat{S}_T^{(2)}$ is very undersized.

The results for DGP 4 are summarized in the three columns under the heading of DGP 4 in Table 6.3. As for HNC, the figures in the table are hardly interpretable. It may uncover even sloppier size or it may be exhibiting the weak power. Quite the contrary, the figures in the column of $\hat{S}_T^{(1)}$ are all close to the size (5%), and the power in the columns of $\hat{S}_T^{(2)}$ grows when the sample size T gets large. Here we note that the figures in $\hat{S}_T^{(2)}$ column are the empirical rejection frequencies divided by the number of cases for which $\hat{S}_T^{(1)}$ did not reject the null. For example in $T = 1000$ case, 72 out of 1000 times $\hat{S}_T^{(1)}$ rejected the null of Granger non-causality in mean.

For the accepted 928 cases we perform the test by $\hat{S}_T^{(2)}$. It has lead to 386 times rejection of non-causality in 2nd moment, namely the 'power' is 0.416.

As a final case, we consider the case where x_t depends on y_{t-1} in 2nd moment as well as in mean.

DGP 5: $x_t = 0.65x_{t-1} + 0.2y_{t-1} + \sqrt{1 + y_{t-1}^2}\eta_t, \quad y_t = -0.3y_{t-1} + \epsilon_t$

This is a challenging situation for the test we propose. We expect $\hat{S}_T^{(1)}$ to detect causality from $0.2y_{t-1}$ term, but then it will stop the procedure and we do not proceed to second stage with $\hat{S}_T^{(2)}$. Hence just the same as in DGP 4, we apply $\hat{S}_T^{(2)}$ when $\hat{S}_T^{(1)}$ fails to detect the causality in mean. The three columns under the heading of DGP 5 in Table 6.3 show that $\hat{S}_T^{(1)}$ suffers at small sample stages, but finally has moderate power at $T = 1000$. When the sample size is large, $\hat{S}_T^{(2)}$ manages to detect the causality in the 2nd moment even if $\hat{S}_T^{(1)}$ failed to detect the causality in mean. As for HNC, again it shows considerably good performance because the true DGP includes linear causality from y_{t-1}. Though the overall size control could be difficult, combining HNC with $\hat{S}_T^{(2)}$ might help in some practical situation.

	DGP 0	DGP 4			DGP 5		
T	$\hat{S}_T^{(1)}$ & $\hat{S}_T^{(2)}$	$\hat{S}_T^{(1)}$	$\hat{S}_T^{(2)}$	HNC	$\hat{S}_T^{(1)}$	$\hat{S}_T^{(2)}$	HNC
100	0.052	0.075	0.003	0.166	0.178	0.023	0.435
200	0.067	0.046	0.047	0.185	0.271	0.167	0.646
300	0.084	0.059	0.072	0.208	0.388	0.265	0.797
400	0.075	0.066	0.128	0.211	0.466	0.434	0.867
500	0.088	0.080	0.162	0.217	0.605	0.597	0.918
1000	0.099	0.072	0.416	0.257	0.920	1.000	0.995

Table 3: Empirical size and power of successive tests when causality in 2nd moment exists (DGP4) and when causality both in mean and in variance exists (DGP 5).

6.4 Application to Price-Volume Causality

As an application to real data, we investigate the causality between stock price and traded volume. Price-Volume charting has been very popular among technical trading, and it is often said that the change in volume precedes the price movement. Ideally, the rise of price will be followed by the increase in the volume, while the fall of price often observed after the shrink in the traded volume. It is sometimes called 'counter clockwise' relationship between volume and price.

We use weekly price and volume data of Nikkei 225 futures. The sample period starts at 1st week of September 1988 and ends in the 4th week of October 2007, of which sample size is $T = 1000$. To ensure the stationarity of the data, both price and volume are transformed to the changes in percentage. We call these variables in change dp_t and dv_t respectively.

Our primal concern is to investigate if there is any causality from past volume change to the present price change. Hence we give dv_{t-1} the role of y_{t-1}, and perform the nonparametric causality test in mean using $\hat{S}_T^{(1)}$. The result is $\hat{S}_T^{(1)} = 11.54 < 14.38$, which leads to accepting the null of non-causality in mean. Then we proceed to the second stage, and regress dp_t^2 nonparametrically on dv_{t-1}. The resulting test statistic is $\hat{S}_T^{(2)} = 20.81 > 14.38$, which implies the rejection of non-causality in 2nd moment.

To make sure, we check the reverse causal relationship, exchanging the role of dp_t and dv_t. The results are $\hat{S}_T^{(1)} = 9.95 < 14.38$ and $\hat{S}_T^{(2)} = 9.30 < 14.38$, hence both causality in mean and in 2nd moment are not detected. Combining these results, we may conclude that the volatility of Nikkei 225 futures is caused by its past volume change, but there seems to be no evidence for the reverse order causality.

7 Concluding Remarks

Granger causality tests proposed previously do not detect certain 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 know only that the error terms are uncorrelated with the series of interest, rather than independent. Observing this fact, we proposed a nonparametric testing procedure for Granger-type causality in the case of any form of nonlinear dependence.

The test can be viewed as a test for omitted variables in time series regression. We show that the test has nontrivial power against $\sqrt{T}$-local alternatives like ICM. But it has an advantage over other omitted variables test in that the power properties are clear, and thus we can incorporate a priori information, if ever, in forming a test. Furthermore, it is possible show that those previously proposed omitted variables tests can be represented in terms of the present test asymptotically.

The Monte Carlo study shows that the test performs well with decent empirical size, in general, and good power property for both causality in mean and causality up to second moment. Compared with the Hidalgo's test, $\hat{S}_T$ does slightly worse when the series exhibit linear dependence in mean in fact, but in the case of nonlinear dependence, especially nonlinear dependence in conditional variance, it obviously outperforms the Hidalgo's test. We also show an empirical example for finanaicial data.

Appendix

A Proofs of Theorem

(*Proof of Theorem 1*) Denote

$$\boldsymbol{a} = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) \boldsymbol{h}(Z_{t-1}) = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) M^{-\frac{1}{2}} Q(Z_{t-1}),$$

$$\hat{\boldsymbol{a}} = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \widehat{u_t f}(X_{t-1}) \hat{\boldsymbol{h}}(Z_{t-1}) = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \widehat{u_t f}(X_{t-1}) \hat{M}^{-\frac{1}{2}} \hat{Q}(Z_{t-1}),$$

and

$$W = diag(w_1, \cdots, w_{k_T}).$$

We would like to show that

$$\boldsymbol{a}' W \boldsymbol{a} - \sum_{i=1}^{\infty} w_i \epsilon_i^2 \xrightarrow{p} 0 \tag{A.1}$$

and

$$\hat{\boldsymbol{a}}' W \hat{\boldsymbol{a}} - \boldsymbol{a}' W \boldsymbol{a} \xrightarrow{p} 0 \tag{A.2}$$

as $T \to \infty$. We first prove (A.1). For $\forall \epsilon > 0$ and $\forall \delta > 0$, we can choose m such that $\sum_{i=m+1}^{\infty} w_i < \delta\epsilon/2$ because $\{w_i\}$ is a summable sequence. As m is a finite number independent of T, $m < k_T < T$ holds for sufficiently large T. Then we have

$$\begin{aligned} P\left(\left| \sum_{i=m+1}^{k_T} w_i a_i^2 - \sum_{i=m+1}^{\infty} w_i \epsilon_i^2 \right| > \delta \right) &\leq \frac{E\left|\sum_{i=m+1}^{k_T} w_i a_i^2\right| + E\left|\sum_{i=m+1}^{\infty} w_i \epsilon_i^2\right|}{\delta} \\ &= \frac{\sum_{i=m+1}^{k_T} w_i + \sum_{i=m+1}^{\infty} w_i}{\delta} \\ &< \frac{2\sum_{i=m+1}^{\infty} w_i}{\delta} < \epsilon \end{aligned}$$

where the first inequality is due to Markov inequality, the equality uses $E(a_i^2) = E(\epsilon_i^2) = 1$. We look at the remaining first m elements. We only need to show

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \xrightarrow{d} N(0, I_m). \tag{A.3}$$

Let $\lambda = (\lambda_1, \cdots, \lambda_m)'$ be any fixed $m-$vector, then

$$\sum_{i=1}^{m} \lambda_i a_i = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) \lambda_i h_i(Z_{t-1}) = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} v_t$$

where $E(v_t|Z_{t-1}) = 0$, $V(v_t) = \sum_{i=1}^{m} \lambda_i^2$. Appealing to the CLT for stationary Martingale differences (see, e.g. Theorem 7.11 of Bierens (2004)) with Cramér-Wold device, we have (A.3), which completes the proof for (A.1).

We prove (A.2) next. Noting that

$$\hat{\boldsymbol{a}}' W \hat{\boldsymbol{a}} - \boldsymbol{a}' W \boldsymbol{a} = (\hat{\boldsymbol{a}} - \boldsymbol{a})' W (\hat{\boldsymbol{a}} - \boldsymbol{a}) + (\hat{\boldsymbol{a}} - \boldsymbol{a})' W \boldsymbol{a} + \boldsymbol{a}' W (\hat{\boldsymbol{a}} - \boldsymbol{a})$$

and

$$|(\hat{\boldsymbol{a}} - \boldsymbol{a})' W \boldsymbol{a}|^2 = |\boldsymbol{a}' W (\hat{\boldsymbol{a}} - \boldsymbol{a})|^2 \leq ||(\hat{\boldsymbol{a}} - \boldsymbol{a})' W^{\frac{1}{2}}||^2 \, ||W^{\frac{1}{2}} \boldsymbol{a}||^2 \leq (\hat{\boldsymbol{a}} - \boldsymbol{a})' W (\hat{\boldsymbol{a}} - \boldsymbol{a}) \cdot \boldsymbol{a}' W \boldsymbol{a},$$

it suffices to show

$$(\hat{\boldsymbol{a}} - \boldsymbol{a})' W (\hat{\boldsymbol{a}} - \boldsymbol{a}) \xrightarrow{p} 0 \tag{A.4}$$

as $\boldsymbol{a}' W \boldsymbol{a} = O_p(1)$ by (A.1). Write further

$$\begin{aligned}
\hat{\boldsymbol{a}} - \boldsymbol{a} &= \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \{\widehat{u_t f}(X_{t-1}) \hat{M}^{-\frac{1}{2}} \hat{Q}(Z_{t-1}) - u_t f(X_{t-1}) M^{-\frac{1}{2}} Q(Z_{t-1})\} \\
&= \frac{1}{\sqrt{T}} \sum_{t=r}^{T} \{\widehat{u_t f}(X_{t-1}) - u_t f(X_{t-1})\} M^{-\frac{1}{2}} Q(Z_{t-1}) \\
&\quad + (\hat{M}^{-\frac{1}{2}} - M^{-\frac{1}{2}}) \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) Q(Z_{t-1}) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=r}^{T} u_t f(X_{t-1}) M^{-\frac{1}{2}} \{\hat{Q}(Z_{t-1}) - Q(Z_{t-1})\} \\
&\quad + \text{smaller order terms} \\
&= A_{1T} + A_{2T} + A_{3T} + \text{smaller order terms}.
\end{aligned}$$

Similarly as above, we immediately know

$$A_{1T}' W A_{1T} \xrightarrow{p} 0 \tag{A.5}$$

$$A_{2T}' W A_{2T} \xrightarrow{p} 0 \tag{A.6}$$

and

$$A_{3T}' W A_{3T} \xrightarrow{p} 0 \tag{A.7}$$

imply (A.4). Therefore, (A.5), (A.6) and (A.7) suffice to show (A.2). We start with (A.6). Because

$$A_{2T} = (\hat{M}^{-\frac{1}{2}} - M^{-\frac{1}{2}})\frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_t f(X_{t-1})Q(Z_{t-1}),$$

we have

$$\begin{aligned} A_{2T}'WA_{2T} &\leq ||A_{2T}'W^{\frac{1}{2}}||^2 \leq ||A_{2T}||^2||W^{\frac{1}{2}}||^2 \\ &\leq ||\hat{M}^{-\frac{1}{2}} - M^{-\frac{1}{2}}||^2||\frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_t f(X_{t-1})Q(Z_{t-1})||^2||W^{\frac{1}{2}}||^2. \end{aligned}$$

As $\{w_i\}$ are summable, $||W^{\frac{1}{2}}||^2 = \sum_{i=1}^{k_T} w_i < \infty$. By Lemmas 5, 6, and Assumption A7,

$$||\hat{M}^{-\frac{1}{2}} - M^{-\frac{1}{2}}||^2 \leq \frac{k_T||\hat{M} - M||^2}{\lambda_{min}^3(M)} = O_p(\frac{\zeta_0(k_T)^2 k_T^2}{T}).$$

Also,

$$\begin{aligned} E||\frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_t f(X_{t-1})\boldsymbol{Q}(Z_{t-1})||^2 &= \sum_{i=1}^{k_T} E\{\frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_t f(X_{t-1})Q_i(Z_{t-1})\}^2 \\ &= \sum_{i=1}^{k_T} E\{u_t^2 f(X_{t-1})^2 Q_i(Z_{t-1})^2\} \\ &\leq 4C_q^2 k_T E\{u_t^2 f(X_{t-1})^3\} \\ &\leq Ck_T. \end{aligned}$$

Combining them, we have

$$A_{2T}'WA_{2T} = O_p(\frac{\zeta_0(k_T)^2 k_T^3}{T}),$$

and thus, (A.6) holds by Assumption A5.

Because

$$\begin{aligned} E(A_{1T}'WA_{1T}) &= \sum_{i=1}^{k_T} w_i E\left[\frac{1}{\sqrt{T}}\sum_{t=r}^{T}\{\widehat{u_t f}(X_{t-1}) - u_t f(X_{t-1})\}h_i(Z_{t-1})\right]^2 \\ &\leq C \sup_{1\leq i\leq k_T} E\left[\frac{1}{\sqrt{T}}\sum_{t=r}^{T}\{\widehat{u_t f}(X_{t-1}) - u_t f(X_{t-1})\}h_i(Z_{t-1})\right]^2, \end{aligned}$$

it is sufficient to show, by Markov inequality,

$$E\left[\frac{1}{\sqrt{T}}\sum_{t=r}^{T}\{\widehat{u_t f}(X_{t-1}) - u_t f(X_{t-1})\}h_i(Z_{t-1})\right]^2 \to 0 \tag{A.8}$$

as $T \to \infty$ uniformly in i in order to prove (A.5). We rewrite the quantity in the square brackets into a U statistic form as follows.

$$\frac{1}{\sqrt{T}}\sum_{t=r}^{T}\{\widehat{u_t f}(X_{t-1}) - u_t f(X_{t-1})\}h_i(Z_{t-1}) = \sqrt{T}\begin{pmatrix} T \\ 2 \end{pmatrix}^{-1}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T} U_i(W_t, W_s),$$

where $W_t = (x_t, Z_{t-1})$ and

$$\begin{aligned} U_i(W_t, W_s) &= \frac{1}{2}\left[\{(x_t - x_s)\frac{1}{h^p}K(\frac{X_{t-1} - X_{s-1}}{h}) - u_t f(X_{t-1})\}h_i(Z_{t-1})\right. \\ &\quad \left. +\{(x_s - x_t)\frac{1}{h^p}K(\frac{X_{t-1} - X_{s-1}}{h}) - u_s f(X_{s-1})\}h_i(Z_{s-1})\right]. \end{aligned}$$

We consider the projection which will be used later. For $W = (x, Z)$ and $u = x - g(X)$,

$$\begin{aligned} u_{1i}(W) &= E\{U_i(W_t, W)\} \\ &= -\frac{1}{2}h_i(Z)\int g(X + hu)f(X + hu)K(u)du \\ &\quad +\frac{1}{2}xh_i(Z)\int f(X + hu)K(u)du - \frac{1}{2}uf(X)h_i(Z). \end{aligned}$$

In order for the Hoefdding decomposition, put $\phi_i(W_t, W_s) = U_i(W_t, W_s) - u_{1i}(W_t) - u_{1i}(W_s)$ and write

$$\begin{aligned} &\sqrt{T}\begin{pmatrix} T \\ 2 \end{pmatrix}^{-1}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T} U_i(W_t, W_s) \\ &= \frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_{1i}(W_t) + \sqrt{T}\begin{pmatrix} T \\ 2 \end{pmatrix}^{-1}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_i(W_t, W_s) \end{aligned}$$

Then we want to show, in order to prove (A.8), that

$$V\{\frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_{1i}(W_t)\} \to 0, \tag{A.9}$$

and

$$E\left\{\sqrt{T}\begin{pmatrix} T \\ 2 \end{pmatrix}^{-1}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_i(W_t,W_s)\right\}^2 \to 0 \tag{A.10}$$

uniformaly in i. We first prove (A.16). We immediately know that

$$E\{u_{1i}(W_t)|W_s\} = 0 \tag{A.11}$$

for $t > s$ because $E(u_t|Z_{t-1}) = 0$, $E(x_t|Z_{t-1}) = g(X_{t-1})$ under H_0, and $E\{h_i(Z_{t-1})|X_{t-1}\} = 0$. Also, by Assumptions A3 and A6, we can show

$$u_{1i}(W) = h^L k_i(W) + o(h^L)$$

with

$$k_{1i}(W) = \frac{h_i(Z)}{2L!}\sum_{\substack{0 \le l_1,\cdots,l_p \le L \\ l_1+\cdots+l_p = L}}\cdots\sum\left\{\int\prod_{i=1}^{p}u_i^{l_i}K(u)du\right\} \times\left\{xf^{(l_1,\cdots,l_p)}(X) + (gf)^{(l_1,\cdots,l_p)}(X)\right\}$$

where $f^{(l_1,\cdots,l_p)}(X) = \partial f(X)/\partial x_1^{l_1}\cdots\partial x_p^{l_p}$. Based on the above, it is straightforward to prove (A.16) because $E\{u_{1i}(W_t)\} = 0$, $V\{u_{1i}(W_t)\} = O(h^{2L})$ by Assumptions A4, A6 and (A.11).

We follow the same track as Yoshihara (1976) to prove (A.17). The squared sum $\left\{\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_i(W_t,W_s)\right\}^2$ involves a double summation term

$$\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_i(W_t,W_s)^2,$$

terms with triple summation such as

$$\sum_{t=r}^{T-2}\sum_{s=t+1}^{T-1}\sum_{u=s+1}^{T}\phi_i(W_t,W_s)\phi_i(W_t,W_u)$$

and those with quadruple summation of the form

$$\sum_{t=r}^{T-3}\sum_{s=t+1}^{T-2}\sum_{u=s+1}^{T-1}\sum_{v=u+1}^{T}\phi_i(W_t,W_s)\phi_i(W_u,W_v).$$

We evaluate the expectation of these terms individually. We follow the same track as Yoshihara (1976) to obtain the bounds for the expectations of these quantities. Define $F_{t_1,\cdots,t_m}(w_1,\cdots,w_m)$ be the joint distribution function of $W_{t_1},\cdots,W_{t_m}$. Firstly, under Assumption A2, A4, we have

$$\begin{aligned}
& \int |\phi_i(w_1,w_2)|^{2+\delta} dF(w_1)dF(w_2) \\
\leq & \int |U_i(w_1,w_2)|^{2+\delta} dF(w_1)dF(w_2) + C \qquad \text{(A.12)} \\
\leq & C\int |x_1-x_2|^{2+\delta} \frac{1}{h^{p(2+\delta)}} |K(\frac{X_1-X_2}{h})|^{2+\delta} |h_i(Z_1)|^{2+\delta} \\
& +|\{x_1-g(X_1)\}f(X_1)h_i(Z_1)|^{2+\delta} dF(w_1)dF(w_2) + C \qquad \text{(A.13)} \\
\leq & C\{h^{-p(1+\delta)}+1\} = C_{1h}. \qquad \text{(A.14)}
\end{aligned}$$

The first inequality holds because $E|u_{1i}(W_t)|^{2+\delta}$ is of smaller order than $E|U_i(W_t,W_s)|^{2+\delta}$. The last inequality is valid uniformly in i because $|h_i(Z)| \leq 2C_q f(X)$ due to Assumption A4. Lemma 4 yields

$$|\int \phi(w_1,w_2)^2 dF_{t,s}(w_1,w_2) - \int \phi_i(w_1,w_2)^2 dF(w_1)dF(w_2)| \leq 4C_{1h}^{\frac{2}{2+\delta}} \beta(s-t)^{\frac{\delta}{2+\delta}}$$

but (A.14) holds also for $\delta = 0$, and thus $\int \phi_i(w_1,w_2)^2 dF(w_1)dF(w_2) = O(h^{-p})$. Combining them, we have

$$\begin{aligned}
E\{\phi_i(W_t,W_s)^2\} &= \int \phi_i(w_1,w_2)^2 dF_{t,s}(w_1,w_2) \\
&\leq C\{h^{-p} + h^{-\frac{2p(1+\delta)}{2+\delta}} \beta(s-t)^{\frac{\delta}{2+\delta}}\}.
\end{aligned}$$

Because

$$\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \beta(s-t)^{\frac{\delta}{2+\delta}} = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T} (s-t)^{-\frac{(2+\eta)\delta}{\eta(2+\delta)}} = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T} (s-t)^{-1-\gamma} \leq CT,$$

for $\gamma = 2(\delta-\eta)/\{\eta(2+\delta)\} > 0$, we have

$$E\left\{T\begin{pmatrix} T \\ 2 \end{pmatrix}^{-2} \sum_{t=r}^{T-1}\sum_{s=t+1}^{T} \phi_i(W_t,W_s)^2\right\} \leq C(\frac{1}{Th^p} + \frac{1}{T^2 h^{\frac{2p(1+\delta)}{2+\delta}}}) = C(\frac{1}{Th^p} + \frac{h^{\frac{2p}{2+\delta}}}{T^2 h^{2p}})$$

which goes to zero because of Assumption A3.

We consider

$$T\begin{pmatrix} T \\ 2 \end{pmatrix}^{-2} \sum_{t=r}^{T-2}\sum_{s=t+1}^{T-1}\sum_{u=s+1}^{T} \phi_i(W_t,W_s)\phi_i(W_t,W_u)$$

next. We have, similarly to the computation of (A.14),

$$\begin{aligned} & \int |\phi_i(w_1,w_2)\phi_i(w_1,w_3)|^{1+\frac{\delta}{2}} dF_{t,s}(w_1,w_2)dF_u(w_3) \\ \leq & \int |\phi_i(w_1,w_2)|^{1+\frac{\delta}{2}} \{\int |\phi_i(w_1,w_3)|^{1+\frac{\delta}{2}} dF_u(w_3)\} dF_{t,s}(w_1,w_2) \\ \leq & C(h^{-p\delta}+1) = C_{2h}. \end{aligned}$$

Then, by Lemma 4, we have

$$|\int \phi_i(w_1,w_2)\phi_i(w_1,w_3)dF_{t,s,u}(w_1,w_2,w_3) - \int \phi_i(w_1,w_2)\phi_i(w_1,w_3)dF_{t,s}(w_1,w_2)dF(w_3)|$$
$$\leq 4C_{2h}^{\frac{2}{2+\delta}} \beta(s-t)^{\frac{\delta}{2+\delta}}.$$

Noting $\int \phi_i(w_1,w_3)dF(w_3) = 0$ by construction, we have

$$|E\{\phi_i(W_t,W_s)\phi_i(W_t,W_u)\}| \leq 4C_{2h}^{\frac{2}{2+\delta}} \beta(s-t)^{\frac{\delta}{2+\delta}} \leq Ch^{-\frac{2p\delta}{2+\delta}}(s-t)^{-1-\gamma}$$

which yields

$$|E\left\{T\begin{pmatrix} T \\ 2 \end{pmatrix}^{-2} \sum_{t=r}^{T-2} \sum_{s=t+1}^{T-1} \sum_{u=s+1}^{T} \phi_i(W_t,W_s)\phi_i(W_t,W_u)\right\}| \leq C\frac{h^{\frac{2-\delta}{2+\delta}}}{Th^p}.$$

We can pick $\delta < 2$ as long as $E|x_t|^{2+\delta}$ exists. Thus, the expectation of the triple sum decays to zero.

For the quadruple summation, we consider the cases (i) $t < s < u < v$ and $s-t > v-u$ and (ii)$t < s < u < v$ and $s-t \leq v-u$ as in Yoshihara (1976). Because both cases can be handled similarly, we consider only the first case (i). Note that

$$\begin{aligned} & \int |\phi_i(w_1,w_2)\phi_i(w_3,w_4)|^{1+\frac{\delta}{2}} dF_{t,s,u}(w_1,w_2,w_3)dF(w_4) \\ \leq & \int |\phi_i(w_1,w_2)|^{1+\frac{\delta}{2}} \{\int |\phi_i(w_3,w_4)|^{1+\frac{\delta}{2}} dF(w_4)\} dF_{t,s,u}(w_1,w_2,w_3) \\ \leq & C(h^{-p\delta}+1) = C_{3h}. \end{aligned}$$

Then

$$\begin{aligned} & |\int \phi_i(w_1,w_2)\phi_i(w_3,w_4)dF_{t,s,u,v}(w_1,w_2,w_3,w_4) \\ & - \int \phi_i(w_1,w_2)\phi_i(w_3,w_4)dF_{t,s,u}(w_1,w_2,w_3)dF(w_4)| \leq 4C_{3h}^{\frac{2}{2+\delta}} \beta(v-u)^{\frac{\delta}{2+\delta}}. \end{aligned}$$

Noting $\int \phi_i(w_3, w_4)dF(w_4) = 0$, we have

$$|E\{\phi_i(W_t, W_s)\phi_i(W_u, W_v)\}| \leq 4C_{3h}^{\frac{2}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}},$$

and thus

$$|E\left\{T\left(\begin{array}{c} T \\ 2 \end{array}\right)^{-2}\sum\sum_{\substack{t<s<u<v \\ s-t<v-u}}\sum\sum\phi_i(W_t, W_s)\phi_i(W_u, W_v)\right\}| \leq CT^{-\gamma}h^{-\frac{2p\delta}{2+\delta}} = \frac{Ch^{\frac{2p(\delta-\eta-\eta\delta)}{\eta(2+\delta)}}}{(Th^p)^{\frac{2(\delta-\eta)}{\eta(2+\delta)}}}.$$

By assumption ($\eta < 1$ & $\delta > \eta/1-\eta$) and ($Th^p \to \infty$), this decays to zero as $T \to \infty$. Combining these results, we have (A.17).

Now to prove (A.7), similarly to the proof for (A.5), it suffices to show

$$E(A_{3Ti}^2) \to 0 \tag{A.15}$$

uniformly in i where $A_{3Ti}$is the $i-$th element of A_{3T}, namely,

$$A_{3Ti} = \frac{1}{\sqrt{T}}\sum_{t=r}^{T} u_t f(X_{t-1})m_i'\{\hat{\boldsymbol{Q}}(Z_{t-1}) - \boldsymbol{Q}(Z_{t-1})\}$$

where $m_i$indicates the $i-$th row of $M^{-\frac{1}{2}}$. It is easily seen that $||m_i||^2 \leq 1/\lambda_{min} < C$ by Assumption A7, where λ_{min} is the minimum eigenvalue of M. We write A_{3Ti} in the U-statistic form,

$$A_{3Ti} = \sqrt{T}\left(\begin{array}{c} T \\ 2 \end{array}\right)^{-1}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T} V_i(W_t, W_s),$$

where

$$\begin{aligned} V_i(W_t, W_s) &= \frac{1}{2}\Big[\{u_t f(X_{t-1}) - u_s f(X_{s-1})\}m_i'\{Q(Z_{t-1}) - Q(Z_{s-1})\}\frac{1}{h^p}K(\frac{X_{t-1} - X_{s-1}}{h}) \\ &\quad -\{u_t f(X_{t-1})h_i(Z_{t-1}) + u_s f(X_{s-1})h_i(Z_{s-1})\}\Big]. \end{aligned}$$

The projection is, for $W = (x, Z)$, $u = x - g(X)$, $q(Z) = Q(Z)/f(X)$ and $r(X) = E\{q(Z_{t-1})|X_{t-1} = X\}$,

$$\begin{aligned} v_{1i}(W) &= E\{V_i(W_t, W)\} \\ &= \frac{1}{2}uf(X)m_i'\int\{q(Z) - r(X + hu)\}f(X + hu)K(u)du \\ &\quad +\frac{1}{2}uf(X)^2m_i'\{q(Z) - r(X)\}f(X). \end{aligned}$$

In order for the Hoefdding decomposition, put $\psi_i(W_t, W_s) = V_i(W_t, W_s) - v_{1i}(W_t) - v_{1i}(W_s)$ and write

$$\sqrt{T}\left(\begin{array}{c} T \\ 2 \end{array}\right)^{-1} \sum_{t=r}^{T-1} \sum_{s=t+1}^{T} V_i(W_t, W_s) = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} v_{1i}(W_t) + \sqrt{T}\left(\begin{array}{c} T \\ 2 \end{array}\right)^{-1} \sum_{t=r}^{T-1} \sum_{s=t+1}^{T} \psi_i(W_t, W_s)$$

Then we want to show, in order to prove (A.15), that

$$V\{\frac{1}{\sqrt{T}} \sum_{t=r}^{T} v_{1i}(W_t)\} \to 0, \tag{A.16}$$

and

$$E\left\{\sqrt{T}\left(\begin{array}{c} T \\ 2 \end{array}\right)^{-1} \sum_{t=r}^{T-1} \sum_{s=t+1}^{T} \psi_i(W_t, W_s)\right\}^2 \to 0 \tag{A.17}$$

uniformly in i. It is straightforward to prove that $\{v_{1i}(W_t)\}$ is a martingale difference sequence, we can further show

$$v_{1i}(W) = h^L k_{3i}(W) + o(h^L)$$

with

$$k_{3i}(W) = \frac{\{x - g(X)\} f(X)}{2L!} \sum_{\substack{0 \le l_1, \cdots, l_p \le L \\ l_1 + \cdots + l_p = L}} \cdots \sum \left\{\int \prod_{i=1}^{p} u_i^{l_i} K(u) du\right\}$$
$$\times \left\{m_i' q(Z) f^{(l_1, \cdots, l_p)}(X) - (m_i' r f)^{(l_1, \cdots, l_p)}(X)\right\}.$$

Therefore, (A.16) is true. For (A.17), we basically need to look at $V_i(W_t, W_s)$ as $v_{1i}(W_t)$ are of much smaller order. We omit the proof for (A.17) because it can be treated in exactly the same manner as (A.10).

B Lemmas

The following lemma by Yoshihara (1976) is useful to prove Theorem 1. The proof is omitted.

Lemma 4 *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 $\delta > 0$,*

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

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

The next two lemmas are used to prove the theorem.

Lemma 5 *Under A1-A6,*

$$E\left\|\hat{M}-M\right\|^2 \leq C_1\frac{\zeta_0^4(k_T)}{T}+C_2\frac{k_T^2}{T}+C_3\frac{\zeta_0^2(k_T)k_T}{T}.$$

Proof. Recall $r_j(X_{t-1}) = E[q_j(Z_{t-1})|X_{t-1}]$ for $j = 1, \cdots, k_T$, and $u_t = x_t - g(X_{t-1})$ where $g(X_{t-1}) = E[x_t|X_{t-1}]$, and let $\tilde{q}_j(Z_{t-1}) = q_j(Z_{t-1}) - r_j(X_{t-1})$. The (j, l) element of $\hat{M} - M$ is,

$$A_{j,l} = \frac{1}{T}\sum \hat{u}_t^2\hat{f}^4(X_{t-1})\hat{\tilde{q}}_j(Z_{t-1})\hat{\tilde{q}}_l(Z_{t-1}) - E\left[u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\right]$$

where

$$\hat{f}(X_{t-1}) = \frac{1}{Th^p}\sum_{s=2}^{T} K\left(\frac{X_{s-1}-X_{t-1}}{h}\right),$$

$$\hat{\tilde{q}}_j(Z_{t-1}) = q_j(Z_{t-1}) - \frac{1}{Th^p}\sum_{s=2}^{T} q_j(Z_{s-1})K\left(\frac{X_{s-1}-X_{t-1}}{h}\right)/\hat{f}(X_{t-1})$$

and

$$\hat{u}_t = x_t - \frac{1}{Th^p}\sum_{s=2}^{T} x_s K\left(\frac{X_{s-1}-X_{t-1}}{h}\right)/\hat{f}(X_{t-1}).$$

Decompose $A_{j,l}$ as the followings,

$$
\begin{aligned}
& A_{j,l} \\
= \; & \frac{2}{T}\sum_{t=2}^{T} x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\{\hat{f}(X_{t-1}) - f(X_{t-1})\} \\
& -\frac{2}{T}\sum_{t=1}^{T} u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\{\hat{g}(X_{t-1})\hat{f}(X_{t-1}) - g(X_{t-1})f(X_{t-1})\} \\
& +\frac{1}{T}\sum_{t=1}^{T} u_t^2 f^3(X_{t-1})\tilde{q}_l(Z_{t-1})q_j(Z_{t-1})\{\hat{f}(X_{t-1}) - f(X_{t-1})\} \\
& -\frac{1}{T}\sum_{t=1}^{T} u_t^2 f^3(X_{t-1})\tilde{q}_l(Z_{t-1})\{\hat{r}_j(X_{t-1})\hat{f}(X_{t-1}) - r_j(X_{t-1})f(X_{t-1})\} \\
& +\frac{1}{T}\sum_{t=1}^{T} u_t^2 f^3(X_{t-1})\tilde{q}_j(Z_{t-1})q_l(Z_{t-1})\{\hat{f}(X_{t-1}) - f(X_{t-1})\} \\
& -\frac{1}{T}\sum_{t=1}^{T} u_t^2 f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\{\hat{r}_l(X_{t-1})\hat{f}(X_{t-1}) - r_l(X_{t-1})f(X_{t-1})\} \\
& +\frac{1}{T}\sum_{t=1}^{T} u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1}) - E\left[u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\right] \\
& +(\text{smaller order terms}) \\
= \; & A_{1j,l} + A_{2j,l} + A_{3j,l} + A_{4j,l} + A_{5j,l} + A_{6j,l} + A_{7j,l} + (\text{smaller order terms}).
\end{aligned}
$$

Let W_t be $W_t = (x_t, Z_{t-1})$, and define $U_1(W_t, W_s)$ and its projection $u_1^c(w)$ as follows,

$$
\begin{aligned}
U_1(W_t, W_s) \;=\; & \frac{1}{2}\left\{ x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\frac{1}{Th^p}K\left(\frac{X_{s-1} - X_{t-1}}{h}\right)\right. \\
& \left. +x_s u_s f^3(X_{s-1})\tilde{q}_j(Z_{s-1})\tilde{q}_l(Z_{s-1})\frac{1}{Th^p}K\left(\frac{X_{s-1} - X_{t-1}}{h}\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
u_1^c(W_t) \;=\; & E[U_1(W, W_s)]|_{W=W_t} \\
=\; & \frac{1}{2}\left\{ x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})f(X_{t-1})\right. \\
& \left. +E[x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]f(X_{t-1}) + O_p(h^L)\right\}.
\end{aligned}
$$

Note that

$$
E[x_s u_s f^3(X_{s-1})\tilde{q}_j(Z_{s-1})\tilde{q}_l(Z_{s-1})|X_{s-1} = X] = E[x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1} = X]
$$

by the stationality of W_t. In order for the Hoefdding decomposition, put $\phi_1(W_t, W_s) = U_1(W_t, W_s) - u_1^c(W_t) - u_1^c(W_s)$. With $U_1(W_t, W_s)$ and its projection, $A_{1j,l}$ is represented as

$$\begin{aligned}
A_{1j,l} &= \frac{2}{T^2}\sum_t\sum_s U_1(W_t, W_s) - \frac{2}{T}\sum_t x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})f(X_{t-1}) \\
&= \frac{2}{T}\sum_t u_1^c(W_t) - \frac{2}{T}\sum_t x_t u_t f^3(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})f(X_{t-1}) + \frac{1}{T^2}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_1(W_t, W_s) \\
&= \frac{2}{T}\sum_t \left\{E\left[x_t u_t \tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}\right] - x_t u_t \tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\right\} f^4(X_{t-1}) + O(h^L) \\
&\quad + \frac{1}{T^2}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_1(W_t, W_s). \qquad \text{(B.1)}
\end{aligned}$$

First we show that the 3rd term of R.H.S is a smaller order term than the first term of R.H.S. $V\left(\frac{1}{T^2}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_1(W_t, W_s)\right)$ includes $E\left[\phi_1^2(W_t, W_s)\right]$, $E\left[\phi_1(W_t, W_s)\phi_1(W_t, W_u)\right]$ and $E\left[\phi_1(W_t, W_s)\phi_1(W_u, W_v)\right]$, we evaluate each term.

$$\begin{aligned}
&\int |\phi_1(w_1, w_2)|^{2+\delta} dF(w_1)dF(w_2) \\
\leq\ &\int |U_1(w_1, w_2)|^{2+\delta} dF(w_1)dF(w_2) + C \qquad \text{(B.2)} \\
\leq\ &C\int |x_1 u_1 f^4(X_1)|^{2+\delta}\frac{1}{h^{p(2+\delta)}}|K(\frac{X_1 - X_2}{h})|^{2+\delta} dF(w_1)dF(w_2) + C \\
\leq\ &C\{h^{-p(1+\delta)} + 1\} = C_{1h}. \qquad \text{(B.3)}
\end{aligned}$$

Lemma 4 yields

$$\left|\int \phi_1(w_1, w_2)^2 dF_{t,s}(w_1, w_2) - \int \phi_1(w_1, w_2)^2 dF(w_1)dF(w_2)\right| \leq 4C_{1h}^{\frac{2}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}}$$

but (B.3) holds also for $\delta = 0$, and thus $\int \phi_1(w_1, w_2)^2 dF(w_1)dF(w_2) = O(h^{-p})$. Combining them, we have

$$\begin{aligned}
E\{\phi_1(W_t, W_s)^2\} &= \int \phi_1(w_1, w_2)^2 dF_{t,s}(w_1, w_2) \\
&\leq C\{h^{-p} + h^{-\frac{2p(1+\delta)}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}}\}.
\end{aligned}$$

Because

$$\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\beta(s-t)^{\frac{\delta}{2+\delta}} = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(s-t)^{-\frac{(2+\eta)\delta}{\eta(2+\delta)}} = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(s-t)^{-1-\gamma} \leq CT,$$

for $\gamma = 2(\delta-\eta)/\{\eta(2+\delta)\} > 0$, we have

$$E\left\{\binom{T}{2}^{-2}\sum_{t=r}^{T-1}\sum_{s=t+1}^{T}\phi_1(W_t,W_s)^2\right\} \leq \frac{1}{T}C(\frac{1}{Th^p}+\frac{1}{T^2h^{\frac{2p(1+\delta)}{2+\delta}}}) = \frac{1}{T}C(\frac{1}{Th^p}+\frac{h^{\frac{2p}{2+\delta}}}{T^2h^{2p}}) = o\left(\frac{1}{T}\right).$$

We consider $\binom{T}{2}^{-2}\sum_{t=r}^{T-2}\sum_{s=t+1}^{T-1}\sum_{u=s+1}^{T}\phi_1(W_t,W_s)\phi_1(W_t,W_u)$ next. We have, similarly to the computation of (B.3),

$$\begin{aligned}
&\int |\phi_1(w_1,w_2)\phi_1(w_1,w_3)|^{1+\frac{\delta}{2}}dF_{t,s}(w_1,w_2)dF_u(w_3)\\
\leq\ &\int |\phi_1(w_1,w_2)|^{1+\frac{\delta}{2}}\{\int |\phi_1(w_1,w_3)|^{1+\frac{\delta}{2}}dF_u(w_3)\}dF_{t,s}(w_1,w_2)\\
\leq\ &Ch^{-p\delta} = C_{2h}.
\end{aligned}$$

Then, by Lemma 4, we have

$$\begin{aligned}
&|\int \phi_1(w_1,w_2)\phi_1(w_1,w_3)dF_{t,s,u}(w_1,w_2,w_3) - \int \phi_1(w_1,w_2)\phi_1(w_1,w_3)dF_{t,s}(w_1,w_2)dF(w_3)|\\
\leq\ &4C_{2h}^{\frac{2}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}}
\end{aligned}$$

Noting $\int \phi_1(w_1,w_3)dF(w_3) = 0$ by construction, we have

$$|E\{\phi_1(W_t,W_s)\phi_1(W_t,W_u)\}| \leq 4C_{2h}^{\frac{2}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}} \leq Ch^{-\frac{2p\delta}{2+\delta}}(s-t)^{-1-\gamma}$$

which yields

$$\left|E\left[\binom{T}{2}^{-2}\sum_{t=r}^{T-2}\sum_{s=t+1}^{T-1}\sum_{u=s+1}^{T}\phi_1(W_t,W_s)\phi_1(W_t,W_u)\right]\right| \leq \frac{1}{T}C\frac{h^{\frac{2-\delta}{2+\delta}}}{Th^p} = o\left(\frac{1}{T}\right).$$

For the quadruple summation, we consider the cases (i) $t<s<u<v$ and $s-t > v-u$ and (ii)$t<s<u<v$ and $s-t \leq v-u$ as in Yoshihara (1976). Because both cases can be handled similarly, we consider only the first case (i). Note that

$$\begin{aligned}
&\int |\phi_1(w_1,w_2)\phi_1(w_3,w_4)|^{1+\frac{\delta}{2}}dF_{t,s,u}(w_1,w_2,w_3)dF(w_4)\\
\leq\ &\int |\phi_1(w_1,w_2)|^{1+\frac{\delta}{2}}\{\int |\phi_1(w_3,w_4)|^{1+\frac{\delta}{2}}dF(w_4)\}dF_{t,s,u}(w_1,w_2,w_3)\\
\leq\ &C(h^{-p\delta}+1) = C_{3h}.
\end{aligned}$$

Then

$$\begin{aligned}
&\left|\int \phi_1(w_1,w_2)\phi_1(w_3,w_4)dF_{t,s,u,v}(w_1,w_2,w_3,w_4) - \int \phi_1(w_1,w_2)\phi_1(w_3,w_4)dF_{t,s,u}(w_1,w_2,w_3)dF\right.\\
\leq\ &4C_{3h}^{\frac{2}{2+\delta}}\beta(v-u)^{\frac{\delta}{2+\delta}}.
\end{aligned}$$

Noting $\int \phi_i(w_3, w_4)dF(w_4) = 0$, we have

$$|E\left[\phi_1(W_t, W_s)\phi_1(W_u, W_v)\right]| \leq 4C_{3h}^{\frac{2}{2+\delta}}\beta(s-t)^{\frac{\delta}{2+\delta}},$$

and thus

$$\begin{aligned}\left|E\left[\begin{pmatrix} T \\ 2 \end{pmatrix}^{-2} \sum\sum_{\substack{t<s<u<v \\ s-t<v-u}}\sum\sum\sum\sum \phi_1(W_t, W_s)\phi_1(W_u, W_v)\right]\right| &\leq \frac{1}{T}CT^{-\gamma}h^{-\frac{2p\delta}{2+\delta}}. \\ &= \frac{1}{T}\frac{Ch^{\frac{2p(\delta-\eta-\eta\delta)}{\eta(2+\delta)}}}{(Th^p)^{\frac{2(\delta-\eta)}{\eta(2+\delta)}}} = o\left(\frac{1}{T}\right).\end{aligned}$$

Since the first term of the right hand side of (B.1) is a sum of stationary martingale difference random variables,

$$\begin{aligned}E\left[(A_{1j,l})^2\right] &= \frac{4}{T}E\left[\left(\left\{E\left[x_t u_t \tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}\right] - x_t u_t \tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\right\} f^4(X_{t-1})\right)^2\right] + O(h^{2L} \\ &\leq \frac{4}{T}E\left[x_t^2 u_t^2 \tilde{q}_j^2(Z_{t-1})\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] + o\left(\frac{1}{T}\right),\end{aligned}$$

then

$$\begin{aligned}\sum_{j=1}^{k_T}\sum_{l=1}^{k_T} E\left[(A_{1j,l})^2\right] &\leq \frac{4}{T}E\left[x_t^2 u_t^2 \sum_{j=1}^{k_T}\tilde{q}_j^2(Z_{t-1})\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \\ &\leq 4\frac{\zeta_0^4(k_T)}{T}E\left[x_t^2 u_t^2 f^8(X_{t-1})\right] \\ &= o(1).\end{aligned}$$

Similarly,

$$\begin{aligned}A_{2j,l} &= \frac{2}{T}\sum_t \left\{u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})m(X_{t-1})f^4(X_{t-1})\right. \\ &\quad \left. -E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_t f^4(X_{t-1})\right\} + O(h^L) \\ &\quad +(\text{small order term}).\end{aligned}$$

The first term of the right hand side is a sum of martingale difference random variables since $x_t = u_t + m(X_{t-1})$ and

$$\begin{aligned}&u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})m(X_{t-1})f^4(X_{t-1}) - E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_t f^4(X_{t-1}) \\ =&\left\{u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})m(X_{t-1}) - E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]m(X_{t-1})\right\} f^4(X_{t-1}) \\ &+ u_t E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_t f^4(X_{t-1}),\end{aligned}$$

both terms of right hand side are martingale differences. Thus

$$\begin{aligned}
& E\left[(A_{2j,l})^2\right] \\
= \; & \frac{4}{T}E\left[\left(u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})g(X_{t-1})f^4(X_{t-1})\right)^2\right] \\
& -\frac{8}{T}E\left[\left(u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})g(X_{t-1})f^4(X_{t-1})\right)E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_tf^4(X_{t-1})\right] \\
& +\frac{4}{T}E\left[\left(E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_tf^4(X_{t-1})\right)^2\right]+o\left(\frac{1}{T}\right) \\
\leq \; & \frac{4}{T}E\left[u_t^2\tilde{q}_j^2(Z_{t-1})\tilde{q}_l^2(Z_{t-1})g^2(X_{t-1})f^8(X_{t-1})\right] \\
& +\frac{8}{T}\left(E\left[\left(u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})g(X_{t-1})f^4(X_{t-1})\right)^2\right]\right)^{1/2} \\
& \left(E\left[\left(E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]x_tf^4(X_{t-1})\right)^2\right]\right)^{1/2} \\
& +\frac{4}{T}E\left[E[u_t\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})|X_{t-1}]^2x_t^2f^8(X_{t-1})\right].
\end{aligned}$$

And

$$\begin{aligned}
\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{2j,l})^2\right] \leq \; & \frac{4}{T}E\left[u_t^2\sum_{j=1}^{k_T}\tilde{q}_j^2(Z_{t-1})\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})g^2(X_{t-1})f^8(X_{t-1})\right] \\
& +\frac{8}{T}\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}\left(4C_q^4E\left[u_t^2g^2(X_{t-1})f^8(X_{t-1})\right]\right)^{1/2} \\
& \left(4C_q^4E\left[E[u_t^2|X_{t-1}]x_t^2f^8(X_{t-1})\right]\right)^{1/2} \\
& +\frac{4}{T}E\left[E[u_t^2\sum_{j=1}^{k_T}\tilde{q}_j^2(Z_{t-1})\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})|X_{t-1}]x_t^2f^8(X_{t-1})\right] \\
\leq \; & 4\frac{\zeta_0^4(k_T)}{T}E\left[u_t^2g^2(X_{t-1})f^8(X_{t-1})\right] \\
& +32C_p^4\frac{k_T^2}{T}\left(E\left[u_t^2g^2(X_{t-1})f^8(X_{t-1})\right]\right)^{1/2}\left(E\left[E[u_t^2|X_{t-1}]x_t^2f^8(X_{t-1})\right]\right)^{1/2} \\
& +4\frac{\zeta_0^4(k_T)}{T}E\left[E[u_t^2|X_{t-1}]x_t^2f^8(X_{t-1})\right] \\
= \; & o\left(\frac{1}{T}\right)
\end{aligned}$$

Similarly

$$A_{3j,l} = -\frac{1}{T}\sum_t\left\{u_t^2\tilde{q}_l(Z_{t-1})q_j(Z_{t-1})-E\left[u_t^2\tilde{q}_l(Z_{t-1})q_j(Z_{t-1})|X_{t-1}\right]\right\}f^4(X_{t-1})+O(h^L),$$

$$\begin{aligned} E\left[(A_{3j,l})^2\right] &\leq \frac{1}{T}E\left[u_t^4\tilde{q}_l^2(Z_{t-1})q_j^2(Z_{t-1})f^8(X_{t-1})\right] + o\left(\frac{1}{T}\right) \\ &\leq C_q^2\frac{1}{T}E\left[u_t^4\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \end{aligned}$$

and

$$\begin{aligned} \sum_{j=1}^{k_T}\sum_{l=2}^{k_T} E\left[(A_{3j,l})^2\right] &\leq C_q^2\frac{1}{T}\sum_{j=1}^{k_T} E\left[u_t^4\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \\ &\leq C_q^2\frac{\zeta_0^2(k_T)k_T}{T}E\left[u_t^4 f^8(X_{t-1})\right] \\ &= o(1). \end{aligned}$$

Since $A_{5j,l} = A_{3j,l}$

$$\sum_{j=1}^{k_T}\sum_{l=2}^{k_T} E\left[(A_{5j,l})^2\right] = o(1).$$

$$A_{4j,l} = \frac{1}{T}\sum_t \left\{u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1}) - E\left[u_t^2\tilde{q}_l(Z_{t-1})|X_{t-1}\right]q_j(Z_{t-1})\right\}f^4(X_{t-1}) + O\left(h^L\right)$$

The first term of the right hand side is a sum of martingale difference random variables since $q_j(Z_{t-1}) = \tilde{q}_j(Z_{t-1}) + r_j(X_{t-1})$ and

$$\begin{aligned} & \left\{u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1}) - E\left[u_t^2\tilde{q}_l(Z_{t-1})|X_{t-1}\right]q_j(Z_{t-1})\right\}f^4(X_{t-1}) \\ = & \left\{u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1}) - E\left[u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1})|X_{t-1}\right]\right\}f^4(X_{t-1}) \\ & -\tilde{q}_j(Z_{t-1})E\left[u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1})|X_{t-1}\right]f^4(X_{t-1}). \end{aligned}$$

Thus

$$\begin{aligned}
E\left[(A_{4j,l})^2\right] &= \frac{1}{T}E\left[u_t^4\tilde{q}_l^2(Z_{t-1})r_j^2(X_{t-1})f^8(X_{t-1})\right] \\
&\quad -\frac{2}{T}E\left[u_t^2\tilde{q}_l(Z_{t-1})r_j(X_{t-1})E\left[u_t^2\tilde{q}_l(Z_{t-1})|X_{t-1}\right]q_j(Z_{t-1})f^8(X_{t-1})\right] \\
&\quad +\frac{1}{T}E\left[\left(E\left[u_t^2\tilde{q}_l(Z_{t-1})|X_{t-1}\right]\right)^2 q_j(Z_{t-1})f^8(X_{t-1})\right]+o\left(\frac{1}{T}\right) \\
&\leq C_q^2\frac{1}{T}E\left[u_t^4\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \\
&\quad +\frac{2}{T}\left(E\left[u_t^4\tilde{q}_l^2(Z_{t-1})g_j^2(X_{t-1})f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad \left(E\left[\left(E\left[u_t^2\tilde{q}_l(Z_{t-1})|X_{t-1}\right]\right)^2 f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad +C_q^2\frac{1}{T}E\left[E\left[u_t^4\tilde{q}_l^2(Z_{t-1})|X_{t-1}\right]f^8(X_{t-1})\right] \\
&\leq C_q^2\frac{1}{T}E\left[u_t^4\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \\
&\quad +4C_q^4\frac{1}{T}\left(E\left[u_t^4 f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad \left(E\left[\left(E\left[u_t^2|X_{t-1}\right]\right)^2 f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad +C_q^2\frac{1}{T}E\left[E\left[u_t^4\tilde{q}_l^2(Z_{t-1})|X_{t-1}\right]f^8(X_{t-1})\right].
\end{aligned}$$

Then,

$$\begin{aligned}
\sum_{j=1}^{k_T}\sum_{l=1}^{k_T} E\left[(A_{4j,l})^2\right] &\leq C_q^2\frac{1}{T}\sum_{j=1}^{k_T} E\left[u_t^4\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})f^8(X_{t-1})\right] \\
&\quad +4C_q^4\frac{1}{T}\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}\left(E\left[u_t^4 f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad \left(E\left[E\left[u_t^4|X_{t-1}\right]f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad +C_q^2\frac{1}{T}\sum_{j=1}^{k_T} E\left[E\left[u_t^4\sum_{l=1}^{k_T}\tilde{q}_l^2(Z_{t-1})|X_{t-1}\right]f^8(X_{t-1})\right] \\
&\leq C_q^2\frac{\zeta_0^2(k_T)k_T}{T}E\left[u_t^4 f^8(X_{t-1})\right] \\
&\quad +4C_q^4\frac{k_T^2}{T}\left(E\left[u_t^4 f^8(X_{t-1})\right]\right)^{1/2}\left(E\left[E\left[u_t^4|X_{t-1}\right]f^8(X_{t-1})\right]\right)^{1/2} \\
&\quad +C_q^2\frac{\zeta_0^2(k_T)k_T}{T}E\left[E\left[u_t^4|X_{t-1}\right]f^8(X_{t-1})\right] \\
&= o(1).
\end{aligned}$$

$\sum_{j=1}^{k_T}\sum_{l=1}^{k_T} E\left[(A_{6j,l})^2\right] = o(1)$, since $A_{6j,l} = A_{4l,j}$.

Denote $m_{j,l}(W_t)$ be

$$m_{j,l}(W_t) \equiv u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1}).$$

The last term is

$$\begin{aligned}
A_{7j,l} &= \frac{1}{T}\sum_{t=1}^{T} u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1}) - E\left[u_t^2 f^4(X_{t-1})\tilde{q}_j(Z_{t-1})\tilde{q}_l(Z_{t-1})\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} m_{j,l}(W_t) - E[m_{j,l}(W_t)],
\end{aligned}$$

and

$$\begin{aligned}
E\left[(A_{7j,l})^2\right] &= \frac{2}{T^2}\sum_{t}\sum_{s\leq t}\mathrm{Cov}(m_{j,l}(W_t), m_{jl}(W_s)) \\
&= \frac{2}{T^2}\sum_{t}\sum_{i=0}^{t-1}\mathrm{Cov}(m_{j,l}(W_t), m_{jl}(W_{t-i})).
\end{aligned}$$

Since an absolute regular process is anα-mixing process and $\alpha_m \leq \beta_m$, mixing inequality implies

$$\begin{aligned}
\mathrm{Cov}(m_{j,l}(W_t), m_{j,l}(W_{t-i})) &\leq 2(2^{1-1/(2+\delta)}+1)\alpha_i^{1-2/(2+\delta)}\left\|m_{j,l}(W_t)\right\|_{2+\delta}^2 \\
&\leq 2(2^{1-1/(2+\delta)}+1)\beta_i^{1-2/(2+\delta)}\left\|m_{j,l}(W_t)\right\|_{2+\delta}^2
\end{aligned}$$

where $\|m_{j,l}(W_t)\|_{2+\delta} = \left(E\left[|m_{j,l}(W_t)|^{2+\delta}\right]\right)^{1/(2+\delta)}$. Thus

$$\begin{aligned} E\left[(A_{7j,l})^2\right] &= \frac{2}{T^2}\sum_t\sum_{i=0}^{t-1}\mathrm{Cov}(m_{j,l}(W_t), m_{jl}(W_{t-i})) \\ &\leq 4(2^{1-1/(2+\delta)}+1)\,\|m_{j,l}(W_t)\|_{2+\delta}^2\,\frac{1}{T^2}\sum_t\sum_{i=0}^{t}\beta_i^{1-2/(2+\delta)} \\ &\leq C_q^4 4(2^{1-1/(2+\delta)}+1)\left(E\left[\left(x_t^2 f^4(X_{t-1})\right)^{2+\delta}\right]\right)^{2/(2+\delta)}\frac{1}{T^2}\sum_t\sum_{i=0}^{t}\beta_i^{1-2/(2+\delta)}, \end{aligned}$$

and

$$\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{7j,l})^2\right] \leq \frac{k_T^2}{T}C_q^4 4(2^{1-1/(2+\delta)}+1)\left(E\left[\left(x_t^2 f^4(X_{t-1})\right)^{2+\delta}\right]\right)^{2/(2+\delta)}\frac{1}{T}\sum_t\sum_{i=0}^{t}\beta_i^{1-2/(2+\delta)}.$$

Combining above results, we have,

$$\begin{aligned} & E\left\|\hat{M}-M\right\|^2 \\ \leq\ & \sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{1j,l})^2\right]+\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{2j,l})^2\right]+\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{3j,l})^2\right] \\ & +\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{4j,l})^2\right]+\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{5j,l})^2\right]+\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{6j,l})^2\right] \\ & +\sum_{j=1}^{k_T}\sum_{l=1}^{k_T}E\left[(A_{7j,l})^2\right] \\ \leq\ & C_1\frac{\zeta_0^4(k_T)}{T}+C_2\frac{k_T^2}{T}+C_3\frac{\zeta_0^2(k_T)k_T}{T} \\ \leq\ & C\max\left(\frac{\zeta_0^4(k_T)}{T},\frac{k_T^2}{T},\frac{\zeta_0^2(k_T)k_T}{T}\right) \\ =\ & o(1). \end{aligned}$$

Lemma 6 *Let X be a symmetric and positive definite $k\times k$ matrix. It is possible to define symmetric matrices $X^{\frac{1}{2}}$ and $X^{-\frac{1}{2}}$ satisfying $X^{\frac{1}{2}}X^{\frac{1}{2}}=X$ and $X^{-\frac{1}{2}}=(X^{\frac{1}{2}})^{-1}$. If $X^{-\frac{1}{2}}$ is differentiable with respect to X,*

$$(i)\ \ \frac{\partial vec(X^{-\frac{1}{2}})}{\partial vec(X)'} = -(X^{-\frac{1}{2}}\otimes X^{-\frac{1}{2}})(X^{\frac{1}{2}}\otimes I_k + I_k\otimes X^{\frac{1}{2}})^{-1}.$$

Suppose further X *has eigenvalues* $\lambda_k(X) \geq \cdots \geq \lambda_1(X) > 0$ *and corresponding* $k \times k$ *eigenvector matrix* E_X *satisfying* $E_X E_X' = E_X' E_X = I_k$. *Similarly, let* M *be a* $k \times k$ *symmetric and positive definite matrix with eigenvalue decomposition* $M = E\Lambda E'$ *where* $\Lambda = diag(\lambda_k, \cdots, \lambda_1)$ *with* $\lambda_k \geq \cdots \geq \lambda_1 > 0$ *and* $EE' = E'E = I_k$. *Define* $X^{-\frac{1}{2}} = E_X \Lambda_X^{-\frac{1}{2}} E_X'$ *with* $\Lambda_X = diag(\lambda_1(X), \cdots, \lambda_k(X))$, $M^{-\frac{1}{2}} = E\Lambda^{-\frac{1}{2}}E'$ *and* $M^{\frac{1}{2}} = E\Lambda^{\frac{1}{2}}E'$. *When* $X^{-\frac{1}{2}}$ *is differentiable at* $X = M$, *then we have*

$$(ii)\ \ vec(X^{-\frac{1}{2}} - M^{-\frac{1}{2}}) = -(M^{-\frac{1}{2}} \otimes M^{-\frac{1}{2}})(M^{\frac{1}{2}} \otimes I_k + I_k \otimes M^{\frac{1}{2}})^{-1} vec(X - M) + o(||X - M||)$$

and

$$(iii)\ \ ||X^{-\frac{1}{2}} - M^{-\frac{1}{2}}||^2 \leq \frac{K||X - M||^2}{\lambda_1^3} + o(||X - M||^2).$$

(Proof)

Let $X^{\frac{1}{2}} = E_X \Lambda_X^{\frac{1}{2}} E_X'$, then $X^{\frac{1}{2}} X^{-\frac{1}{2}} = I_k$ and $X^{\frac{1}{2}} X^{\frac{1}{2}} = X$. Differentiating both equalities, we have

$$(dX^{\frac{1}{2}})X^{-\frac{1}{2}} + X^{\frac{1}{2}}(dX^{-\frac{1}{2}}) = dI_k = 0$$

and

$$(dX^{\frac{1}{2}})X^{\frac{1}{2}} + X^{\frac{1}{2}}(dX^{\frac{1}{2}}) = dX.$$

Because $vec(ABC) = (C' \otimes A)vec(B)$ and $(A \otimes B)(C \otimes D) = AC \otimes BD$ for comformable matrice A, B, C, D, we obtain,

$$vec(dX^{-\frac{1}{2}}) = -(I_k \otimes X^{\frac{1}{2}})^{-1}(I_k \otimes X^{-\frac{1}{2}})vec(dX^{\frac{1}{2}}) = -(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}})vec(dX^{\frac{1}{2}}) \quad \text{(B.4)}$$

and

$$vec(dX^{\frac{1}{2}}) = (I_k \otimes X^{\frac{1}{2}} + X^{\frac{1}{2}} \otimes I_k)^{-1} vec(dX). \quad \text{(B.5)}$$

Substituting (B.5) into (B.4), we have

$$vec(dX^{-\frac{1}{2}}) = -(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}})(I_k \otimes X^{\frac{1}{2}} + X^{\frac{1}{2}} \otimes I_k)^{-1} vec(dX)$$

which completes the proof for (i). It is straightforward that (ii) holds if $X^{-\frac{1}{2}}$ is differentiable at $X = M$. To prove (iii), using (ii), we have

$$\begin{aligned}
& ||X^{-\frac{1}{2}} - M^{-\frac{1}{2}}||^2 = vec(X^{-\frac{1}{2}} - M^{-\frac{1}{2}})' vec(X^{-\frac{1}{2}} - M^{-\frac{1}{2}}) \\
= \ & vec(X - M)'(I_k \otimes M^{\frac{1}{2}} + M^{\frac{1}{2}} \otimes I_k)^{-1}(M^{-1} \otimes M^{-1})(M^{\frac{1}{2}} \otimes I_k + I_k \otimes M^{\frac{1}{2}})^{-1} vec(X - M) \\
& + o(||X - M||^2) \\
\leq \ & ||vec(X - M)||^2 ||(I_k \otimes M^{\frac{1}{2}} + M^{\frac{1}{2}} \otimes I_k)^{-1}(M^{-1} \otimes M^{-1})(M^{\frac{1}{2}} \otimes I_k + I_k \otimes M^{\frac{1}{2}})^{-1}|| \\
& + o(||X - M||^2).
\end{aligned}$$

We evaluate the norms in the last quantity. We obviously have $||vec(X-M)||^2 = ||X-M||^2$. Because $M^{\frac{1}{2}} = E\Lambda^{\frac{1}{2}}E'$ and $M^{-1} = E\Lambda^{-1}E'$, we have,

$$M^{\frac{1}{2}} \otimes I_k + I_k \otimes M^{\frac{1}{2}} = (E \otimes E)(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})(E' \otimes E')$$

and

$$M^{-1} \otimes M^{-1} = (E \otimes E)(\Lambda^{-1} \otimes \Lambda^{-1})(E' \otimes E').$$

Therefore, using $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ for invertible matrices A, B, we have

$$||(I_k \otimes M^{\frac{1}{2}} + M^{\frac{1}{2}} \otimes I_k)^{-1}(M^{-1} \otimes M^{-1})(M^{\frac{1}{2}} \otimes I_k + I_k \otimes M^{\frac{1}{2}})^{-1}||^2$$

$$\begin{aligned}
= \;& ||(E \otimes E)(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-1}(E' \otimes E') \\
& \times (E \otimes E)(\Lambda^{-1} \otimes \Lambda^{-1})(E' \otimes E') \\
& \times (E \otimes E)(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-1}(E' \otimes E')||^2 \\
= \;& ||(E \otimes E)(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-1}(\Lambda^{-1} \otimes \Lambda^{-1})(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-1}(E' \otimes E')||^2 \\
\leq \;& tr((\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-2}(\Lambda^{-2} \otimes \Lambda^{-2})(\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}})^{-2}) \\
= \;& \sum_{i=1}^{k}\sum_{i=1}^{k} \frac{1}{(\sqrt{\lambda_i} + \sqrt{\lambda_j})^4} \cdot \frac{1}{\lambda_i^2 \lambda_j^2} \leq \frac{k^2}{\lambda_1^6}.
\end{aligned}$$

We use that $\Lambda^{\frac{1}{2}} \otimes I_k + I_k \otimes \Lambda^{\frac{1}{2}}$ and $\Lambda^{-1} \otimes \Lambda^{-1}$ are diagonal matrices in the last inequality. Thus

$$||X^{-\frac{1}{2}} - M^{-\frac{1}{2}}||^2 \leq \frac{k||X-M||^2}{\lambda_1^3} + o(||X-M||^2).$$

References

[1] Bierens, H.J. (2004) Introduction to the Mathematical and Statistical Foundations of Econometrics, Cambridge University Press.

[2] Bierens, H.J. and W. Ploberger (1997) "Asymptotic Theory of Integrated Conditional Moment Tests," Econometrica, 65, 5, 1129-1151.

[3] Breitung, J. and B. Candelon (2006) "Testing for Short- and Long-run Causality: A Frequency-domain Approach," Journal of Econometrics, 132, 363-378.

[4] Chao, J., V. Corradi and N. R. Swanson (2001), "Out-of-Sample Tests for Granger Causality," Macroeconomic Dynamics, 5, 598-620.

[5] Chen, X. and Y. Fan (1999) "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," Journal of Econometrics, 91, 373-401.

[6] Comte, F. and O. Lieberman (2000), "Second-Order Noncausality in Multivariate GARCH Processes," Journal of Time Series Analysis, 21, 5, 535-557.

[7] De Jong and H.J. Bierens (1994) "On th Limit Behavior of a Chi-Squared Type Test if the Number of Conditional Moments Tested Approaches to Infinity," Econometric Theory, 9, 70-90.

[8] 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.

[9] Diks, C. and V. Panchenko (2005), "A note on the Hiemstra-Jones test for Granger noncausality," Studies in Nonlinear Dynamics & Econometrics, 9, 2, 1-7.

[10] Doornik, J.A. (2007), Object-Oriented Matrix Programming Using Ox, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

[11] Dufour, J., D. Pelletier and É. Renault (2006), "Short Run and Long Run Causality in Time Series: Inference," Journal of Econometrics, 132, 337-362.

[12] Dufour, J., and É. Renault (1998) "Short Run and Long Run Causality in Time Series: Theory," Econometrica, 66, 1099-1125.

[13] Fan, Y. and Q. Li (1996) (1996) "Consistent Model Specification Tests: Omitted variables and Semiparametric Functional Forms," Econometrica, 64, 4, 865-890.

[14] Gao, J. and M. King (2004) "Adaptive Estimation in Continuous-Time Diffusion Models," Econometric Therory, 20, 844-882.

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

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

[17] Hansen, B.E. (2008) "Uniform Convergence Rates for Kernel Estimation with Dependent Data," Econometric Theory, 24, 3, 726-748.

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

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

[20] Hitomi, K. (2000) "Common Structure of Consistent Misspecification Tests and a New Test," mimeo.

[21] Hong, Y. and A. White (1995) "Consistent Specification Testing via Nonparametric Series Regression," Econometrica, 63, 5, 1133-1159.

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

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

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

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

[26] McCrorie, J. R. and M.J. Chambers (2006) "Granger Causality and the Sampling of Economic Processes," Journal of Econometrics, 132, 311-336.

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

[28] Péguin-Feissolle, A. and T. Teräsvirta (1999), "A General Framework for Testing the Granger Noncausality Hypothesis," SSE/EFI Working Paper Series in Economics and Finance, No. 343. Stockholm School of Economics.

[29] Qiao, Z, M. McAleer and W.K. Wong (2009) "Linear and Nonlinear Causality between Changes in Consumption and Consumer Attitudes," Economics Letters, 102, 161-164.

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

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

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

[33] 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.

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

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

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

[37] Von Neumann, J. (1941) "Distribution of the Ratio of the Mean Squared successive Difference to the Variance," Annals of Mathematical Statistics, 12, 367-395.

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