12-2007

A consistent characteristic function-based test for conditional independence

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Halbert WHITE

A consistent characteristic function-based test for conditional independence

Article *in* Journal of Econometrics · December 2007

DOI: 10.1016/j.jeconom.2006.11.006

Liangjun Su
Singapore Management University

A Consistent Characteristic Function-Based Test for Conditional Independence

Liangjun Su[a], Halbert White[b*]
[a]Guanghua School of Management, Peking University, Beijing 100871, China
[b]Department of Economics, UCSD, La Jolla, CA 92093-0508, USA

Abstract

Y is conditionally independent of Z given X if $\Pr\{f(y|X,Z) = f(y|X)\} = 1$ for all y on its support, where $f(\cdot|\cdot)$ denotes the conditional density of Y given (X,Z) or X. This paper proposes a nonparametric test of conditional independence based on the notion that two conditional distributions are equal if and only if the corresponding conditional characteristic functions are equal. We extend the test of Su and White (2005) in two directions: (1) our test is less sensitive to the choice of bandwidth sequences; (2) our test has power against deviations on the full support of the density of (X,Y,Z). We establish asymptotic normality for our test statistic under weak data dependence conditions. Simulation results suggest that the test is well behaved in finite samples. Applications to stock market data indicate that our test can reveal some interesting nonlinear dependence that a traditional linear Granger causality test fails to detect.

JEL classification: C12; C14; C22

Keywords: Conditional characteristic function; Conditional independence; Granger non-causality; Nonparametric regression; U-statistics

*Corresponding author. Tel.: +1-858-534-3502; fax: +1-858-534-7040.
E-mail address: hwhite@econ.ucsd.edu (H. White).

1 Introduction

In this paper, we investigate a nonparametric test of conditional independence. Let X, Y and Z be random variables. As in Su and White (2005, "SW"), we write

$$Y \perp Z \mid X \tag{1.1}$$

to denote that Y is independent of Z given X, i.e., $\Pr\{f(y|X,Z) = f(y|X)\} = 1$ for all y on its support, where $f(y|x,z)$ is the conditional density of Y given $(X,Z) = (x,z)$ and $f(y|x)$ is that of Y given $X = x$.

In comparison with the number of nonparametric tests of independence or serial independence in the literature, there are few nonparametric tests for conditional independence of continuous variables. Tests previously given include those of Linton and Gozalo (1997, "LG"), Fernandes and Flores (1999), and Delgado and González-Manteiga (2001, "DG"). More recently, SW have proposed a test for conditional independence based on a weighted version of the Hellinger distance between the two conditional densities $f(y|x,z)$ and $f(y|x)$, and they show that the asymptotic null distribution of their test statistic is normal. Although this test is straightforward to implement, it has two limitations. First, it uses the same bandwidth sequence in estimating all required joint and marginal densities nonparametrically; this is unsatisfactory when the dimension of (X,Y,Z) exceeds three. Second, their test can only detect deviations from conditional independence on a compact subset of the support of the joint density of (X,Y,Z).

Here we study the use of conditional characteristic functions (CCFs) to test for conditional independence, motivated by the following considerations: (1) the ability of CCFs to characterize conditional independence, based on the fact that two conditional distributions are identical if and only if their respective CCFs are equal; (2) the demonstrated ability of empirical characteristic functions (ECFs) to yield well-behaved, powerful tests for other important distributional hypotheses, such as goodness-of-fit, symmetry, homogeneity, independence, and serial independence (see Hong (1999) for a brief account); (3) the appeal of obtaining a test with power complementary to that of previous tests; (4) the desirability of obtaining a computationally convenient test, based on limiting normal or chi-squared distributions, as in Hong (1999) or Brett and Pinkse (1997) and Pinkse (1998, 2000), who use characteristic function-based approaches to test for independence, serial independence, and spatial independence; and (5) the appeal of obtaining a test statistic whose limiting distribution does not depend on the presence of estimated parameters.

Concerning our last three motivations, we began with only a strong suspicion, based on the previously cited work, that a CCF approach would yield tests with these appealing properties. As we prove, however, this approach does indeed deliver the desired properties. The extreme generality of the alternative hypothesis here makes it correspondingly difficult to study the global efficiency (e.g., rate-optimality or minimaxity) of any particular test. Indeed, as an Associate Editor has noted, it is possible to construct a large variety of different tests by employing sample analogs of characterizations of conditional independence (e.g., integral transforms) other than CCF. Nevertheless, because each such test necessarily exploits certain features of the data generating process at the expense of others, complementarities between tests can easily arise. For example, DG's test effectively uses only certain low-frequency information, as it is based on the empirical distribution function; in contrast, our test is more powerful against high frequency alternatives. This validates our third motivation.

We achieve the goal of our fourth motivation by exploiting ideas from the approaches of Hong (1999) and Pinkse (1998, 2000). Like them, we base our test upon the properties of characteristic functions and use a weighted integral approach. Unlike them, we test for conditional independence instead of (serial or spatial) independence; the conditioning significantly complicates matters. We also exploit ideas from Bierens (1982, 1990) on consistent specification testing; our test is thus consistent against all deviations from conditional independence on the full support of the density of (X, Y, Z). Unlike Bierens' tests and those of LG, Fernandez-Florez (1999), and DG, and like the test of SW, our test statistic has a normal null distribution asymptotically.

Finally, we prove that the asymptotic null distribution of our CCF-based statistic is not affected by $\sqrt{n}$-consistent estimation of unknown parameters. In contrast, DG's test yields a statistic whose asymptotic null distribution typically is affected, as it is based on the Cramér-von Mises criterion. This validates our fifth motivation.

Our paper offers a convenient approach to testing for distributional hypotheses via an infinite number of conditional moment regressions, and by relying on the properties of CCFs, it unifies the two branches of the literature in an insightful way. A variety of interesting and important hypotheses other than conditional independence in economics and finance, including conditional goodness-of-fit, conditional homogeneity, conditional quantile restrictions, and conditional symmetry, can also be studied using our approach. These tests are naturally suited to answering such questions as "Are the distributions of assets, consumption, or income implied by a particular dynamic macroeconomic model close to the actual distributions in the data?" "Is there any significant difference in wage distributions between blacks and whites (or any two of the ethnics) conditional on their characteristics such as age, education and experience?" or "Does the stock market react symmetrically to positive and negative shocks after taking into account the influence of all fundamentals?"

It is well known that distributional Granger non-causality (Granger, 1980) is a particular case of conditional independence. Our test can be directly applied to test for Granger non-causality without the need to specify a particular linear or non-linear model. Additionally, our test can be applied to the situation where not all variables of interest are continuously valued or observable. In particular, our test applies to situations where limited dependent variables or discrete conditioning variables are involved. Further, it is common in econometrics that conditional independence tests would be conducted using estimated residuals or other estimated random variables, which are a function of the observed data and some parameter estimators. It is straightforward to show that parameter estimation error has no effect on the asymptotic null distribution of our test statistic. For other motivational examples and potential applications of our test, see LG and SW.

The remainder of this paper is organized as follows. In Section 2, we describe the basic framework for our nonparametric test for conditional independence when there is no parameter estimation involved and all random variables are continuously valued. In section 3 we study the asymptotic null distribution of the test statistic and discuss the local power properties of our test. We examine the finite sample performance of our test via Monte Carlo simulation in Section 4. We apply our test to stock market data in Section 5. Final remarks are contained in Section 6. All technical details are relegated to Appendices

A through C.

2 Basic Framework

In this paper, we are interested in the question of whether Y and Z are independent conditional on X, where X, Y and Z are vectors of dimension d_1, d_2 and d_3, respectively. The data consist of n identically distributed but weakly dependent observations $(X_t,\ Y_t,\ Z_t)$, $t = 1, ..., n$.

The joint density (cumulative distribution function) of $(X_t,\ Y_t,\ Z_t)$ is denoted by f (F). Below we make reference to several marginal densities from $f(x, y, z)$ which we denote simply using the list of their arguments – for example $f(x, y) = \int f(x, y, z)dz$, $f(x, z) = \int f(x, y, z)dy$ and $f(x) = \int f(x, y, z)dydz$ where $\int$ denotes integration on the full range of the argument of integration. This notation is compact, and, we hope, sufficiently unambiguous.

Further, let $f(\cdot|\cdot)$ denote the conditional density of one random vector given another. The null of interest is that conditional on X, the random vectors Y and Z are independent, i.e.,

$$H_0 : \ \Pr\{f(y|X, Z) = f(y|X)\} = 1 \ \forall y \in \mathbb{R}^{d_2}. \tag{2.1}$$

The alternative hypothesis is

$$H_1 : \ \Pr\{f(y|X, Z) = f(y|X)\} < 1 \text{ for some } y \in \mathbb{R}^{d_2}. \tag{2.2}$$

The proposed test is based on CCFs. It is well known that two conditional distribution functions are equal almost everywhere (a.e.) if and only if their respective conditional characteristic functions are equal a.e.. To state this precisely, let ψ be the difference between the CCF $\phi_{Y|X,Z}$ of Y conditional on (X, Z) and the CCF $\phi_{Y|X}$ of Y conditional on X, i.e.,

$$\begin{aligned} \psi(u; x, z) &\equiv \phi_{Y|X,Z}(u; x, z) - \phi_{Y|X}(u; x) \\ &= E[\exp(iu'Y)|X = x, Z = z] - E[\exp(iu'Y)|X = x], \end{aligned}$$

where $i = \sqrt{-1}$ and $u \in \mathbb{R}^{d_2}$ is a real-valued vector. Y and Z are independent conditional on X if and only if $\psi(u; x, z) = 0$ a.e.-(x, z) for every $u \in \mathbb{R}^{d_2}$.

Consider the following smooth functional

$$\Gamma \equiv \int_S \int_A \left| \int \psi(u; x, z) e^{i\tau' u} dG_0(u) \right|^2 a(x, z) dF(x, z) dG(\tau), \tag{2.3}$$

where $a(x, z)$ is a given known nonnegative weighting function with full support on $\mathbb{R}^{d_1+d_3}$; and $dG_0(u) = g_0(u)du$ and $dG(\tau) = g(\tau)d\tau$, where we choose g_0 to be a density function with full support on $\mathbb{R}^{d_2}$ and the choice for g is arbitrary except that it must be nonnegative, integrable, and bounded with full support on $\mathbb{R}^{d_2}$.

The choice of the above functional is intuitive. Under the null, $\psi(u; x, z) = 0$ a.e.-(x, z) for every $u \in \mathbb{R}^{d_2}$, and consequently $\Gamma = 0$. The following lemma says that the converse is also true.

Lemma 2.1 $\int \psi(u;x,z)e^{i\tau' u}dG_0(u) = 0$ *a.e.-F on* $\mathbb{R}^{d_1+d_3}$ *for every* $\tau \in \mathbb{R}^{d_2}$ *if and only if* $\psi(u;x,z) = 0$ *a.e.-$G_0 \times F$ on* $\mathbb{R}^{d_2} \times \mathbb{R}^{d_1+d_3}$.

The proof is given in Appendix A. It is an extension of the proof of Theorem 1 in Bierens (1982). Bierens (1982, 1990) proposes consistent tests for functional form of nonlinear regression models based on a Fourier transform of conditional expectations. Consider a generic regression $Y = g(X) + \varepsilon$, where Y is the dependent variable (with $d_2 = 1$), X is the independent variable and ε is the error term. Suppose one has specified the regression function $g(x)$ as $f(x, \theta_0)$, where $f(x, \theta)$ defines a known real-valued Borel measurable function on $\mathbb{R}^{d_2} \times \Theta$ and Θ is a parameter space containing the unknown "true" parameter θ_0 if the specification is correct. Under the null of correct specification, i.e., $\Pr[g(X) = f(X, \theta_0)] = 1$ for some $\theta_0 \in \Theta$, Bierens (1982) shows that the test based on the sample analogue of $E[(Y - f(X,\theta_0))e^{i\tau' X}]$ (which is 0 for every $\tau \in \mathbb{R}^{d_1}$ under the null) is consistent. The test function $e^{i\tau' X}$ depends on the nuisance parameter τ. Stinchcombe and White (1998) generalize this idea to allow the test function to be any non-polynomial analytic function.

An important point concerning (2.3) is that it is straightforward to develop asymptotic theory for the resulting test statistic. Under some regularity conditions (to allow the change of order of integration), one can write $\int \psi(u;x,z)e^{i\tau' u}dG_0(u) = \int\int e^{iu'(y+\tau)}[f(y|x,z) - f(y|x)]dG_0(u)dy$. Define $H(y) \equiv \int e^{iu' y}dG_0(u)$, the characteristic function of the probability measure $dG_0(u)$. Then one can write

$$\Gamma = \int\int |E[H(Y+\tau)|x,z] - E[H(Y+\tau)|x]|^2 \, a(x,z)dF(x,z)dG(\tau). \tag{2.4}$$

This integral facilitates application of the convenient asymptotic distribution theory for U-statistics.

To introduce the test statistic of interest, we first introduce kernel estimators for the unknown conditional expectations above. For a kernel function K and bandwidth $h \equiv h(n)$, we define

$$K_h(u) \equiv h^{-d}K(u/h),$$

where d is the dimension of the vector u. Let $m(x,z;\tau) \equiv E[H(Y+\tau)|X = x, Z = z]$ and $m(x;\tau) \equiv E[H(Y+\tau)|X = x]$. We estimate the latter two conditional expectations by the standard Nadaraya-Watson (NW) leave-one-out kernel regression technique:

$$\widehat{m}_{h_1}(X_t, Z_t;\tau) \equiv \left\{(n-1)^{-1} \sum_{s=1, s\neq t}^{n} K_{h_1}(X_t - X_s, Z_t - Z_s)H(Y_s+\tau)\right\} / \widehat{f}_{h_1}(X_t, Z_t),$$

and

$$\widehat{m}_{h_2}(X_t;\tau) \equiv \left\{(n-1)^{-1} \sum_{s=1, s\neq t}^{n} K_{h_2}(X_t - X_s)H(Y_s+\tau)\right\} / \widehat{f}_{h_2}(X_t),$$

where $\widehat{f}_{h_1}(X_t, Z_t) \equiv (n-1)^{-1}\sum_{s=1,s\neq t}^{n} K_{h_1}(X_t - X_s, Z_t - Z_s)$, and $\widehat{f}_{h_2}(X_t) \equiv (n-1)^{-1}\sum_{s=1,s\neq t}^{n} K_{h_1}(X_t - X_s, Z_t - Z_s)$. Note that we have used different bandwidths in estimating the two conditional expectations. In the sequel we will refer to $\widehat{m}_{h_1}(x,z;\tau)$ as the unrestricted regression estimator and $\widehat{m}_{h_2}(x;\tau)$ as the restricted regression estimator. A natural test statistic immediately follows as

$$\Gamma_{1n} \equiv \frac{1}{n}\sum_{t=1}^{n}\int |\widehat{m}_{h_1}(X_t, Z_t;\tau) - \widehat{m}_{h_2}(X_t;\tau)|^2 a(X_t, Z_t)dG(\tau). \tag{2.5}$$

Three main issues arise in analyzing Γ_{1n}: (1) bias reduction, (2) the random denominator, and (3) the choice of $a(\cdot)$. The latter two are closely tied to each other. For the first issue, as demonstrated in an earlier version of this paper, there are three bias terms to be corrected when using Γ_{1n} as a test statistic, two of which can be removed by appealing to the clever centering device of Härdle and Mammen (1993). Given $\widehat{m}_{h_2}(x;\tau)$, we can compute a smoothed version, $\widehat{sm}_{h_1}(x,z;\tau)$, of $\widehat{m}_{h_2}(x;\tau)$ by regressing $\widehat{m}_{h_2}(X_t;\tau)$ on (X_t,Z_t) , and basing the test on the difference between $\widehat{m}_{h_1}(x,z;\tau)$ and $\widehat{sm}_{h_1}(x,z;\tau)$. For the moment, assume the data are i.i.d. We are thus led to replace $\widehat{m}_{h_2}(X_t;\tau)$ in (2.5) with

$$\widehat{sm}_{h_1}(X_t,Z_t;\tau) \equiv \left\{(n-1)^{-1}\sum_{s=1,s\neq t}^{n} K_{h_1}(X_t-X_s,Z_t-Z_s)\widehat{m}_{h_2}(X_s;\tau)\right\}/\widehat{f}_{h_1}(X_t,Z_t) \tag{2.6}$$

to form

$$\Gamma_{2n} \equiv \frac{1}{n}\sum_{t=1}^{n}\int \left[\widehat{m}_{h_1}(X_t,Z_t;\tau)-\widehat{sm}_{h_1}(X_t,Z_t;\tau)\right]^2 a(X_t,Z_t)dG(\tau). \tag{2.7}$$

In principle, one can choose any positive weighting function a that has support on $\mathbb{R}^{d_1+d_3}$. Nevertheless, we would like to choose a so that we can avoid the random denominator issue. If we were to choose $a(X_t,Z_t)$ to be $\widehat{f}^2_{h_1}(X_t,Z_t)$, after multiplication by $\widehat{f}_{h_1}(X_t,Z_t)$ the random denominators in both $\widehat{m}_{h_1}(X_t,Z_t;\tau)$ and $\widehat{sm}_{h_1}(X_t,Z_t;\tau)$ would disappear. But we still have the third random denominator built into $\widehat{m}_{h_2}(X_s;\tau)$, which is used to form $\widehat{sm}_{h_1}(X_t,Z_t;\tau)$ (see (2.6)). There seems to be no choice of a that would enable us to avoid this.

Note that we can rewrite (2.4) as

$$\Gamma = \int\int \left|E[H(Y+\tau)f(X)|x,z]-E[H(Y+\tau)f(X)|x]\right|^2 \widetilde{a}(x,z)dF(x,z)dG(\tau), \tag{2.8}$$

where $\widetilde{a}(x,z)\equiv a(x,z)/f^2(x)$. We then consider the functional

$$\Gamma_{3n} = \frac{1}{n}\sum_{t=1}^{n}\int |\widetilde{m}_{h_1}(X_t,Z_t;\tau)-\widetilde{sm}_{h_1}(X_t,Z_t;\tau)|^2\widehat{f}^2_{h_1}(X_t,Z_t)dG(\tau), \tag{2.9}$$

where

$$\widetilde{m}_{h_1}(X_t,Z_t;\tau) \equiv \left\{(n-1)^{-1}\sum_{s=1,s\neq t}^{n} K_{h_1}(X_t-X_s,Z_t-Z_s)H(Y_s+\tau)\widehat{f}_{h_2}(X_s)\right\}/\widehat{f}_{h_1}(X_t,Z_t), \tag{2.10}$$

and

$$\widetilde{sm}_{h_1}(X_t,Z_t;\tau) \equiv \left\{(n-1)^{-1}\sum_{s=1,s\neq t}^{n} K_{h_1}(X_t-X_s,Z_t-Z_s)\widehat{m}_{h_2}(X_s;\tau)\widehat{f}_{h_2}(X_s)\right\}/\widehat{f}_{h_1}(X_t,Z_t). \tag{2.11}$$

In other words, $\widetilde{m}_{h_1}(x,z;\tau)$ is an estimator of $E[H(Y+\tau)f(X)|x,z]$ and $\widetilde{sm}_{h_1}(x,z;\tau)$ is a smoother version of the usual kernel estimator of $E[H(Y+\tau)f(X)|x]$. Due to the use of the clever device of Härdle and Mammen (1993), a simple "outer" weighting function a will not suffice for our purpose. We need to use both an "outer" weighting function $\widetilde{a}=\widehat{f}^2_{h_1}$ and an "inner" weighting function $\widehat{f}_{h_2}$ in forming (2.9).

After some simple algebra, we have

$$\begin{aligned}\Gamma_{3n} &= \frac{1}{n(n-1)^2}\sum_{t_1=1}^{n}\int\left[\sum_{t_2\neq t_1}K_{1t_1t_2}\widehat{f}_{2,t_2}\left\{H(Y_{t_2}+\tau)-\widehat{m}_{h_2}(X_{t_2};\tau)\right\}\right]^2 dG(\tau)\\ &= \frac{1}{n(n-1)^2}\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\sum_{t_3\neq t_1}\int K_{1t_1t_2}K_{1t_1t_3}\widehat{f}_{2,t_2}\widehat{f}_{2,t_3}\left[H(Y_{t_2}+\tau)-\widehat{m}_{h_2}(X_{t_2};\tau)\right]\\ &\quad\times\left[H(Y_{t_3}+\tau)-\widehat{m}_{h_2}(X_{t_3};\tau)\right]dG(\tau). \end{aligned} \tag{2.12}$$

where $K_{1ts} \equiv K_{h_1}(X_t - X_s, Z_t - Z_s)$ and $\widehat{f}_{2,s} \equiv \widehat{f}_{h_2}(X_s)$. The above statistic is simple to compute and offers a natural way to test H_0. Nevertheless, we propose a bias-adjusted test statistic, namely

$$\Gamma_n \equiv \frac{n-1}{n-2}\left(\Gamma_{3n} - B_n\right), \tag{2.13}$$

where $B_n \equiv n^{-1}(n-1)^{-2}\sum_{t_1=1}^{n}\sum_{t_2\neq t_1}\int K^2_{1t_1t_2}\widehat{f}^2_{2,t_2}\left[H(Y_{t_2}+\tau)-\widehat{m}_{h_2}(X_{t_2};\tau)\right]^2 dG(\tau)$. In effect, our test statistic Γ_n removes all the "diagonal" ($t_2 = t_3$) terms from Γ_{3n} in (2.12), thus reducing the bias of the statistic. A similar idea has been used in Lavergne and Vuong (2000).

We will show that after being appropriately scaled, Γ_n is asymptotically normally distributed under suitable assumptions.

3 The Asymptotic Distribution of the Test Statistic

In this section we first focus on the case of a stochastic process that has an observable series of continuously-valued realizations. Cases for which a subset of the random vector $(X', Y', Z')'$ is discretely valued or unobserved are discussed at the end of this section.

3.1 Asymptotic Null Distribution

We work with the dependence notion of β-mixing. Let $\{V_t,\ t \geq 0\}$ be a strictly stationary stochastic process and F_s^t denote the sigma algebra generated by $(V_s, ..., V_t)$ for $s \leq t$. The process is called β-mixing or absolutely regular, if as $k \to \infty$,

$$\beta(k) \equiv \sup_{s\in\mathbb{N}} E\left[\sup_{A\epsilon\mathcal{F}_{s+k}^{\infty}}\left|P(A|\mathcal{F}_{-\infty}^{s}) - P(A)\right|\right] \to 0.$$

Our assumptions are as follows.

Assumption A.1 (Data Generating Process (DGP))

(i) $\{W_t \equiv (X_t', Y_t', Z_t')',\ t \geq 1\}$ is a strictly stationary absolutely regular process on $\mathbb{R}^{d_1+d_2+d_3} \equiv \mathbb{R}^d$ with mixing coefficients $\beta(k)$ that satisfy $\sum_{k=1}^{\infty} k^2 \beta^{\delta/(1+\delta)}(k) < \infty$ for some $0 < \delta \leq 1/3$.

(ii) $f(\cdot,\cdot,\cdot) \in \mathcal{G}_{v-1}^{\infty}$, $m(\cdot;\tau) \in \mathcal{G}_v^{2(1+\delta)}$, and $m(\cdot,\cdot;\tau) \in \mathcal{G}_v^{2(1+\delta)}$ for each $\tau \in \mathbb{R}^{d_2}$, where $v \geq 2$ is an integer and $\mathcal{G}_\mu^\alpha$ is a class of functions defined in Robinson (1988, p. 939). Furthermore, f and the $m's$ satisfy global Lipschitz conditions: $|f(w_0 + w) - f(w_0)| \leq D_f(w_0)||w||$, $|m(u_0 + u;\tau) - m(u_0;\tau)| \leq$

$D_m(u_0;\tau)||u||$ for $u=(x,z)$ or x, where $\int|D_f(w)|^{2(1+\delta)}dF(w)<\infty$, $\int|D_m(u;\tau)|^{2(1+\delta)}dF(u)dG(\tau)<\infty$, and $||\cdot||$ is the Euclidean norm.

(iii) For $1\leq l\leq 10$, the probability density function (pdf) $f_{t_1,...,t_l}$ of $(W_{t_1},...,W_{t_l})$ is bounded and satisfies a Lipschitz condition: $|f_{t_1,...,t_l}(w_1+u_1,\cdots,w_l+u_l)-f_{t_1,...,t_l}(w_1,\cdots,w_l)|\leq D_{t_1,...,t_l}(w_1,...,w_l)||u||$, where $u\equiv(u_1,...,u_l)$ and $D_{t_1,...,t_l}$ is integrable and satisfies the conditions that $\int_{\mathbb{R}^{dl}}D_{t_1,...,t_l}(w_1,...,w_l)$ $||w||^{2(1+\delta)}dw<\overline{M}<\infty$, and $\int_{\mathbb{R}^{dl}}D_{t_1,...,t_l}(w_1,...,w_l)\ f_{t_1,...,t_l}(w_1,\cdots,w_l)dw<\overline{M}<\infty$.

Assumption A.2 (Kernel and bandwidth)

(i) The kernel K is a product of a univariate kernel $k:\mathbb{R}\to\mathbb{R}$ such that $k(\cdot)$, $\int_{\mathbb{R}}u^ik(u)du=\delta_{i0}$ $(i=0,1,...,r-1)$, and $k(u)=O((1+|u|^{r+1+\epsilon})^{-1})$ for some $\epsilon>0$, where δ_{ij} is Kronecker's delta.

(ii) As $n\to\infty$, the bandwidth sequences h_1 and h_2 are such that $nh_1^{(d_1+d_3)/2}h_2^{2r}\to 0$, $nh_1^{2(d_1+d_3)}\to\infty$, and $h_1^{(d_1+d_3)}\ll h_2^{d_1}\ll h_1^{d_1}$.

Assumption A.3 (Weight functions)

(i) The weight function g_0 has full support on $\mathbb{R}^{d_2}$; is bounded, even, integrable, and everywhere positive; and is chosen such that its corresponding characteristic function H is real-valued and boundedly $(r+1)$- differentiable.

(ii) The weight function g is uniformly bounded, integrable, and nonnegative everywhere on $\mathbb{R}^{d_2}$.

Remarks. Assumption A.1(i) requires that $\{W_t\}$ be a stationary absolutely regular process with algebraic decay rate. This is standard for application of a central limit theorem for U-statistics for weakly dependent data (e.g., Tenreiro, 1997). A.1(ii) imposes smoothness and moment conditions on f and the $m's$. For instance, if μ is a positive integer, then $g\in\mathcal{G}_\mu^\alpha$ means that g is differentiable up to order μ, has Taylor expansion with the remainder satisfying a local Lipschitz condition, and g has finite αth moment. A.1(iii) imposes smoothness and moment conditions on $f_{t_1,...,t_l}$. Similar conditions are imposed in Li (1999). Assumption A.2(i) requires that the kernel be of second order or higher and it implies $\int_{\mathbb{R}}u^rk(u)du<\infty$. Unless $d_1+d_3=2$, a higher order kernel is needed, which is nevertheless common in the literature (e.g., Robinson (1988), Li (1999), Fan and Li (1999)). Assumption A.2(ii) specifies conditions on the choice of bandwidth sequences. Assumption A.3(i) is not as strict as it appears. The uniform boundedness of H comes free as one important property of characteristic functions. That H is real-valued and boundedly $(r+1)$- differentiable is also easily met in practice by choosing g_0 appropriately. For example, g_0 can be either a normal density function on $\mathbb{R}^{d_2}$, or a double exponential density function. A potential opportunity created by Assumption A.3 is to choose g_0 and g in applications so that any numerical integration can be done quickly or one can work out the integration analytically. We return to this point in Section 4.

Now let $V(x,z;\tau,\tau')\equiv\mathrm{cov}(H(Y+\tau),H(Y+\tau')\mid X=x,Z=z)$. Define $\sigma^2\equiv 2C_3^{(d_1+d_3)}\int\int\int V^2(x,z;\tau,\tau')$ $f^4(x)f^4(x,z)dG(\tau)dG(\tau')d(x,z)$, where $C_3\equiv\int_{\mathbb{R}}\left[\int_{\mathbb{R}}k(u+v)k(u)du\right]^2dv$. Our main result is

Theorem 3.1 *Under Assumptions A.1-A.3 and under H_0, $nh_1^{(d_1+d_3)/2}\Gamma_n\xrightarrow{d}N(0,\sigma^2)$.*

The proof is tedious and is relegated to Appendix A. To implement the test, we require a consistent estimate of the variance σ^2. Let $\widehat{\sigma}^2\equiv 2C_3^{(d_1+d_3)}n^{-2}\sum_{t=1}^n\sum_{s=1}^n\int\int\widehat{f}_{h_1}(X_t,Z_t)\widehat{f}_{h_1}(X_s,Z_s)\widehat{f}_{h_2}^2(X_t)$ $\widehat{f}_{h_2}^2(X_s)\widehat{\varepsilon}_t(\tau)\widehat{\varepsilon}_t(\tau')\widehat{\varepsilon}_s(\tau)\widehat{\varepsilon}_s(\tau')\,dG(\tau)dG(\tau')$, where $\widehat{\varepsilon}_t(\tau)\equiv H(Y_t+\tau)-\widehat{m}_{h_2}(X_t;\tau)$. It is easy to show that $\widehat{\sigma}^2$ is consistent for σ^2 under H_0. We then compare

$$T_n \equiv nh_1^{(d_1+d_3)/2}\Gamma_n/\sqrt{\widehat{\sigma}^2} \tag{3.1}$$

with the one-sided critical value z_α from the standard normal distribution, and reject the null when $T_n > z_\alpha$.

3.2 Asymptotic Local Power Properties

To examine the asymptotic local power property of our test, we let $f^{[n]}(x,y,z)$ denote a sequence of densities, $f^{[n]}(x,y) \equiv \int f^{[n]}(x,y,z)dz$, $f^{[n]}(x,z) \equiv \int f^{[n]}(x,y,z)dy$, and $f^{[n]}(x) \equiv \int f^{[n]}(x,y,z)dydz$. Assume that $||f^{[n]}(x,y,z) - f(x,y,z)||_\infty \to 0$ as $n \to \infty$. Let $\alpha_n \to 0$ as $n \to \infty$. Let E_n denote expectation under the law associated with $f^{[n]}$. Define $m^{[n]}(x,z;\tau) \equiv E_n[H(Y+\tau)|X=x, Z=z]$ and $m^{[n]}(x;\tau) \equiv E_n[H(Y+\tau)|X=x]$. Given our setup, local alternatives can be specified as

$$H_1(\alpha_n) : m^{[n]}(x,z;\tau) = m^{[n]}(x;\tau) + \alpha_n\Delta(x,z;\tau), \tag{3.2}$$

where $\Delta(x,z;\tau)$ satisfies

$$\gamma \equiv \int\int \triangle(x,z;\tau)^2 f^2(x)f^3(x,z)d(x,z)dG(\tau) < \infty.$$

The following proposition shows that our test can distinguish local alternatives $H_1(\alpha_n)$ at rate $\alpha_n = n^{-1/2}h_1^{-(d_1+d_3)/4}$ while maintaining a constant level of asymptotic power.

Proposition 3.2 *Under Assumptions A.1–A.3, suppose that* $\alpha_n = n^{-1/2}h_1^{-(d_1+d_3)/4}$ *in* $H_1(\alpha_n)$. *Then, the power of the test satisfies* $\Pr(T_n \geq z_\alpha|H_1(\ \alpha_n)) \to 1 - \Phi(z_\alpha - \gamma/\sigma)$.

3.3 Remarks

Theorem 3.1 covers the asymptotic null distribution of the test statistic when the null hypothesis involves a stochastic process that has observed continuously-valued realizations. While this case suffices for many empirical applications (e.g., a nonparametric test of Granger non-causality), our testing procedure is potentially applicable to a much wider range of situations. We now discuss several of these.

1. **Conditional independence test with unobservables.** When $W = (X', Y', Z')'$ has to be estimated from the data, two cases are possible. First, if W is estimated by using a finite-dimensional $\sqrt{n}$-consistent parameter estimator, one can show straightforwardly that the results in Theorem 3.1 and Proposition 3.2 continue to hold, and we say our test is "free of parameter estimation error". Second, when W is estimated nonparametrically, say by $\widehat{W}$, a sufficiently fast convergence rate is required. For brevity, we leave this for future research.

2. **Limited dependent variables and discrete conditioning variables.** As mentioned in the introduction, our test is also applicable to situations where not all variables in (X, Y, Z) are continuously valued. Although we have made reference to the joint density $f(x, y, z)$ to facilitate the presentation, there is no explicit use of the continuity of the random variable Y in our derivations. In particular,

the joint density $f(x, y, z)$ can be replaced everywhere by $f(x, z)dF(y|x, z)$ without changing any of the derivations. This is more than a superficial change, as it allows the application of our test to any situation involving discretely distributed variables. For example, Y may be a discrete response, or a more complicated censored or truncated version of a continuous (latent) variable. Also, one can treat a mixture of continuous and discrete conditioning variables with more complicated notation.

3. **Testing for independence.** It is possible to extend our procedure to the case where $d_1 = 0$, i.e., testing for independence between Y and Z. In this case, the null hypothesis reduces to

$$H_0^* : \Pr\{f(y|Z) = f(y)\} = 1 \ \forall y \in \mathbb{R}^{d_2}.$$

To test H_0^*, we can replace $\widehat{m}_{h_2}(X_t; \tau)$ in equations (2.12) and (2.13) by $\overline{H}(\tau) = n^{-1} \sum_{s=1}^{n} H(Y_s + \tau)$. One can readily modify the other assumptions in Section 3 and show easily that Theorem 3.1 and Proposition 3.2 continue to hold. For brevity, we don't repeat the argument.

4 Monte Carlo Experiments

In this section we report results of some Monte Carlo simulation experiments designed to examine the finite sample performance of our nonparametric conditional independence test. Specifically, we conduct simulation experiments focused on testing the order of nonlinear autoregressive (NLAR) processes. For each DGP under study, we standardize the data $\{(X_t, Y_t, Z_t),\ t = 1, ...n\}$ before implementing our test so that each variable has mean zero and variance one.

4.1 Motivation

During the last two decades, interest in nonlinear models in economics, econometrics and statistics has increased significantly. One area of wide interest is nonlinear time series model identification, and more specifically, lag selection. See Auestad and Tjostheim (1990), Cheng and Tong (1992), Tjostheim and Auestad (1994), Tschernig and Yang (2000), Finkenstädt et al. (2001), Lobato (2003), among others. These methods investigate the order d of a strictly stationary β-mixing univariate autoregressive time series model of the form

$$Y_t = g(Y_{t-1}, Y_{t-2}, ..., Y_{t-d}, \varepsilon_t), \tag{4.1}$$

where the function g is unknown and $\{\varepsilon_t\}$ is a noise process.

In contrast, our theory pertains to the entire conditional distribution, not just the conditional location or conditional standard deviation. As before, let $f(\cdot|\cdot)$ be the conditional density of one random variable given another. The null of interest is

$$H_0(d) : f(Y_t|Y_{t-1}, ..., Y_{t-d-1}) = f(Y_t|Y_{t-1}, ..., Y_{t-d}), \tag{4.2}$$

i.e., conditioning on $(Y_{t-1}, ..., Y_{t-d})$, the random variable Y_{t-d-1} has no explanatory power for Y_t. If d^* is the minimum of d such that (4.2) is true, we say the nonlinear time series is of order d^*. In the following, we write $H_0(d) : d^* = d$ to represent (4.2). In the special case when $d = 0$, the test reduces to a test of serial independence of first order: $H_0(0) : f(Y_t|Y_{t-1}) = f(Y_t)$.

4.2 Simulation Design and Practical Issues

We consider the following DGPs in our Monte Carlo study.

DGP1: $Y_t = 0.3Y_{t-1} + \varepsilon_t$;

DGP2: $Y_t = (-0.5Y_{t-1} + \varepsilon_t)1(Y_{t-1} \leq 1) + (0.4Y_{t-1} + \varepsilon_t)1(Y_{t-1} > 1)$;

DGP3: $Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$;

DGP4: $Y_t = 0.6\Phi(Y_{t-1})Y_{t-1} + \varepsilon_t$, where Φ represents the cumulative distribution of a standard normal distribution;

DGP5: $Y_t = -0.5Y_{t-1} + 0.5Y_{t-2}\left\{1 + \exp(-0.5Y_{t-1})\right\}^{-1} + \varepsilon_t$;

DGP6: $Y_t = 0.1\log(Y_{t-1}^2) + \sqrt{0.1 + 0.9Y_{t-2}^2}\varepsilon_t$;

DGP7: $Y_t = \exp(-Y_{t-1}^2) + |0.1Y_{t-2}(16 - Y_{t-2})|\varepsilon_t$;

DGP8: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \sqrt{0.3 + |Y_{t-3}|}\varepsilon_t$;

DGP9: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.8Y_{t-1}^2 + 0.64Y_{t-2}^2 + 0.512Y_{t-3}^2$;

DGP10: $Y_t = \sqrt{h_t}\varepsilon_t$, $h_t = 0.01 + 0.8h_{t-1} + 0.15Y_{t-1}^2$;

where $\{\varepsilon_t\}$ are i.i.d. $N(0,1)$ in DGPs 1-5 and 8-10, they are the i.i.d. sum of 30 uniformly independently distributed random variables each over the range [-0.1, 0.1] in DGP 6, and the i.i.d. sum of 10 uniformly independently distributed random variables each over the range [-1/7, 1/7] in DGP 7. DGPs 1 through 3 are studied in Hong and White (2005) in testing for serial independence. DGPs 4 and 5 are studied in Lobato (2003) in testing for nonlinear autoregression. DGPs 6-7 are used in Finkenstädt et al. (2001) in determining the order of nonlinear time series. Clearly, DGPs 1-4 are of order 1, DGPs 5-7 are of order 2, and DGPs 8–10 are of order 3 or higher. Note that all DGPs except DGP 1 are nonlinear in the mean or in the variance or in both.

We test for $H_0(d) : d^* = d$, where $d = 1$ or 2. We use a fourth order kernel in estimating all required quantities: $k(u) = (3 - u^2)\varphi(u)/2$, where $\varphi(u)$ is the pdf of the standard normal distribution. We choose both $g_0(\cdot)$ and $g(\cdot)$ (see Assumption A.3) to be a standard normal pdf. For this particular g_0, the corresponding characteristic function $H(y) \equiv \int e^{iuy}dG_0(u)$ has the simple form $H(y) = \exp(-y^2/2)$. Given our choice of g_0 and g, we can work out the integration analytically so that no numerical integration over $dG(\tau)$ is required.

Since we have two bandwidth parameters to choose, h_1 and h_2, and it is difficult to pin down the optimal bandwidth sequences, we choose h_1 and h_2 separately by cross validation in our simulation. Specifically, we set

$$h_1 = h_1^* n^{\frac{1}{8+d_{13}}} n^{-\frac{1}{4+d_{13}}} \text{ and } h_2 = h_2^* n^{\frac{1}{8+d_1}} n^{-\frac{1}{4+d_1}}, \tag{4.3}$$

where h_1^* and h_2^* are the least-squares cross-validated bandwidths for estimating the conditional expectation of Y_t given (X_t, Z_t) and X_t, respectively. Note that given the fourth order kernel we use, h_1^* and h_2^* converge at rates $n^{-1/(8+d_{13})}$ and $n^{-1/(8+d_1)}$, respectively. Undersmoothing is required for our test. We use Lee (2003, p. 16) to adjust h_1^* and h_2^* appropriately in (4.3) to make sure Assumption A2 is met.

It is well known that a nonparametric test that relies on the asymptotic normal approximation may perform poorly in finite samples. An alternative approach is to use bootstrap approximation. Based upon Paparoditis and Politis's (2000) local bootstrap procedure, SW propose a smoothed local bootstrap procedure to obtain the bootstrap data $\{X_t^*, Y_t^*, Z_t^*\}$. In the following we follow SW's method to obtain

Table 1: Empirical rejection frequency of the tests

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9	DGP10
$H_0\,(1)$										
n=100										
5%	0.050	0.065	0.060	0.055	0.305	0.300	0.160	0.410	0.865	0.230
10%	0.095	0.120	0.095	0.090	0.405	0.465	0.240	0.600	0.910	0.310
n=200										
5%	0.070	0.060	0.055	0.045	0.450	0.400	0.210	0.765	0.910	0.275
10%	0.115	0.105	0.110	0.100	0.600	0.610	0.385	0.910	0.945	0.410
$H_0\,(2)$										
n=100										
5%	0.040	0.050	0.060	0.035	0.055	0.065	0.030	0.395	0.655	0.230
10%	0.105	0.110	0.095	0.075	0.115	0.125	0.050	0.615	0.750	0.300
n=200										
5%	0.045	0.050	0.055	0.030	0.045	0.070	0.035	0.700	0.780	0.325
10%	0.100	0.120	0.100	0.070	0.120	0.125	0.065	0.875	0.830	0.485

Note: DGPs 1-4 satisfy $H_0\,(1)$ and DGPs 1-7 satisfy $H_0\,(2)$, whereas the other DGPs satisfy neither $H_0\,(1)$ nor $H_0\,(2)$.

the bootstrap resamples. One can follow SW to verify the validity of their bootstrap method in our framework.

4.3 Results

Table 1 reports the empirical rejection frequency of the 5% and 10% test for $H_0\,(1)$ and $H_0\,(2)$. For brevity we only study sample sizes $n = 100$ and 200; we use 200 Monte Carlo replications for each experiment. The number of bootstrap resamples is also set to 200 for each scenario. From Table 1 we see that the size of our test is well behaved in that most of the empirical frequencies are close to the nominal significance level when the null hypothesis is true. The test has reasonable power when the null hypothesis is not true. For example, in testing $H_0\,(1)$, the 10% test powers for both DGP8 and DGP9 are above 0.90 for as small a sample as 200. Similarly, in testing $H_0\,(2)$, the 10% test powers for both DGP8 and DGP9 are above 0.80 for as small a sample as 200. We view a sample of 200 as small, given the fact that densities of dimension two or three must be estimated in constructing the test.

5 Application to Stock Market Data

Although many studies conducted during the 1980s and 1990s report that financial time series such as exchange rates and stock prices exhibit nonlinear dependence (e.g., Hsieh, 1989; Sheedy 1998), researchers often neglect this when they test for possible dependence. As documented by Hiemstra and Jones (1994), all prior studies of causal relationship rely exclusively on the traditional linear Granger causality test, which unfortunately may have little power in detecting nonlinear relationships. In this section, we use both our test and a traditional linear Granger causality test to study the dynamic linkage between three US stock market price indices (Dow Jones 65 composite, Nasdaq, and S&P 500) and the trading volumes on the New York Stock Exchange (NYSE), Nasdaq, and NYSE markets, respectively.

We obtain daily data for the three major stock market price indices and trading volumes from Yahoo Finance for the sample period from January 3rd, 2000 to January 10th, 2003. After excluding weekends and holidays, the total numbers of observations are 759 for the Dow Jones 65 composite and Nasdaq series and 761 for the S&P 500 series. Following the literature, we let P_t and V_t stand for the natural logarithm of stock price indices and volumes multiplied by 100, respectively.

We first employ the augmented Dickey-Fuller test to check for stationarity of $\{P_t\}$ and $\{V_t\}$. The test results indicate that there is a unit root in all level series but not in the first differenced series. Therefore, both Granger causality tests will be conducted on the first differenced data, which we denote as ΔP_t and ΔV_t below. Next, we employ Johansen's likelihood ratio method to examine whether P_t and V_t are cointegrated or not. We find no evidence of cointegration. Consequently, we include no error correction term in our linear Granger causality test.

For the linear Granger non-causality tests, we are interested in whether ΔP_t and ΔV_t Granger-cause each other linearly. For example, in testing whether ΔP_t Granger causes ΔV_t linearly, one would typically check if the null hypothesis $H_{0,L} : \beta_1 = ... = \beta_{L_x} = 0$ holds with

$$\Delta V_t = \alpha_0 + \alpha_1 \Delta V_{t-1} + ... + \alpha_{L_v} \Delta V_{t-L_v} + \beta_1 \Delta P_{t-1} + ...\beta_{L_p} \Delta P_{t-L_p} + \epsilon_t, \tag{5.1}$$

where $\epsilon_t \sim i.i.d.(0, \sigma^2)$ under $H_{0,L}$. Nevertheless, to permit a direct comparison with our nonparametric test for nonlinear Granger causality, we focus on testing $H^*_{0,L} : \beta = 0$ in

$$\Delta V_t = \alpha_0 + \alpha_1 \Delta V_{t-1} + ... + \alpha_{L_v} \Delta V_{t-L_v} + \beta \Delta P_{t-i} + \epsilon_t, \ i = 1, ..., L_p. \tag{5.2}$$

To implement our nonparametric test, we set all smoothing parameters according to those used in the simulations in the last section. To mitigate the curse of dimensionality, we focus on testing

$$H^*_{0,NL} : \Pr(f(\Delta V_t | \Delta V_{t-1}, ..., \Delta V_{t-L_v}; \Delta P_{t-i}) = f(\Delta V_t | \Delta V_{t-1}, ..., \Delta V_{t-L_v})) = 1, i = 1, ..., L_p. \tag{5.3}$$

in checking the Granger causal direction from ΔP_t to ΔV_t, and similarly for the reverse direction.

The results of linear and nonlinear Granger causality tests between ΔP_t and ΔV_t are given in Table 2, where we choose L_v and L_p to be 1, 2 or 3. For example, when L_v is 1, we also choose L_p to be 1 so that we only check whether ΔP_{t-1} should enter (5.2) or not. This corresponds to the first row in each panel of Table 2. When L_v is 2, we choose L_p to be 2. In this case, we check whether ΔP_{t-1} or ΔP_{t-2} (but not both) should enter (5.2) or not, which corresponds to the second and third rows in each panel of Table 2. The case for $L_v = 3$ is done analogously, corresponding to the fourth to sixth rows in each panel of Table 2. The case for testing whether ΔV_t Granger causes ΔP_t is done similarly.

The results of the linear Granger causality test between stock prices and volumes are given in Panel A of Table 2. At all levels of L_v, we find causal links from stock prices to trading volumes for the Nasdaq

Table 2: Granger non-causality tests between stock prices and trading volumes

Panel A: Linear Granger non-causality test between $\triangle P$ and $\triangle V$

H_0 : $\triangle P$ does not Granger cause $\triangle V$				H_0 : $\triangle V$ does not Granger cause $\triangle P$			
	Dow Jones	Nasdaq	S&P 500		Dow Jones	Nasdaq	S&P 500
L_v=1, ΔP_{t-1}	0.910	0.001	0.007	L_p=1, ΔV_{t-1}	0.211	0.953	0.979
L_v=2, ΔP_{t-1}	0.504	0.002	0.005	L_p=2, ΔV_{t-1}	0.209	0.812	0.871
L_v=2, ΔP_{t-2}	0.369	0.011	0.018	L_p=2, ΔV_{t-2}	0.957	0.564	0.758
L_v=3, ΔP_{t-1}	0.374	0.004	0.004	L_p=3, ΔV_{t-1}	0.210	0.816	0.855
L_v=3, ΔP_{t-2}	0.201	0.008	0.008	L_p=3, ΔV_{t-2}	0.970	0.591	0.789
L_v=3, ΔP_{t-3}	0.231	0.719	0.241	L_p=3, ΔV_{t-3}	0.969	0.662	0.983

Panel B: Nonlinear Granger non-causality test between $\triangle P$ and $\triangle V$

H_0 : $\triangle P$ does not Granger cause $\triangle V$				H_0 : $\triangle V$ does not Granger cause $\triangle P$			
	Dow Jones	Nasdaq	S&P 500		Dow Jones	Nasdaq	S&P 500
L_v=1, ΔP_{t-1}	0.045	0	0.015	L_p=1, ΔV_{t-1}	0.125	0.370	0.380
L_v=2, ΔP_{t-1}	0	0	0	L_p=2, ΔV_{t-1}	0.060	0.340	0.420
L_v=2, ΔP_{t-2}	0	0	0	L_p=2, ΔV_{t-2}	0.055	0.355	0.665
L_v=3, ΔP_{t-1}	0.020	0.010	0.020	L_p=3, ΔV_{t-1}	0.280	0.490	0.545
L_v=3, ΔP_{t-2}	0.005	0.025	0.015	L_p=3, ΔV_{t-2}	0.240	0.460	0.605
L_v=3, ΔP_{t-3}	0.020	0.005	0.030	L_p=3, ΔV_{t-3}	0.205	0.475	0.550

Note: Numbers in the main entries are the p-values. For the nonlinear Granger non-causality test, the number of bootstrap resamples is $B = 200$ in each case.

and S&P 500 data but not for the Dow Jones at the 5% nominal significance level. Unambiguously, we find no Granger causality from trading volume to stock price using the linear causality test.

The results for our nonparametric test are reported in Panel B of Table 2. From Panel B, we see that at the 5% nominal significance level, stock prices lead trading volumes for all three datasets and this is true at all lags of our study. Further, like the linear Granger causality test results, our nonparametric test results find no evidence of Granger causality from trading volumes to stock prices.

6 Concluding remarks

This paper develops asymptotic distribution theory for a consistent nonparametric conditional independence test. It is based upon properties of the conditional characteristic functions and transforms the notion of conditional independence into the equivalence of two infinite collections of conditional moment restrictions. Together with the previous work of SW, this addresses the long standing need in econometrics for an asymptotic theory for a practical and powerful nonparametric test for conditional independence. We extend the test of SW in two directions: our test is less sensitive to the choice of bandwidth, and it has power in detecting deviations from conditional independence in the full support of the density.

To improve the asymptotic approximation to the finite sample distribution of the test statistic, one could consider higher order refinements, which may offer a solution to the choice of optimal bandwidth. However, it is well known that estimation of higher order refinements is tedious and may not necessarily provide a sufficiently good approximation in finite samples. Another topic not addressed here, and a suitable subject for future research, is the optimality of the test.

Acknowledgements

We gratefully thank Graham Elliott, Clive W. J. Granger, Peter Robinson, an associate editor, and two anonymous referees for very constructive comments on an earlier version of this paper. The first author gratefully acknowledges financial support from the NSFC under the grant number 70501001. The usual disclaimers apply.

Appendix

Throughout this appendix, C is a generic constant that may vary from case to case. Denote $W_t \equiv (X_t', Y_t, Z_t')'$, $f_{1t} \equiv f(X_t, Z_t)$, $\widehat{f}_{1t} \equiv \widehat{f}_{h_1}(X_t, Z_t)$, $f_{2t} \equiv f(X_t)$, $\widehat{f}_{2t} \equiv \widehat{f}_{h_2}(X_t)$, $K_{1ts} \equiv K_{h_1}(X_t - X_s, Z_t - Z_s)$, $K_{2ts} \equiv K_{h_2}(X_t - X_s)$, $K_{(x,z),t} \equiv K_{h_1}(x - X_t, z - Z_t)$, $K_{x,t} \equiv K_{h_2}(x - X_t)$, and $d_{13} \equiv d_1 + d_3$. Let

$$\sum_{t \neq s} = \sum_{s=1}^{n} \sum_{t=1, t \neq s}^{n}, \quad \sum_{t_1 \neq t_2 \neq t_3} \equiv \sum_{t_1=1}^{n} \sum_{t_2=1, t_2 \neq t_1}^{n} \sum_{t_3=1, t_3 \neq t_1, t_3 \neq t_2}^{n},$$

$$\sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2} \equiv \sum_{t_1 \neq t_2 \neq t_3} \sum_{t_4=1, t_4 \neq t_2}^{n}, \text{ and } \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2, t_5 \neq t_3} \equiv \sum_{t_1 \neq t_2 \neq t_3} \sum_{t_4=1, t_4 \neq t_2}^{n} \sum_{t_5=1, t_5 \neq t_3}^{n}.$$

Let $n_3 \equiv n(n-1)(n-2)$, $n_4 = n_3(n-1)$, and $n_5 = n_3(n-1)^2$. Let $E_s[K_{2ts}] \equiv \int K_{h_2}(X_t - x_s)dF(x_s)$. Further, let the bar notation denote an *i.i.d.* process. For example, $\{\overline{W}_t,\ t \geq 0\}$ is an *i.i.d.* sequence having the same marginal distributions as $\{W_t,\ t \geq 0\}$.

A Proofs

Proof of Lemma 2.1. The "if" part is trivial. Now suppose that $\int \psi(u; x, z)e^{i\tau' u}dG_0(u) = 0$ a.e.-F on $\mathbb{R}^{d_{13}}$ for every $\tau \in \mathbb{R}^{d_2}$; we follow Bierens (1982) closely to show that $\psi(u; x, z) = 0$ a.e.-$G_0 \times F$ on $\mathbb{R}^{d_2} \times \mathbb{R}^{d_{13}}$.

Denote $\mathrm{Re}(\psi)$ and $\mathrm{Im}(\psi)$ as the real and imaginary part of ψ respectively. Put $\psi_1(\cdot) = \max(\mathrm{Re}(\psi(\cdot)), 0)$, $\psi_2(\cdot) = \max(-\mathrm{Re}(\psi(\cdot)), 0)$, $\psi_3(\cdot) = \max(\mathrm{Im}(\psi(\cdot)), 0)$, and $\psi_4(\cdot) = \max(-\mathrm{Im}(\psi(\cdot)), 0)$. Then obviously ψ_j, $j = 1, ..., 4$, are nonnegative Borel measurable real functions on $\mathbb{R}^d$ satisfying $\mathrm{Re}(\psi) = \psi_1 - \psi_2$ and $\mathrm{Im}(\psi) = \psi_3 - \psi_4$.

Now assume for the moment that $c_j = \int \psi_j(u; x, z)dG_0(u) > 0$ for $j = 1, ..., 4$. We define four conditional probability measures

$$v_j(B; x, z) = \int_B \psi_j(u; x, z)dG_0(u)/c_j,\ j = 1, ..., 4, \text{ where } B \text{ is a Borel set on } \mathbb{R}^{d_2}. \tag{A.1}$$

Writing $dv_j(u; x, z) \equiv v_j(du; x, z)$ for $j = 1, 2, 3$, and 4, we have

$$\begin{aligned}
&\int \psi(u; x, z)e^{i\tau' u}dG_0(u)\\
&= \left[\int \psi_1(u; x, z)e^{i\tau' u}dG_0(u) - \int \psi_2(u; x, z)e^{i\tau' u}dG_0(u)\right] + i\ \left[\int \psi_3(u; x, z)e^{i\tau' u}dG_0(u) - \int \psi_4(u; x, z)\right.\\
&\left.\times e^{i\tau' u}dG_0(u)\right]\\
&= \left[c_1 \int e^{i\tau' u}dv_1(u; x, z) - c_2 \int e^{i\tau' u}dv_2(u; x, z)\right] + i\ \left[c_3 \int e^{i\tau' u}dv_3(u; x, z) - c_4 \int e^{i\tau' u}dv_4(u; x, z)\right]\\
&= [c_1\eta_1(\tau; x, z) - c_2\eta_2(\tau; x, z)] + i\ [c_3\eta_3(\tau; x, z) - c_4\eta_4(\tau; x, z)],
\end{aligned}$$

where $\eta_j(\tau; x, z) \equiv \int e^{i\tau' u}dv_j(u; x, z)$, $j = 1, ..., 4$, are conditional characteristic functions of the conditional probability measures v_j respectively.

If $\int \psi(u; x, z)e^{i\tau' u}dG_0(u) = 0$ a.e.-F on $\mathbb{R}^{d_{13}}$ for every $\tau \in \mathbb{R}^{d_2}$, $c_1\eta_1(\tau; x, z) = c_2\eta_2(\tau; x, z)$ and $c_3\eta_3(\tau; x, z) = c_4\eta_4(\tau; x, z)$ a.e.-(x, z) for every $\tau \in \mathbb{R}^{d_2}$. Note that $\eta_1(0; x, z) = \eta_2(0; x, z) = \eta_3(0; x, z) = \eta_4(0; x, z) = 1$, so

$$c_1 = c_2,\ c_3 = c_4, \tag{A.2}$$

and

$$\eta_1(\tau; x, z) = \eta_2(\tau; x, z) \text{ and } \eta_3(\tau; x, z) = \eta_4(\tau; x, z) \text{ a.e.-}F \text{ on } \mathbb{R}^{d_{13}} \text{ for every } \tau \in \mathbb{R}^{d_2}. \tag{A.3}$$

Consequently, for every Borel set B on $\mathbb{R}^{d_2}$, we have

$$v_1(B; x, z) = v_2(B; x, z) \text{ and } v_3(B; x, z) = v_4(B; x, z) \text{ a.e.-}F \text{ on } \mathbb{R}^{d_{13}}.$$

From (A.1), (A.2) and (A.3), we obtain that for every Borel set B on $\mathbb{R}^{d_2}$,

$$\int_B \psi_1(u; x, z)dG_0(u) = \int_B \psi_2(u; x, z)dG_0(u),$$

$$\int_B \psi_3(u;x,z)dG_0(u) = \int_B \psi_4(u;x,z)dG_0(u),$$

and consequently,

$$\int_B \psi(u;x,z)dG_0(u) = 0.$$

Note that $B_1 \equiv \{u \in \mathbb{R}^{d_2} : \text{Re}(\ \psi(u;x,z)) > 0\}$ is a Borel set, and $\int_{B_1} \text{Re}(\ \psi(u;x,z))dG_0(u) = 0$, which is only possible if B_1 is a null set with respect to $dG_0(u)$ a.e.-F on $\mathbb{R}^{d_{13}}$. Similarly, one concludes that the Borel sets $B_2 \equiv \{u \in \mathbb{R}^{d_2} : \text{Re}(\psi(u;x,z)) < 0\}$, $B_3 \equiv \{u \in \mathbb{R}^{d_2} : \text{Im}(\psi(u;x,z)) > 0\}$ and $B_4 \equiv \{u \in \mathbb{R}^{d_2} : \text{Im}(\psi(u;x,z)) < 0\}$ are all null sets with respect to $dG_0(u)$ a.e.-F on $\mathbb{R}^{d_{13}}$. Hence, $\cup_{i=1}^4 B_i = \{u \in \mathbb{R}^{d_2} : \psi(u;x,z) \neq 0\}$ is a null set with respect to $dG_0(u)$ a.e.$-F$ on $\mathbb{R}^{d_{13}}$. This means $\psi(u;x,z) = 0$ a.e.$-G_0 \times F$ on $\mathbb{R}^{d_2} \times \mathbb{R}^{d_{13}}$. If $c_j = \int \psi_j(u;x,z)dG_0(u) = 0$ for some $j \in \{1,2,3,4\}$, our conclusion still holds as an easy exercise. This completes the "only if" part of Lemma 2.1. ■

Proof of Theorem 3.1. Let $\varepsilon_t(\tau) \equiv H(Y_t+\tau) - m(X_t, Z_t;\tau)$, and $\widehat{e}_t(\tau) \equiv m(X_t;\tau) - \widehat{m}_{h_2}(X_t;\tau)$. Under H_0, $H(Y_t+\tau) - \widehat{m}_{h_2}(X_t;\tau) = \varepsilon_t(\tau) + \widehat{e}_t(\tau)$, and from equations (2.12) and (2.13), we have

$$\begin{aligned}\Gamma_n &= \frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\widehat{f}_{2,t_2}\widehat{f}_{2,t_3}\left[H(Y_{t_2}+\tau) - \widehat{m}_{h_2}(X_{t_2};\tau)\right] \\ &\qquad\times\left[H(Y_{t_3}+\tau) - \widehat{m}_{h_2}(X_{t_3};\tau)\right]dG(\tau) \\ &= \frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\widehat{f}_{2,t_2}\widehat{f}_{2,t_3}\left\{\varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) + 2\varepsilon_{t_2}(\tau)\widehat{e}_{t_3}(\tau) + \widehat{e}_{t_2}(\tau)\widehat{e}_{t_3}(\tau)\right\}dG(\tau) \\ &\equiv \Gamma_{n1} + 2\Gamma_{n2} + \Gamma_{n3}.\end{aligned}$$

We complete the proof of Theorem 3.1 by showing that $nh_1^{d_{13}/2}\Gamma_{n1} \xrightarrow{d} N(0,\sigma^2)$, and $\Gamma_{ni} = o_p(n^{-1}h_1^{-d_{13}/2})$ for $i = 2$, and 3. These results are established in Lemmas A.1 to A.3. ■

Lemma A.1 $nh_1^{d_{13}/2}\Gamma_{n1} \xrightarrow{d} N(0,\sigma^2)$.

Proof. First, write

$$\begin{aligned}\Gamma_{n1} &= \frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}f_{2t_2}f_{2t_3}\varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau)dG(\tau) \\ &\quad+\frac{2}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\left(\widehat{f}_{2,t_2} - f_{2t_2}\right)f_{2t_3}\varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau)dG(\tau) \\ &\quad+\frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\left(\widehat{f}_{2,t_2} - f_{2t_2}\right)\left(\widehat{f}_{2,t_3} - f_{2t_3}\right)\varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau)dG(\tau) \\ &\equiv \Gamma_{n11} + 2B_{n1} + B_{n2}.\end{aligned}$$

By Lemma B.1, $B_{n1} = o_p(n^{-1}h_1^{-d_{13}/2})$. By Lemma B.2, $B_{n2} = o_p(n^{-1}h_1^{-d_{13}/2})$. Recall $W_t = (X_t', Y_t', Z_t')'$. Let $\phi_{ts} \equiv \phi(W_t, W_s) \equiv h_1^{d_{13}/2}\int\int K_{(x,z)t}K_{(x,z)s}f_{2t}f_{2s}\varepsilon_t(\tau)\varepsilon_s(\tau)dG(\tau)dF(x,z)$, and $nh_1^{d_{13}/2}\ \widetilde{\Gamma}_{n11} \equiv 2(n-1)^{-1}\sum_{1\leq t<s\leq n}\phi_{ts}$. By Lemma B.5, $\Gamma_{n11} = \widetilde{\Gamma}_{n11} + o_p(n^{-1}h_1^{-d_{13}/2})$. So it suffices to show $nh_1^{d_{13}/2}\widetilde{\Gamma}_{n11} \xrightarrow{d} N(0,\sigma^2)$.

Clearly, ϕ is symmetric in its argument, and $E\left[\phi(w, W_s)\right] = E\left[\phi(W_s, w)\right] = 0$. Now $nh_1^{d_{13}/2}\widetilde{\Gamma}_{n11}$ is a second order degenerate U-statistic. As in the proof of Lemma B.4 of Su and White (2005), it is easy to verify that Conditions (iii)-(vii) in Theorem 1 of Tenreiro (1997) are satisfied, so that a central limit theorem applies to $nh_1^{d_{13}/2}\widetilde{\Gamma}_{n11}$. The asymptotic variance is given by $\sigma^2 \equiv \underset{n\to\infty}{\text{p}\lim} 2E\left[\phi(\overline{W}_1, \overline{W}_2)^2\right] = 2C_3^{d_{13}} \int\int\int V^2(x,z;\tau,\tau')f^4(x,z)f^4(x)dG(\tau)dG(\tau')d(x,z)$, where $C_3 \equiv \int_{\mathbb{R}}\left[\int_{\mathbb{R}} k(u+v)k(u)du\right]^2 dv$. The proof of Lemma A.1 is complete. ■

Lemma A.2 $\Gamma_{n2} = o_p(n^{-1}h_1^{-d_{13}/2})$.

Proof. Write

$$\begin{aligned}\Gamma_{n2} &= \frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}f_{2t_2}\varepsilon_{t_2}(\tau)\widehat{f}_{2,t_3}\widehat{e}_{t_3}(\tau)dG(\tau)\\ &\quad+\frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\left(\widehat{f}_{2t_2}-f_{2t_2}\right)\varepsilon_{t_2}(\tau)\widehat{f}_{2,t_3}\widehat{e}_{t_3}(\tau)dG(\tau)\\ &\equiv B_{n3}+B_{n4}.\end{aligned}$$

By Lemmas B.3 and B.4, $B_{ni} = o_p(n^{-1}h_1^{-d_{13}/2})$, $i = 3$ and 4. ■

Lemma A.3 $\Gamma_{n3} = o_p(n^{-1}h_1^{-d_{13}/2})$.

Proof. Noting that $\widehat{f}_{2,t}\widehat{e}_t(\tau) = (n-1)^{-1}\sum_{s=1,s\neq t}^n K_{2ts}\left[m(X_t;\tau) - H(Y_s+\tau)\right]$, we have

$$\begin{aligned}\Gamma_{n3} &= \frac{1}{n_5}\sum_{\substack{t_1\neq t_2\neq t_3,\\ t_4\neq t_2, t_5\neq t_3}}\int K_{1t_1t_2}K_{1t_1t_3}K_{2t_2t_4}K_{2t_3t_5}\left[H(Y_{t_4}+\tau)-m(X_{t_2};\tau)\right]\left[H(Y_{t_5}+\tau)-m(X_{t_3};\tau)\right]dG(\tau)\\ &= B_{n5}+2B_{n6}+B_{n7},\end{aligned}$$

where

$$B_{n5} = \frac{1}{n_5}\sum_{\substack{t_1\neq t_2\neq t_3,\\ t_4\neq t_2, t_5\neq t_3}}\int K_{1t_1t_2}K_{1t_1t_3}K_{2t_2t_4}K_{2t_3t_5}\left[H(Y_{t_4}+\tau)-m(X_{t_4};\tau)\right]\left[H(Y_{t_5}+\tau)-m(X_{t_5};\tau)\right]dG(\tau),$$

$$B_{n6} = \frac{1}{n_5}\sum_{\substack{t_1\neq t_2\neq t_3,\\ t_4\neq t_2, t_5\neq t_3}}\int K_{1t_1t_2}K_{1t_1t_3}K_{2t_2t_4}K_{2t_3t_5}\left[H(Y_{t_4}+\tau)-m(X_{t_4};\tau)\right]\left[m(X_{t_5};\tau)-m(X_{t_3};\tau)\right]dG(\tau),$$

and

$$B_{n7} = \frac{1}{n_5}\sum_{\substack{t_1\neq t_2\neq t_3,\\ t_4\neq t_2, t_5\neq t_3}}\int K_{1t_1t_2}K_{1t_1t_3}K_{2t_2t_4}K_{2t_3t_5}\left[m(X_{t_4};\tau)-m(X_{t_2};\tau)\right]\left[m(X_{t_5};\tau)-m(X_{t_3};\tau)\right]dG(\tau).$$

Noting that $E\left[H\left(Y_t+\tau\right)|X_t\right] = m\left(X_t;\tau\right)$, one can follow the proof of Lemma B.2 to get $B_{ni} = o_p(n^{-1}h_1^{-d_{13}/2})$, $i = 5$, 6, and 7. ■

Proof of Proposition 3.2. Let $\triangle_t(\tau) \equiv \triangle(X_t, Z_t;\tau)$. Using the fact that $H(Y_{t_2}+\tau)-\widehat{m}_{h_2}(X_{t_2};\tau) = \varepsilon_t(\tau)+\alpha_n\triangle_t(\tau)+\widehat{e}_t(\tau)$ under $H_1(\alpha_n)$, we have from equations (2.12) and (2.13) that

$$\begin{aligned}\Gamma_n &= \frac{1}{n_3}\sum_{t_1\neq t_2\neq t_3}\int K_{1t_1t_2}K_{1t_1t_3}\widehat{f}_{2,t_2}\widehat{f}_{2,t_3}\{\varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau)+2\varepsilon_{t_2}(\tau)\widehat{e}_{t_3}(\tau)+\widehat{e}_{t_2}(\tau)\widehat{e}_{t_3}(\tau)\\ &\qquad+\alpha_n^2\triangle_{t_2}(\tau)\triangle_{t_3}(\tau)+2\alpha_n\varepsilon_{t_2}(\tau)\triangle_{t_3}(\tau)+2\alpha_n\triangle_{t_2}(\tau)\widehat{e}_{t_3}(\tau)\}\,dG(\tau)\\ &\equiv \Gamma_{n1}+2\Gamma_{n2}+\Gamma_{n3}+\Gamma_{n4}+2\Gamma_{n5}+2\Gamma_{n6},\end{aligned}$$

where Γ_{ni}, $i = 1, 2, 3$, are as defined in the proof of Theorem 3.1. It is straightforward to show that for $\alpha_n = n^{-1/2}h_1^{-d_{13}/4}$, $nh_1^{d_{13}}\Gamma_{n4} \to^p \gamma \equiv \int\int \triangle(x,z;\tau)^2 f^2(x) f^3(x,z) d(x,z) dG(\tau)$, $\Gamma_{ni} = o_p(n^{-1}h_1^{-d_{13}/2})$, $i = 5$ and 6. Also, $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ under $H_1(n^{-1/2}h_1^{-d_{13}/4})$. Consequently, $\Pr(T_n \geq z_\alpha | H_1(\alpha_n)) \to 1 - \Phi(z_\alpha - \gamma/\sigma))$. ■

B Some Useful Lemmas

Let $0 < \delta \leq 1/3$ be as defined in Assumption A.1(i). Below we frequently use the facts that (1) $\delta/(1+\delta) \leq 1/4$ and $(2+4\delta)/(1+\delta) \leq 5/2$; and (2) $|E[K_{h_2}(x - X_t) - f(x)]| \leq h_2^r G_f(x)$ by Lemma 4 of Robinson (1988). To save space, we denote $v_t \equiv (X_t', Y_t, Z_t')'$.

Lemma B.1 $B_{n1} \equiv \frac{1}{n_3}\sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left(\widehat{f}_{2,t_2} - f_{2t_2}\right) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau) = o_p(n^{-1}h_1^{-d_{13}/2})$.

Proof. Write $B_{n1} = B_{n11} + B_{n12}$, where

$$B_{n11} = \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left(E_{t_0}\left[K_{2t_2t_0}\right] - f_{2t_2}\right) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau),$$

and

$$B_{n12} = \frac{1}{n_4} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2} \int K_{1t_1t_2} K_{1t_1t_3} \left(K_{2t_2t_4} - E_{t_4}\left[K_{2t_2t_4}\right]\right) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau).$$

First, we want to show

$$B_{n11} = o_p(n^{-1}h_1^{-d_{13}/2}). \tag{B.1}$$

Let $\varphi_0(v_{t_1}, v_{t_2}, v_{t_3}) \equiv \int K_{1t_1t_2} K_{1t_1t_3} \left(E_{t_0}\left[K_{2t_2t_0}\right] - f_{2t_2}\right) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau)$. Because φ_0 is not symmetric in its arguments, we need to symmetrize it in order to apply Lemmas C.1 and C.2. A symmetrized version of φ_0 is $\varphi(v_{t_1}, v_{t_2}, v_{t_3}) = (1/3)\{\varphi_0(v_{t_1}, v_{t_2}, v_{t_3}) + \varphi_0(v_{t_2}, v_{t_1}, v_{t_3}) + \varphi_0(v_{t_3}, v_{t_1}, v_{t_2})\}$. Noting that φ is of the same order as φ_0, we will apply Lemmas C.1 and C.2 directly to φ_0 to simplify the proofs. This simplification is applied throughout this appendix.

Let M_1 be as defined in Lemma C.1, then $M_1^{1/(1+\delta)} = O(h_1^{-2\delta d_{13}/(1+\delta)} h_2^r)$. So by Lemma C.1 and Assumptions A.1-A.3, $E\left[B_{n11}\right] = O(n^{-1}h_1^{-2\delta d_{13}/(1+\delta)} h_2^r) = o(n^{-1}h_1^{-d_{13}/2})$. Let M_2, R_{24}, R_{25}, and R_{26} be as defined in Lemma C.2. Then $M_2^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)} h_2^{2r})$, $R_{24} = O(n^5 h_1^{-2d_{13}} h_2^{2r})$, $R_{25} = O(n^4 h_1^{-d_{13}} h_2^{2r})$, and $R_{26} = O(n^4 h_1^{-2d_{13}} h_2^{2r})$. So by Lemma C.2, Assumptions A.1-A.3, $E\left[B_{n11}\right]^2 = n^{-6}(O(n^3 M_4^{1/(1+\delta)}) + O(\sum_{s=4}^6 R_{2s})) = o(n^{-2}h_1^{-d_{13}})$. Consequently, $B_{n11} = o_p(n^{-1}h_1^{-d_{13}/2})$ by the Chebyshev inequality.

Next, we show

$$B_{n12} = o_p(n^{-1}h_1^{-d_{13}/2}). \tag{B.2}$$

It is easy to show that the summation of the $t_4 = t_1$ or $t_4 = t_3$ terms in B_{n12} is of order $o_p(n^{-1}h_1^{-d_{13}/2})$ by applying Lemmas C.1 and C.2, and $B_{n12} \backsim_p 24\widetilde{B}_{n12}$, where

$$\widetilde{B}_{n12} \equiv \sum_{1 \leq t_1 < \ldots < t_4 \leq n} \int K_{1t_1t_2} K_{1t_1t_3} \left(K_{2t_2t_4} - E_{t_4}\left[K_{2t_2t_4}\right]\right) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau),$$

and $A \backsim_p B$ means $A = B\{1 + o_p(1)\}$.

Now let $\varphi(v_{t_1}, v_{t_2}, v_{t_3}, v_{t_4}) \equiv \int K_{1t_1t_2} K_{1t_1t_3} (K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}]) f_{2t_3} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau)$. Let M_3 be as defined in Lemma C.3. Then $M_3^{1/(1+\delta)} = O(h_1^{-2\delta d_{13}/(1+\delta)} h_2^{-2\delta d_1/(1+\delta)})$. So by Lemma C.3 and Assumptions A.1-A.3, $E\left[\widetilde{B}_{n12}\right] = O(n^{-2} h_1^{-2\delta d_{13}/(1+\delta)} h_2^{-2\delta d_1/(1+\delta)}) = o(n^{-1} h_1^{-d_{13}/2})$. Let M_4, M_{44}, R_{46}, and R_{47} be as defined in Lemma C.4. Then $M_4^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)} h_2^{-2\delta d_1/(1+\delta)})$, $M_{44}^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)}\ h_2^{-(1+2\delta)d_1/(1+\delta)})$, $R_{46} = O(n^5 h_1^{-2d_{13}} h_2^{-d_1})$, and $R_{47} = O(n^4 h_1^{-2d_{13}} h_2^{-d_1})$. So by Lemma C.4 and Assumptions A.1-A.3, $E\left[\widetilde{B}_{n12}\right]^2 = n^{-8}(O(n^5 M_4^{1/(1+\delta)}) + O(n^4 M_{44}^{1/(1+\delta)}) + O(\sum_{s=6}^{7} R_{4s})) = o(n^{-2} h_1^{-d_{13}})$. Consequently, $B_{n12} = o_p(n^{-1} h_1^{-d_{13}/2})$ by the Chebyshev inequality. ■

Lemma B.2 $B_{n2} \equiv \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left(\widehat{f}_{2,t_2} - f_{2t_2}\right)\left(\widehat{f}_{2,t_3} - f_{2t_3}\right) \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau)$

$= o_p(n^{-1} h_1^{-d_{13}/2})$.

Proof. Write B_{n2}

$$
\begin{aligned}
= & \frac{1}{n_5} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2, t_5 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left[K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}]\right] \left[K_{2t_3t_5} - E_{t_5}[K_{2t_3t_5}]\right] \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau) \\
& + \frac{2}{n_4} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2} \int K_{1t_1t_2} K_{1t_1t_3} \left[K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}]\right] \left[E_{t_0}[K_{2t_3t_0}] - f_{2t_3}\right] \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau) \\
& + \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left[E_{t_0}[K_{2t_2t_0}] - f_{2t_2}\right] \left[E_{t_0}[K_{2t_3t_0}] - f_{2t_3}\right] \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau) \\
\equiv & B_{n21} + 2B_{n22} + B_{n23}.
\end{aligned}
$$

The proofs of $B_{n22} = o_p(n^{-1} h_1^{-d_{13}/2})$ and $B_{n23} = o_p(n^{-1} h_1^{-d_{13}/2})$ are analogous to those of (B.2) and (B.1), respectively. Note that $B_{n21} \backsim_p 120\widetilde{B}_{n21}$, where $\widetilde{B}_{n21}$ is defined as B_{n21} but with summation $\sum_{1 \leq t_1 < \ldots < t_5 \leq n}$ in place of $\sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2, t_5 \neq t_3}$. We are left to show

$$\widetilde{B}_{n21} = o_p(n^{-1} h_1^{-d_{13}/2}). \tag{B.3}$$

Let $\varphi(v_{t_1}, v_{t_2}, v_{t_3}, v_{t_4}, v_{t_5}) \equiv \int K_{1t_1t_2} K_{1t_1t_3} (K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}]) \left[K_{2t_3t_5} - E_{t_5}[K_{2t_3t_5}]\right] \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) dG(\tau)$. Let M_5 be as defined in Lemma C.5. Then $M_5^{1/(1+\delta)} = O(h_1^{-2\delta d_{13}/(1+\delta)} h_2^{-2\delta d_1/(1+\delta)})$. So by Lemma C.5 and Assumptions A.1-A.3, $E\left[\widetilde{B}_{n21}\right] = O(n^{-3} M_5^{1/(1+\delta)}) = o(n^{-1} h_1^{-d_{13}/2})$. Let M_6, M_{64}, M_{65} R_{67}, and R_{68} be as defined in Lemma C.6. Then $M_6^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)} h_2^{-2\delta d_1/(1+\delta)})$, $M_{64}^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)}\ h_2^{-(1+2\delta)d_1/(1+\delta)})$, $M_{65}^{1/(1+\delta)} = O(h_1^{-(2+4\delta)d_{13}/(1+\delta)} h_2^{-(2+2\delta)d_1/(1+\delta)})$, $R_{67} = O(n^6 h_1^{-2d_{13}}\ h_2^{-2d_1})$, and $R_{68} = O(n^5 h_1^{-2d_{13}} h_2^{-2d_1})$. So by Lemma C.6 and Assumptions A.1-A.3, $E\left[\widetilde{B}_{n21}\right]^2 = n^{-10}\ (O(n^7 M_6^{1/(1+\delta)}) + O(n^6 M_{64}^{1/(1+\delta)}) + O(n^5 M_{65}^{1/(1+\delta)}) + O(\sum_{s=7}^{8} R_{6s})) = o(n^{-2} h_1^{-d_{13}})$. Consequently, $B_{n22} = o_p(n^{-1} h_1^{-d_{13}/2})$ by the Chebyshev inequality. ■

Lemma B.3 $B_{n3} \equiv \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} f_{2t_2} \varepsilon_{t_2}(\tau) \widehat{f}_{2,t_3} \widehat{e}_{t_3}(\tau) dG(\tau) = o_p(n^{-1} h_1^{-d_{13}/2})$.

Proof. Write

$$\begin{aligned} B_{n3} &= \frac{1}{n_4} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2} \int K_{1t_1t_2} K_{1t_1t_3} K_{2t_3t_4} f_{2t_2} \varepsilon_{t_2}(\tau) \left[m\left(X_{t_3};\tau\right) - m\left(X_{t_4};\tau\right) \right] dG(\tau) \\ &+ \frac{1}{n_4} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2} \int K_{1t_1t_2} K_{1t_1t_3} K_{2t_3t_4} f_{2t_2} \varepsilon_{t_2}(\tau) \left[m\left(X_{t_4};\tau\right) - H\left(Y_{t_4} + \tau\right) \right] dG(\tau) \\ &\equiv B_{n31} + B_{n32}. \end{aligned}$$

Noting that $E\left[H\left(Y_t + \tau\right)|X_t\right] = m\left(X_t;\tau\right)$ and $|m(x;\tau) - m(x';\tau)| \leq D_m(x;\tau) \|x - x'\|$ with $\int D_m(x;\tau)^{2(1+\delta)} dF(x) dG(\tau) < \infty$, we can modify the proof of (B.2) in Lemma B.1 and show $B_{n3i} = o_p(n^{-1} h_1^{-d_{13}/2})$, $i = 1$ and 2. ■

Lemma B.4 $B_{n4} \equiv \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} \left(\widehat{f}_{2t_2} - f_{2t_2} \right) \varepsilon_{t_2}(\tau) \widehat{f}_{2t_3} \widehat{e}_{t_3}(\tau) dG(\tau) = o_p(n^{-1} h_1^{-d_{13}/2})$.

Proof. Write B_{n4}

$$\begin{aligned} &= \frac{1}{n_5} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2, t_5 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} K_{2t_2t_4} K_{2t_3t_5} \left[K_{2t_2t_4} - f_{2t_2} \right] \left[H(Y_{t_5} + \tau) - m(X_{t_3};\tau) \right] dG(\tau) \\ &= \frac{1}{n_5} \sum_{t_1 \neq t_2 \neq t_3, t_4 \neq t_2, t_5 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} K_{2t_2t_4} K_{2t_3t_5} \left\{ \left[K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}] \right] \left[H(Y_{t_5} + \tau) - m(X_{t_5};\tau) \right] \right) \\ &+ \left[K_{2t_2t_4} - E_{t_4}[K_{2t_2t_4}] \right] \left[m(X_{t_5};\tau) - m(X_{t_3};\tau) \right] + \left[E_{t_4}[K_{2t_2t_4}] - f_{2t_2} \right] \left[H(Y_{t_5} + \tau) - m(X_{t_5};\tau) \right] \\ &+ \left[E_{t_4}[K_{2t_2t_4}] - f_{2t_2} \right] \left[m(X_{t_5};\tau) - m(X_{t_3};\tau) \right] \right\} dG(\tau) \\ &\equiv B_{n41} + B_{n42} + B_{n43} + B_{n44}. \end{aligned}$$

As in Lemmas B.1 and B.2, we can show that each of $B_{n4i}, i = 1, ..., 4$, is $o_p(n^{-1} h_1^{-d_{13}/2})$. ■

Lemma B.5 $\Gamma_{n11} \equiv \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \int K_{1t_1t_2} K_{1t_1t_3} f_{2t_2} f_{2t_3} \varepsilon_{t_2}(\tau) \varepsilon_{t_3}(\tau) dG(\tau) = \widetilde{\Gamma}_{n11} + o_p(n^{-1} h_1^{-d_{13}/2})$, *where* $\widetilde{\Gamma}_{n11} \equiv \frac{2}{n(n-1)} \sum_{1 \leq t < s \leq n} \int \int K_{(x,z)t} K_{(x,z)s} f_{2t} f_{2s} \varepsilon_t(\tau) \varepsilon_s(\tau) dG(\tau) dF(x,z)$.

Proof. This lemma is an analog of Lemma B.6 in Su and White (2005). We simplify their proof by applying a technical lemma given in Appendix C. Let $\Delta_n \equiv \Gamma_{n11} - \widetilde{\Gamma}_{n11}$. Let $\psi(v_{t_1}, v_{t_2}, v_{t_3}) \equiv \int K_{1t_1t_2} K_{1t_1t_3} f_{2t_2} \ f_{2t_3} \varepsilon_{t_2}(\tau) \varepsilon_{t_3}(\tau) dG(\tau)$. Then $\widetilde{\Gamma}_{n11} = E_{v_{t_1}} \left[\psi(v_{t_1}, v_{t_2}, v_{t_3}) \right]$ and

$$\begin{aligned} \Delta_n &= \frac{1}{n_3} \sum_{t_1 \neq t_2 \neq t_3} \left\{ \psi(v_{t_1}, v_{t_2}, v_{t_3}) - E_{v_{t_1}} \left[\psi(v_{t_1}, v_{t_2}, v_{t_3}) \right] \right\} \\ &= \frac{6}{n_3} \sum_{1 \leq t_1 < t_2 < t_3 \leq n} \left\{ E\left[\psi(v_{t_1}, v_{t_2}, v_{t_3}) | v_{t_2}, v_{t_3} \right] - E_{v_{t_1}} \left[\psi(v_{t_1}, v_{t_2}, v_{t_3}) \right] \right\} \\ &+ \frac{6}{n_3} \sum_{1 \leq t_1 < t_2 < t_3 \leq n} \left\{ \psi(v_{t_1}, v_{t_2}, v_{t_3}) - E\left[\psi(v_{t_1}, v_{t_2}, v_{t_3}) | v_{t_2}, v_{t_3} \right] \right\} \\ &\equiv 6\Delta_{n1} + 6\Delta_{n2}. \end{aligned}$$

It suffices to show $\Delta_{ni} = o_p(n^{-1}h_1^{-d_{13}/2})$, $i = 1$, 2. By the triangle inequality,

$$\begin{aligned} E|\Delta_{n1}| &\leq \frac{1}{n_3} \sum_{\substack{1\leq t_1<t_2<t_3\leq n \\ t_2-t_1\geq t_3-t_2}} E\left|E\left[\psi(v_{t_1},v_{t_2},v_{t_3})|v_{t_2},v_{t_3}\right] - E_{v_{t_1}}\left[\psi(v_{t_1},v_{t_2},v_{t_3})\right]\right| \\ &\quad + \frac{1}{n_3} \sum_{\substack{1\leq t_1<t_2<t_3\leq n \\ t_3-t_2\geq t_2-t_1}} E\left|E\left[\psi(v_{t_1},v_{t_2},v_{t_3})|v_{t_2},v_{t_3}\right] - E_{v_{t_1}}\left[\psi(v_{t_1},v_{t_2},v_{t_3})\right]\right| \\ &\equiv \Delta_{11} + \Delta_{12}. \end{aligned}$$

By Assumption A.1(i) and Yoshihara (1989),

$$\begin{aligned} \Delta_{11} &\leq \frac{1}{n_3} \sum_{t_1=1}^{n-2} \sum_{t_2=2}^{n-1} \sum_{t_3=t_2+1}^{t_2+(t_2-t_1)} 4h_1^{-2d_{13}\delta/(1+\delta)} \beta^{\delta/(1+\delta)}(t_2-t_1) \\ &\leq \frac{4nh_1^{-2d_{13}\delta/(1+\delta)}}{n_3} \sum_{k=1}^{n} k\beta^{\delta/(1+\delta)}(k) = o(n^{-1}h_1^{-d_{13}/2}). \end{aligned}$$

Similarly, we can show $\Delta_{12} = o(n^{-1}h_1^{-d_{13}/2})$. Consequently, $\Delta_{n1} = o_p(n^{-1}h_1^{-d_{13}/2})$ by the Markov inequality.

Now let $\varphi_{t_1t_2t_3} \equiv \varphi(v_{t_1},v_{t_2},v_{t_3}) \equiv \psi(v_{t_1},v_{t_2},v_{t_3}) - E\left[\psi(v_{t_1},v_{t_2},v_{t_3})|v_{t_2},v_{t_3}\right]$. Then $\Delta_{n2} = n_3^{-1} \sum_{1\leq t_1<t_2<t_3\leq n} \varphi_{t_1t_2t_3}$. Clearly, $E(\Delta_{n2}) = 0$. By the Chebyshev inequality, it suffices to show $E(\Delta_{n2})^2 = o(n^{-2}h_1^{-d_{13})})$. This follows by an application of Lemma C.2. The conclusion thus follows. ■

C Some Technical Lemmas

Let $\{V_t, t \geq 1\}$ be a d-dimensional stationary absolutely regular process satisfying Assumption A.1(i) in the main text. Let $P(V)$ denote the probability law of a random variable V. Let $1 \leq i_1, i_2, ..., i_k \leq n$ be arbitrary positive integers. For any j $(1 \leq j \leq k)$, define a collection of probability measures $\mathcal{P}_j^k$ by

$$\begin{aligned} \mathcal{P}_j^k(V_{i_1},...,V_{i_k}) \equiv &\Big\{P_j^k(V_{i_1},...,V_{i_k}) \equiv \Pi_{s=1}^j P(\underline{V}_s) : \underline{V}_s \text{ is a subset of } \{V_{i_1},...,V_{i_k}\}, \\ &\cup_{s=1}^j \underline{V}_s = \{V_{i_1},...,V_{i_k}\}, \text{ and } \underline{V}_t \cap \underline{V}_s = \varnothing \text{ for all } 1 \leq t \neq s \leq j\Big\}. \end{aligned}$$

In the following, we frequently suppress the arguments of P_j^k and $\mathcal{P}_j^k$ when no confusion can arise. For example, when $k = 3$, we use $\max_{1\leq j\leq 3} \max_{P_j^3\in\mathcal{P}_j^3} \int_{\mathbb{R}^{3d}} \left|\varphi\left(v_{i_1},v_{i_2},v_{i_3}\right)\right|^{1+\delta} dP_j^3$ to denote

$$\begin{aligned} \max\Big\{&\int_{\mathbb{R}^{3d}} \{|\varphi(v_1,v_2,v_3)|^{1+\delta} dF_{i_1i_2i_3}(v_1,v_2,v_3), \int_{\mathbb{R}^{3d}} |\varphi(v_1,v_2,v_3)|^{1+\delta} dF(v_1)dF_{i_2i_3}(v_2,v_3), \\ &\int_{\mathbb{R}^{3d}} |\varphi(v_1,v_2,v_3)|^{1+\delta} dF(v_2)dF_{i_1i_3}(v_1,v_3), \int_{\mathbb{R}^{3d}} |\varphi(v_1,v_2,v_3)|^{1+\delta} dF(v_3)dF_{i_1i_2}(v_1,v_2), \\ &\int_{\mathbb{R}^{3d}} |\varphi(v_1,v_2,v_3)|^{1+\delta} \Pi_{i=1}^3 dF(v_i)\Big\} \end{aligned}$$

where, e.g., $F_{i_1i_2i_3}$ is the joint distribution of $(V_{i_1}, V_{i_2}, V_{i_3})$.

Below we assume φ is symmetric in its arguments, and state the lemmas without presenting detailed proofs. Note that Lemma C.1 is implied by Lemma B.2 in Fan and Li (1999), and Lemma C.2 is an extension of Lemma A of Hjellvik et. al. (1998). In comparison with Hjellvik et. al. (1998), we don't assume $E\varphi(V_{i_1}, v_{i_2}, ..., v_{i_k}) = 0$, and hence our results are not as succinct as theirs. All lemmas can be proved by using Lemma 1 of Yoshihara (1976) repeatedly.

Lemma C.1 $S_1 \equiv E\left[\sum_{i_1<i_2<i_3} \varphi(V_{i_1}, V_{i_2}, V_{i_3})\right] = O(n^3 E\ [\varphi(\overline{V}_{i_1}, \overline{V}_{i_2}, \overline{V}_{i_3})]) + O(n^2 M_1^{\frac{1}{1+\delta}})$, *where*

$$M_1 \equiv \max_{1<i_1<i_2\leq n} \max_{1\leq j\leq 3} \max_{P_j^3 \in \mathcal{P}_j^3} \int_{\mathbb{R}^{3d}} \left|\varphi\left(v_1, v_{i_1}, v_{i_2}\right)\right|^{1+\delta} dP_j^3.$$

Lemma C.2 $S_2 \equiv E\left[\sum_{i_1<i_2<i_3} \varphi(V_{i_1}, V_{i_2}, V_{i_3})\right]^2 = O(n^3 M_2^{\frac{1}{1+\delta}}) + O(\sum_{s=1}^6 R_{2s})$, *where* $M_2 \equiv \max\{M_{21}, M_{22}, M_{23}\}$,

$$M_{21} \equiv \max_{1\leq i_1<...<i_6\leq n} \max_{1\leq j\leq 3} \max_{P_j^6 \in \mathcal{P}_j^6} \int_{\mathbb{R}^{6d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}\right) \varphi\left(v_{i_4}, v_{i_5}, v_{i_6}\right)\right|^{1+\delta} dP_j^6,$$

$$M_{22} \equiv \max_{1\leq i_1<...<i_5\leq n} \max_{1\leq j\leq 3} \max_{P_j^5 \in \mathcal{P}_j^5} \int_{\mathbb{R}^{5d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}\right) \varphi\left(v_{i_1}, v_{i_4}, v_{i_5}\right)\right|^{1+\delta} dP_j^5,$$

$$M_{23} \equiv \max_{1\leq i_1<...<i_4\leq n} \max_{1\leq j\leq 3} \max_{P_j^4 \in \mathcal{P}_j^4} \int_{\mathbb{R}^{4d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}\right) \varphi\left(v_{i_1}, v_{i_2}, v_{i_4}\right)\right|^{1+\delta} dP_j^4,$$

$$R_{21} \equiv n^6 \max_{i_3<i_5<i_6} \int_{\mathbb{R}^{3d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, V_{i_3}\right) \varphi\left(v_{i_4}, V_{i_5}, V_{i_6}\right)\right] dF(v_{i_1}) dF(v_{i_2}) dF(v_{i_4}),$$

$$R_{22} \equiv n^5 \max_{i_4<i_5} \int_{\mathbb{R}^{3d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}\right) \varphi\left(v_{i_1}, V_{i_4}, V_{i_5}\right)\right] \Pi_{s=1}^3 dF(v_{i_s}),$$

$$R_{23} \equiv n^5 \max_{i_1<i_5} \int_{\mathbb{R}^{3d}} E\left[\varphi\left(V_{i_1}, v_{i_2}, v_{i_3}\right) \varphi\left(V_{i_1}, v_{i_4}, V_{i_5}\right)\right] \Pi_{s=2}^4 dF(v_{i_s}),$$

$$R_{24} \equiv n^5 \max_{i_3<i_5} \int_{\mathbb{R}^{3d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, V_{i_3}\right) \varphi\left(v_{i_1}, v_{i_4}, V_{i_5}\right)\right] dF(v_{i_1}) dF(v_{i_2}) dF(v_{i_4}),$$

$$R_{25} \equiv n^4 E\left[\varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3\right) \varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_4\right)\right],$$

and

$$R_{26} \equiv n^3 \max_{1<i_1<i_2\leq n} E\left[\varphi\left(V_1, V_{i_1}, V_{i_2}\right)\right]^2.$$

Remark. In certain cases, the above results can be simplified: (1) if $\int_{\mathbb{R}^{2d}} \varphi(v_1, v_2, v_3) \Pi_{s=1}^2 dF(v_s) = 0$, $S_2 = O(n^3 M_2^{\frac{1}{1+\delta}}) + O(\sum_{s=4}^6 R_{2s})$, and (2) if $\int_{\mathbb{R}^d} \varphi(v_1, v_2, v_3) dF(v_1) = 0$, $S_2 = O(n^3 M_2^{\frac{1}{1+\delta}}) + O(R_{26})$.

Lemma C.3 $S_3 \equiv E\left[\sum_{i_1<i_2<i_3<i_4} \varphi(V_{i_1}, V_{i_2}, V_{i_3}, V_{i_4})\right] = O(n^4 E\ [\varphi(\overline{V}_{i_1}, \overline{V}_{i_2}, \overline{V}_{i_3}, \overline{V}_{i_4})]) + O(n^3 M_{33}^{\frac{1}{1+\delta}})$ $+O(n^2 M_{32}^{\frac{1}{1+\delta}}) + O(n M_{31}^{\frac{1}{1+\delta}})$, *where*

$$\begin{aligned} M_{31} \quad \equiv \quad & \max_{1\leq i_1<...<i_4\leq n} \max\left\{\int_{\mathbb{R}^{4d}} \left|\varphi(v_1, v_2, v_3, v_4)\right|^{1+\delta} dF_{i_1 i_2 i_3 i_4}(v_1, v_2, v_3, v_4),\right. \\ & \left.\int_{\mathbb{R}^{4d}} \left|\varphi(v_1, v_2, v_3, v_4)\right|^{1+\delta} dF(v_1) dF_{i_2 i_3 i_4}(v_2, v_3, v_4)\right\}, \end{aligned}$$

$$M_{32} \equiv \max_{2\leq i_2<...<i_4\leq n} \max\left\{\int_{\mathbb{R}^{4d}} |\varphi(v_1,v_2,v_3,v_4)|^{1+\delta} dF(v_1)\, dF_{i_2i_3i_4}(v_2,v_3,v_4),\right.$$
$$\left.\int_{\mathbb{R}^{4d}} |\varphi(v_1,v_2,v_3,v_4)|^{1+\delta} dF(v_1)\, dF(v_2) dF_{i_3i_4}(v_3,v_4)\right\}, \text{ and}$$

$$M_{33} \equiv \max_{3\leq i_3<i_4\leq n} \max\left\{\int_{\mathbb{R}^{4d}} |\varphi(v_1,v_2,v_3,v_4)|^{1+\delta} dF(v_1)\, dF(v_2) dF_{i_3i_4}(v_3,v_4),\right.$$
$$\left.\int_{\mathbb{R}^{4d}} |\varphi(v_1,v_2,v_3,v_4)|^{1+\delta} \Pi_{s=1}^4 dF(v_s)\right\}.$$

Remark. In certain cases, the above results can be simplified: (1) if $\int_{\mathbb{R}^{4d}} \varphi(v_1,...,v_4)\, \Pi_{s=1}^4 dF(v_s) = 0$, $S_3 = O(n^3 M_{33}^{\frac{1}{1+\delta}}) + O(n^2 M_{32}^{\frac{1}{1+\delta}}) + O(n M_{31}^{\frac{1}{1+\delta}})$; (2) if $\int_{\mathbb{R}^{2d}} \varphi(v_1,...,v_4)\, dF(v_1) dF(v_2) = 0$, $S_3 = O(n^2 M_{32}^{\frac{1}{1+\delta}}) + O(n M_{31}^{\frac{1}{1+\delta}})$; and (3) if $\int_{\mathbb{R}^d} \varphi(v_1,...,v_4)\, dF(v_1) = 0$, $S_3 = O(n M_{31}^{\frac{1}{1+\delta}})$.

Lemma C.4 $S_4 \equiv E\left[\sum_{i_1<i_2<i_3<i_4} \varphi(V_{i_1},V_{i_2},V_{i_3},V_{i_4})\right]^2 = O(n^5 M_4^{\frac{1}{1+\delta}}) + O(n^4 M_{44}^{\frac{1}{1+\delta}}) + O(\sum_{s=1}^7 R_{4s})$, *where* $M_4 \equiv \max\{M_{41}, M_{42}, M_{43}\}$,

$$M_{41} \equiv \max_{1\leq i_1<...<i_8\leq n} \max_{1\leq j\leq 5} \max_{P_j^8\in\mathcal{P}_j^8} \int_{\mathbb{R}^{8d}} \left|\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(v_{i_5},v_{i_6},v_{i_7},v_{i_8})\right|^{1+\delta} dP_j^8,$$

$$M_{42} \equiv \max_{1\leq i_1<...<i_7\leq n} \max_{1\leq j\leq 5} \max_{P_j^7\in\mathcal{P}_j^7} \int_{\mathbb{R}^{7d}} \left|\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(v_{i_1},v_{i_5},v_{i_6},v_{i_7})\right|^{1+\delta} dP_j^7,$$

$$M_{43} \equiv \max_{1\leq i_1<...<i_6\leq n} \max_{1\leq j\leq 5} \max_{P_j^6\in\mathcal{P}_j^6} \int_{\mathbb{R}^{6d}} \left|\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(v_{i_1},v_{i_2},v_{i_5},v_{i_6})\right|^{1+\delta} dP_j^6,$$

$$M_{44} \equiv \max_{1\leq i_1<...<i_5\leq n} \max_{1\leq j\leq 4} \max_{P_j^5\in\mathcal{P}_j^5} \int_{\mathbb{R}^{5d}} \left|\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_5})\right|^{1+\delta} dP_j^5,$$

$$R_{41} \equiv n^8 \max_{i_4<i_7<i_8} \int_{\mathbb{R}^{5d}} E\left[\varphi(v_{i_1},v_{i_2},v_{i_3},V_{i_4})\,\varphi(v_{i_5},v_{i_6},V_{i_7},V_{i_8})\right] \Pi_{s=1}^3 dF(v_{i_s}) \Pi_{t=5}^6 dF(v_{i_t}),$$

$$R_{42} \equiv n^7 \max_{i_6<i_7} \int_{\mathbb{R}^{5d}} E\left[\varphi(v_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(v_{i_1},v_{i_5},V_{i_6},V_{i_7})\right] \Pi_{s=1}^5 dF(v_{i_s}),$$

$$R_{43} \equiv n^7 \max_{i_1<i_7} \int_{\mathbb{R}^{5d}} E\left[\varphi(V_{i_1},v_{i_2},v_{i_3},v_{i_4})\,\varphi(V_{i_1},v_{i_5},v_{i_6},V_{i_7})\right] \Pi_{s=2}^6 dF(v_{i_s}),$$

$$R_{44} \equiv n^7 \max_{i_4<i_7} \int_{\mathbb{R}^{5d}} E\left[\varphi(v_{i_1},v_{i_2},v_{i_3},V_{i_4})\,\varphi(v_{i_1},v_{i_5},v_{i_6},V_{i_7})\right] \Pi_{s=1}^3 dF(v_{i_s}) \Pi_{t=5}^6 dF(v_{i_t}),$$

$$R_{45} \equiv n^6 E\left[\varphi(\overline{V}_1,\overline{V}_2,\overline{V}_3,\overline{V}_4)\,\varphi(\overline{V}_1,\overline{V}_2,\overline{V}_5,\overline{V}_6)\right],$$

$$R_{46} \equiv n^5 E\left[\varphi(\overline{V}_1,\overline{V}_2,\overline{V}_3,\overline{V}_4)\,\varphi(\overline{V}_1,\overline{V}_2,\overline{V}_3,\overline{V}_5)\right],$$

and

$$R_{47} \equiv n^4 \max_{1<i_1<i_2<i_3\leq n} E\left[\varphi(V_1,V_{i_1},V_{i_2},V_{i_3})\right]^2.$$

Remark. In certain cases, the above results can be simplified: (1) if $\int_{\mathbb{R}^{3d}} \varphi(v_1,...,v_4)\, \Pi_{s=1}^3 dF(v_s) = 0$, $S_4 = O(n^5 M_4^{\frac{1}{1+\delta}}) + O(n^4 M_{44}^{\frac{1}{1+\delta}}) + O(\sum_{s=4}^7 R_{4s})$; (2) if $\int_{\mathbb{R}^{2d}} \varphi(v_1,...,v_4)\, dF(v_1) dF(v_2) = 0$, $S_4 = O(n^5 M_4^{\frac{1}{1+\delta}}) + O(n^4 M_{44}^{\frac{1}{1+\delta}}) + O(\sum_{s=6}^7 R_{4s})$; and (3) if $\int_{\mathbb{R}^d} \varphi(v_1,...,v_4)\, dF(v_1) = 0$, $S_4 = O(n^5 M_4^{\frac{1}{1+\delta}}) + O(n^4 M_{44}^{\frac{1}{1+\delta}}) + O(R_{47})$.

Lemma C.5 $S_5 \equiv E\left[\sum_{i_1<i_2<i_3<i_4<i_5} \varphi\left(V_{i_1}, V_{i_2}, V_{i_3}, V_{i_4}, V_{i_5}\right)\right] = O(n^5 E\left[\varphi\left(\overline{V}_{i_1}, \overline{V}_{i_2}, \overline{V}_{i_3}, \overline{V}_{i_4}, \overline{V}_{i_5}\right)\right]) + O(n^4 M_{54}^{\frac{1}{1+\delta}}) + O(n^3 M_{53}^{\frac{1}{1+\delta}}) + O(n^2 M_{52}^{\frac{1}{1+\delta}}) + O(n M_{51}^{\frac{1}{1+\delta}})$, *where*

$$\begin{aligned} M_{51} &\equiv \max_{1\le i_1<\ldots<i_5\le n} \max\left\{\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} dF_{i_1 i_2 i_3 i_4 i_5}\left(v_1, v_2, v_3, v_4, v_5\right),\right. \\ &\left.\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} dF\left(v_1\right) dF_{i_2 i_3 i_4 i_5}\left(v_2, v_3, v_4\right)\right\}, \end{aligned}$$

$$\begin{aligned} M_{52} &\equiv \max_{2\le i_2<\ldots<i_5\le n} \max\left\{\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} dF\left(v_1\right) dF_{i_2 i_3 i_4 i_5}\left(v_2, v_3, v_4, v_5\right),\right. \\ &\left.\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} dF\left(v_1\right) dF(v_2) dF_{i_3 i_4 i_5}\left(v_3, v_4, v_5\right)\right\}, \end{aligned}$$

$$\begin{aligned} M_{53} &\equiv \max_{3\le i_3<\ldots<i_5\le n} \max\left\{\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} dF\left(v_1\right) dF(v_2) dF_{i_3 i_4 i_5}\left(v_3, v_4, v_5\right),\right. \\ &\left.\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} \Pi_{s=1}^{3} dF\left(v_s\right) dF_{i_4 i_5}\left(v_4, v_5\right)\right\}, \end{aligned}$$

and

$$\begin{aligned} M_{54} &\equiv \max_{4\le i_4<i_5\le n} \max\left\{\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} \Pi_{s=1}^{3} dF\left(v_s\right) dF_{i_4 i_5}\left(v_4, v_5\right),\right. \\ &\left.\int_{\mathbb{R}^{5d}} \left|\varphi\left(v_1, v_2, v_3, v_4, v_5\right)\right|^{1+\delta} \Pi_{s=1}^{5} dF\left(v_s\right)\right\}. \end{aligned}$$

Remark. In certain cases, the above results can be simplified: (1) if $\int_{\mathbb{R}^{5d}} \varphi\left(v_1, \ldots, v_5\right) \Pi_{s=1}^{5} dF(v_s) = 0$, $S_5 = O(n^4 M_{54}^{\frac{1}{1+\delta}}) + O(n^3 M_{53}^{\frac{1}{1+\delta}}) + O(n^2 M_{52}^{\frac{1}{1+\delta}}) + O(n M_{51}^{\frac{1}{1+\delta}})$; (2) if $\int_{\mathbb{R}^{3d}} \varphi\left(v_1, \ldots, v_5\right) \Pi_{s=1}^{3} dF(v_s) = 0$, $S_5 = O(n^3 M_{53}^{\frac{1}{1+\delta}}) + O(n^2 M_{52}^{\frac{1}{1+\delta}}) + O(n M_{51}^{\frac{1}{1+\delta}})$; (3) if $\int_{\mathbb{R}^{2d}} \varphi\left(v_1, \ldots, v_5\right) dF(v_1) dF(v_2) = 0$, $S_5 = O(n^2 M_{52}^{\frac{1}{1+\delta}}) + O(n M_{51}^{\frac{1}{1+\delta}})$; and (4) if $\int_{\mathbb{R}^{d}} \varphi\left(v_1, \ldots, v_5\right) dF(v_1) = 0$, $S_5 = O(n M_{51}^{\frac{1}{1+\delta}})$.

Lemma C.6 $S_6 \equiv E\left[\sum_{i_1<i_2<i_3<i_4<i_5} \varphi\left(V_{i_1}, V_{i_2}, V_{i_3}, V_{i_4}, V_{i_5}\right)\right]^2 = O(n^7 M_6^{\frac{1}{1+\delta}}) + O(n^6 M_{64}^{\frac{1}{1+\delta}}) + O(n^5 M_{65}^{\frac{1}{1+\delta}}) + O(\sum_{s=1}^{8} R_{6s})$, *where* $M_6 \equiv \max\left\{M_{61}, M_{62}, M_{63}\right\}$,

$$M_{61} \equiv \max_{1\le i_1<\ldots<i_{10}\le n} \max_{1\le j\le 7} \max_{P_j^{10}\in\mathcal{P}_j^{10}} \int_{\mathbb{R}^{10d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_6}, v_{i_7}, v_{i_8}, v_{i_9}, v_{i_{10}}\right)\right|^{1+\delta} dP_j^{10},$$

$$M_{62} \equiv \max_{1\le i_1<\ldots<i_9\le n} \max_{1\le j\le 7} \max_{P_j^{9}\in\mathcal{P}_j^{9}} \int_{\mathbb{R}^{9d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_1}, v_{i_6}, v_{i_7}, v_{i_8}, v_{i_9}\right)\right|^{1+\delta} dP_j^{9},$$

$$M_{63} \equiv \max_{1\le i_1<\ldots<i_8\le n} \max_{1\le j\le 7} \max_{P_j^{8}\in\mathcal{P}_j^{8}} \int_{\mathbb{R}^{8d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_1}, v_{i_2}, v_{i_6}, v_{i_7}, v_{i_8}\right)\right|^{1+\delta} dP_j^{8},$$

$$M_{64} \equiv \max_{1\le i_1<\ldots<i_7\le n} \max_{1\le j\le 7} \max_{P_j^{7}\in\mathcal{P}_j^{7}} \int_{\mathbb{R}^{7d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_6}, v_{i7}\right)\right|^{1+\delta} dP_j^{7},$$

$$M_{65} \equiv \max_{1\le i_1<\ldots<i_6\le n} \max_{1\le j\le 6} \max_{P_j^{6}\in\mathcal{P}_j^{6}} \int_{\mathbb{R}^{6d}} \left|\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i6}\right)\right|^{1+\delta} dP_j^{6},$$

$$R_{61} \equiv n^{10} \max_{i_5<i_9<i_{10}} \int_{\mathbb{R}^{7d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, V_{i_5}\right) \varphi\left(v_{i_6}, v_{i_7}, v_{i_8}, V_{i_9}, V_{i_{10}}\right)\right] \Pi_{s=1}^4 dF(v_{i_s}) \Pi_{t=6}^8 dF(v_{i_t}),$$

$$R_{62} \equiv n^9 \max_{i_8<i_9} \int_{\mathbb{R}^{7d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(v_{i_1}, v_{i_6}, v_{i_7}, V_{i_8}, V_{i_9}\right)\right] \Pi_{s=1}^7 dF(v_{i_s}),$$

$$R_{63} \equiv n^9 \max_{i_1<i_7} \int_{\mathbb{R}^{7d}} E\left[\varphi\left(V_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, v_{i_5}\right) \varphi\left(V_{i_1}, v_{i_6}, v_{i_7}, v_{i_8}, V_{i_9}\right)\right] \Pi_{s=2}^8 dF(v_{i_s}),$$

$$R_{64} \equiv n^9 \max_{i_4<i_7} \int_{\mathbb{R}^{7d}} E\left[\varphi\left(v_{i_1}, v_{i_2}, v_{i_3}, v_{i_4}, V_{i_5}\right) \varphi\left(v_{i_1}, v_{i_6}, v_{i_7}, v_{i_8}, V_{i_9}\right)\right] \Pi_{s=1}^4 dF(v_{i_s}) \Pi_{t=6}^8 dF(v_{i_t}),$$

$$R_{65} \equiv n^8 E\left[\varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_4, \overline{V}_5\right) \varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_6, \overline{V}_7, \overline{V}_8\right)\right],$$

$$R_{66} \equiv n^7 E\left[\varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_4, \overline{V}_5\right) \varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_6, \overline{V}_7\right)\right],$$

$$R_{67} \equiv n^6 E\left[\varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_4, \overline{V}_5\right) \varphi\left(\overline{V}_1, \overline{V}_2, \overline{V}_3, \overline{V}_4, \overline{V}_6\right)\right],$$

and

$$R_{68} \equiv n^5 \max_{1<i_1<i_2<i_3<i_4\leq n} E\left[\varphi\left(V_1, V_{i_1}, V_{i_2}, V_{i_3}, V_{i_4}\right)\right]^2.$$

Remark. In certain cases, the above results can be simplified: (1) if $\int_{\mathbb{R}^{4d}} \varphi\left(v_1, ..., v_5\right) \Pi_{s=1}^4 dF(v_s) = 0$, R_{61} through R_{63} vanish in S_6; (2) if $\int_{\mathbb{R}^{3d}} \varphi\left(v_1, ..., v_5\right) \Pi_{s=1}^3 dF(v_s) = 0$, R_{61} through R_{65} vanish in S_6; (3) if $\int_{\mathbb{R}^{2d}} \varphi\left(v_1, ..., v_5\right) dF(v_1) dF(v_2) = 0$, R_{61} through R_{66} vanish in S_6; and (4) if $\int_{\mathbb{R}^{d}} \varphi\left(v_1, ..., v_5\right) dF(v_1) = 0$, R_{61} through R_{67} vanish in S_6.

References

Auestad, B., Tjostheim D., 1990. Identification of nonlinear time series: first order characterization and order determination. Biometrika 77, 669-687.

Bierens H. J., 1982. Consistent model specification test. Journal of Econometrics 20, 105-134.

Bierens, H. J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443-1458.

Brett, C., Pinkse, J., 1997. Those taxes are all over the map! A test of independence of municipal tax rates in British Columbia. International Regional Science Review 20, 131-152.

Cheng, B., Tong, H., 1992. On consistent nonparametric order determination and chaos. Journal of the Royal Statistical Society B 54, 427-449.

Delgado, M. A., González-Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics 29, 1469-1507.

Fan Y., Li, Q., 1999. Root-n-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11, 251-269.

Fernandes, M., Flores, R. G., 2000. Tests for conditional independence, Markovian dynamics, and noncausality. Discussion Paper, European University Institute.

Finkenstädt, B. F., Yao, Q., Tong, H., 2001. A conditional density approach to the order determination of time series. Statistics and Computing 11, 229-240.

Granger, C. W. J., 1980. Testing for causality: a personal viewpoint. Journal of Economic Dynamics and Control 2, 329-352.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926-1947.

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

Hjellvik, V., Yao, Q., Tjostheim, D., 1998. Linearity testing using local polynomial approximation. Journal of Statistical Planning and Inference 68, 295-321.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. Journal of the American Statistical Association 94, 1201-1220.

Hong, Y., White, H., 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. Econometrica 73, 837-901.

Hsieh, D. A., 1989. Testing for nonlinear dependence in daily foreign exchange rates. Journal of Business 62, 339-368.

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. Econometric Theory 16, 576-601.

Lee, S., 2003. Efficient semiparametric estimation of partially linear quantile regression model. Econometric Theory 19, 1-31.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101-147.

Linton, O., Gozalo, P., 1997. Conditional independence restrictions: testing and estimation. Discussion Paper, Cowles Foundation for Research in Economics, Yale University.

Lobato, I., 2003. Testing for nonlinear autoregression. Journal of Business and Economic Statistics 21, 164-173.

Paparoditis, E., Politis, D. N., 2000. The local bootstrap for kernel estimators under general dependence conditions. Annals of the Institute of Statistical Mathematics 52, 139-159.

Pinkse, J., 1998. A consistent nonparametric test for serial independence. Journal of Econometrics 84, 205-231.

Pinkse, J., 2002. Nonparametric misspecification testing. Discussion paper, Dept. of Economics, University of British Columbia.

Robinson, P. M., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 931-954.

Sheedy, E., 1998. Correlation in currency markets: a risk-adjusted perspective. Journal of International Financial Markets, Institutions & Money 8, 59-82.

Stinchcombe, M. B., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-325.

Su, L., White, H., 2005. A Hellinger-metric nonparametric test for conditional independence. Discussion Paper, Dept. of Economics, UCSD.

Tenreiro, C., 1997. Loi Asymptotique des erreurs quadratiques integrees des estimateurs a noyau de la densite et de la regression sous des conditions de dependance. Portugaliae Mathematica 54, 197-213.

Tjostheim, D., Auestad, B., 1994. Nonparametric identification of nonlinear time series: selecting significant lags. Journal of the American Statistical Association 89, 1410-1419.

Tschernig, R., Yang, L., 2000. Nonparametric lag selection for time series. Journal of Time Series Analysis 21, 457-487.

Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Z. Wahrsch. Verw. Gebiete 35, 237-252.

Yoshihara, K., 1989. Limiting behavior of generalized quadratic forms generated by absolutely regular processes. In Mandel, P. and Huskova, M. (eds.): Proceedings of the Fourth Prague Symposium on Asymptotic Statistics, pp. 539-547.