A Consistent Characteristic-Function-Based Test for Conditional Independence

Liangjun Su and Halbert White
Department of Economics, UCSD
La Jolla, CA 92093-0508

July 25, 2003

Abstract

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 use the fuctional delta method to expand the test statistic around the population truth and establish asymptotic normality under $\beta-$mixing conditions. We show that the test is consistent and has power against local alternatives at distance $n^{-1/2}h_1^{-(d_1+d_3)/4}$. The cases for which not all random variables of interest are continuously valued or observable are also treated, and we show that the test is nuisance-parameter free. Simulation results suggest that the test has better finite sample performance than the Hellinger metric test of Su and White (2002) in detecting nonlinear Granger causality in the mean. Applications to exchange rates and to stock prices and trading volumes indicate that our test can reveal some interesting nonlinear causal relations that the traditional linear Granger causality test fails to detect.

Key words: $\beta - mixing$, *Conditional characteristic function, Conditional independence, Functional delta method, Granger non-causality, Nonparametric regression, U-statistics.*

JEL Classification: C12, C14.

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 (2002), we write

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

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

There are many nonparametric tests of independence or serial independence in the literature, starting with Hoeffding (1948), and followed by those based on empirical distribution functions such as Blum et al. (1961), Skaug and Tjostheim (1993) and Delgado (1996), and those based on smoothing methods like Robinson (1991), Skaug and Tjostheim (1996), Zheng (1997) and Hong and White (2000).[1] Nevertheless, there are few nonparametric tests for conditional independence of continuous variables.[2] Linton and Gozalo (1997) propose two nonparametric tests of conditional independence between variables of interest based on a generalization of the empirical distribution function. Since the asymptotic null distributions of their test statistics are complicated functionals of a Gaussian process and depend on the underlying distributions, i.e., neither test is distribution free, a bootstrap procedure is needed for calculating the critical values. This hinders its potential application; to date no applications have appeared that we are aware of. Fernandes and Flores (1999) employ a generalized entropy measure to test conditional independence, but the asymptotic normal null distribution relies heavily on the choice of suitable weighting functions. Simulation results indicate that their test has poor size properties and low or no power against causality in variance. Recently, Su and White (2002) 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 their test is easy to implement, it has some limitations in that it uses the same bandwidth sequence in estimating all required joint and marginal densities nonparametrically, and such a procedure is unsatisfactory when the dimension of (X,Y,Z) is above three.

Motivated by the notion that two conditional distributions are identical if and only if their respective conditional characteristic functions are equal, we propose a new test that is closely related to conditional characteristic functions but avoids estimating them directly. The empirical characteristic function (ECF) has a long history in testing hypotheses in both statistics and econometrics literature. It has been mainly used to test for goodness-of-fit, symmetry and homogeneity (e.g., Feuerverger and Mureika, 1977; Koutrouvelis, 1980; Koutrouvelis and Kellermeier, 1981; Baringhaus and Henze, 1988; Henze and Zirkler, 1990; Ghosh and Ruymgaart, 1992; Heathcote, 1995; Naito, 1996a, 1996b; Fan, 1997; Gurtler and Henze, 2000; Alba et al., 2001). In most cases, the resulting null distribution is not normal or chi-squared, and simulation is required to obtain the critical values. Early papers that use ECF to test for independence include de Silva and Griffiths (1980) and Csörgo (1985). The latter test parallels that of Blum et. al. (1961) and evaluates the difference between the joint and product ECFs where it is most variable. The

[1] It seems that the correlation integral based BDS test (Brock et al., 1987) does not belong to either category here.

[2] For categorical data there are also numerous tests of independence and conditional independence, see Rosenbaum (1984), Agresti (1990) and Yao and Tritchler (1993), among others.

test statistic has an asymptotic χ^2_2 null distribution. Feuerverger (1987) notes the possibility of using the ECF to test serial independence. In a sequence of papers, Brett and Pinkse (1997) and Pinkse (1998, 2000) use a characteristic function principle and a weighted integral approach to test for spatial independence, serial independence and independence, respectively. In all cases, the resulting null distribution is χ^2_1. In contrast, Hong(1999) proposes a generalized spectral density approach with numerical integration over a pair of auxiliary parameters to conduct hypothesis tests in time series. The approach essentially uses ECFs and their derivatives in a time series framework in a clever way, and the resulting null distribution for the test statistic is asymptotically normal.

In this paper we borrow ideas from both Pinkse's and Hong's approach. Like them, we base our test upon the properties of conditional characteristic functions and use a weighted integral approach. We also borrow ideas from Bierens (1982) and its various following papers. As a result, our test is consistent against all deviations from conditional independence on a compact subset of the support of the density of (X, Z). Nevertheless, our approach differs from those of Pinkse, Hong and Bierens in fundamental ways. Specifically, we exploit a particular weighting function and the properties of nonparametric kernel regression to obtain a test statistic that has asymptotically normal null distribution as in Hong (1999), and in contrast to the difficult asymptotic null distributions arising in using the approach of Bierens (1982) and his followers.[3]

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 conditional characteristic functions, 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 quantiles and conditional symmetry, can also be studied using our approach.[4] These tests are naturally suited to helping answer 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, 1969, 1980) is a particular case of conditional independence (see Florens and Mouchart, 1982; Florens and Fougere, 1996). Our test can be directly applied to test for Granger non-causality without the need to specify a particular linear or non-linear models.[5]

[3] Conditional moment tests of a finite number of conditional moments as proposed by Newey (1985) and Tauchen (1985) do not posses the property of consistency against all possible alternatives. Consistency can be accomplished by employing a nuisance parameter not identified under the null as in Bierens (1982, 1984, 1987, 1990), Bierens and Hartog (1988), de Jong(1996), and Stinchcombe and White (1998). The use of the nuisance parameter effectively permits the test to examine an infinite number of conditional moments.

[4] For a different approach, see Inoue (1998) who borrows ideas from Bierens (1990) and de Jong (1996) and proposes a unified approach for consistent testing of linear restrictions on the conditional distribution function of a time series. As in Bierens (1990) and de Jong (1996), the asymptotic null distribution is not standard and the proposed test is conservative for small and moderate sample sizes.

[5] In the same spirit, Baek and Brock (1992) propose a nonparametric test for causality based on the so called correlation integral, an estimator of spatial probabilities across time. Hiemstra and Jones (1994) generalize their approach to allow for data dependence and apply the test to aggregate daily stock prices and trading volumes data, revealing significant nonlinear

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. Also, it is common in econometrics that conditional independence test would naturally be conducted using estimated residuals or other estimated random variables, which are a function of the observed data and some parameter estimators. For motivational examples, see Linton and Gozalo (1997) and Su and White (2002). Under some regularity conditions, we show here that parameter estimation error has no effect on the asymptotic null distribution of our test statistic.

The remainder of this paper is organized as follows. In Section 2, we describe the basic framework for our nonparametric test for conditional independence where there is no parameter estimation problem involved and all random variables are continuously valued. In section 3 we study the asymptotic null distribution of the test statistic and the consistency and local power properties of our test. In Section 4, we extend the results to allow for parameter estimation error and discuss certain other relevant issues. We examine the finite sample performance of our test via Monte Carlo simulation in Section 5. We apply our test to exchange rate data and to stock price and trading volume data in Section 6. Final remarks are contained in Section 7. 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 (resp. cumulative distribution function) of $(X_t,\ Y_t,\ Z_t)$ is denoted by f (resp. 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.,

$$H_0:\ Pr\{f(Y|X, Z) = f(Y|X)\} = 1. \tag{2.1}$$

The alternative hypothesis is that $f(y|x, z) \neq f(y|x)$, over a non-negligible range of the support of the joint density f, or, more precisely,

$$H_1:\ Pr\{f(Y|X, Z) = f(Y|X)\} < 1. \tag{2.2}$$

The proposed test is based on characteristic functions. 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 conditional

causal relations between them.

characteristic functions (c.c.f.) $\phi_{Y|X,Z}$ of Y conditional on (X, Z) and $\phi_{Y|X}$ 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 $u \in R^{d_2}$ 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 $u \in R^{d_2}$.

Consider the following smooth functional[6]

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

where $a(x, z)$ is a given known nonnegative weighting function with compact support A on $R^{d_1+d_3}$; 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 R^{d_2} and the choice for g is arbitrary except that it must be nonnegative, integrable and bounded on R^{d_2} with support S. For the moment, one can take $S = R^{d_2}$.

The choice of the above functional is intuitive. Under the null, $\psi(u; x, z) = 0$ a.e.-(x, z) for every $u \in R^{d_2}$, and consequently $\Gamma(f, F) = 0$. The following lemma says essentially 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 A for every $\tau \in R^{d_2}$ if and only if $\psi(u; x, z) = 0$ a.e.-$G_0 \times F$ on $R^{d_2} \times A$.*

The proof of the above lemma is relegated to Appendix C. It is a modification of the proof of Theorem 1 in Bierens (1982). Bierens (1982, 1990) proposed consistent tests for functional form of nonlinear regression models based on a Fourier transform of conditional expectations. Consider a generic regression model $Y = g(X) + \varepsilon$, 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 $R^{d_2} \times \Theta$ and Θ is a parameter space containing the unknown parameters θ_0 if the specification is true. Under the null of correct specification, i.e., $P[g(X) = f(X, \theta_0)] = 1$ 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 $\tau \in R^{d_1}$ 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-polynominal analytic function.

An especially important point concerning (2.3) is that it is straightforward to develop asymptotic theory for the resulting test statistic. Note that 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 obtain

$$\begin{aligned}\Gamma(f,F) &= \int\int |E[H(Y+\tau)|x,z] - E[H(Y+\tau)|x]|^2 a(x,z) dF(x,z) dG(\tau) \\ &= \int\int \left| \frac{\int H(y+\tau) f(x,y,z)dy}{f(x,z)} - \frac{\int H(y+\tau) f(x,y)dy}{f(x)} \right|^2 a(x,z) dF(x,z) dG(\tau).\end{aligned} \tag{2.4}$$

[6]Alternatively, one can consider the smooth functional $\Gamma(f, F) \equiv \int\int |\psi(u; x, z)|^2 g^2(u)du \; a(x, z)dF(x, z)$ as in Pinkse (1998, 2000). But we find it is difficult to study the asymptotic theory in this case.

This integral facilitates the 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[7] K and bandwidth h, we define[8]

$$K_h(u) \equiv h^{-d}K(u/h), \tag{2.5}$$

where d is the dimension of the vector u. Specifically, we estimate $m(x,z;\tau) \equiv E[H(Y+\tau)|X=x, Z=z]$ by the standard Nadaraya-Watson (NW) kernel regression estimator,

$$\widehat{m}_{h_1}(x,z;\tau) \equiv \widehat{E}[H(Y+\tau)|x,z] \equiv \frac{\sum_{t=1}^{n} K_{h_1}(x-X_t, z-Z_t)H(Y_t+\tau)}{\sum_{t=1} K_{h_1}(x-X_t, z-Z_t)}, \tag{2.6}$$

and $m(x;\tau) \equiv E[H(Y+\tau)|X=x]$ by

$$\widehat{m}_{h_2}(x;\tau) \equiv \widehat{E}[H(Y+\tau)|x] \equiv \frac{\sum_{t=1}^{n} K_{h_2}(x-X_t)H(Y_t+\tau)}{\sum_{t=1} K_{h_2}(x-X_t)}. \tag{2.7}$$

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.

Next, it is convenient to introduce the two density estimates:

$$\widehat{f}_{h_1}(x,y,z) \equiv \frac{1}{n}\sum_{t=1}^{n} K_{h_1}(x-X_t, y-Y_t, z-Z_t), \tag{2.8}$$

$$\widehat{f}_{h_2}(x,y) \equiv \frac{1}{n}\sum_{t=1}^{n} K_{h_2}(x-X_t, y-Y_t). \tag{2.9}$$

Similarly, we define the estimates $\widehat{f}_{h_1}(x,z)$ of $f(x,z)$, calculated with bandwidth h_1, and $\widehat{f}_{h_2}(x)$ of $f(x)$, calculated with bandwidth h_2. Finally, let $\widehat{F}(x,z)$ denote the empirical distribution function of (X,Z). The form of the test statistic we consider is

$$\begin{aligned}\widehat{\Gamma} \equiv \Gamma(\widehat{f},\widehat{F}) &\equiv \int_S\int_A |\widehat{E}[H(Y+\tau)|x,z] - \widehat{E}[H(Y+\tau)|x]|^2 a(x,z)d\widehat{F}(x,z)dG(\tau) \\ &= \frac{1}{n}\sum_{t=1}^{n}\int |\widehat{m}_{h_1}(X_t,Z_t;\tau) - \widehat{m}_{h_2}(X_t;\tau)|^2 a(X_t,Z_t)dG(\tau),\end{aligned} \tag{2.10}$$

that is, we compare $\widehat{m}_{h_1}(x,z;\tau)$ to $\widehat{m}_{h_2}(x;\tau)$ by their integrated squared distance weighted by $a(x,z)g(\tau)$. It is straightforward to show that $\widehat{\Gamma}$ is a consistent estimator of Γ under mild conditions.

Note that because H is a characteristic function and thus uniformly bounded on its support, the appearance of kernel density estimators in the denominator in estimating the conditional expectations does not necessitate using the weighting function, a (see Lemma $B.2$ in Appendix B). Nevertheless, the use of a facilitates the proof of our theorems in several places. Moreover, it allows us to focus the

[7]For simplicity only, we will take the multivariate kernel function K to be a product of the univariate kernel function k.

[8]To keep the notation simple, we do not explicitly indicate the dependence of the bandwidth parameters h on the sample size n. We also adopt the same notational convention for kernel K as for density f, namely, to indicate which kernel by the list of its arguments or by specifying the dimension of its arguments.

conditional independence test on particular ranges of the data. By choosing an appropriate a, the test can be tailored to the empirical question of interest. For example, one may be interested in whether Y and Z are conditionally independent only for positive values of X. Here, we will consider functions a which are bounded with compact support $A \subset R^{d_1+d_3}$ strictly contained in the support of the density $f(x,z)$. In typical examples, a will be either the indicator function of a compact set A or a density-type function with compact support A. As a result, we shall only detect deviations between $f(y|x,z)$ and $f(y|x)$ that arise on A. [9]

We will show that the properties of our test statistic can be derived from the properties of (2.4). Two observations are of particular importance: (a) the first order terms in the functional expansion of $\Gamma(\widehat{f}, F)$ around $\Gamma(f, F)$ are degenerate under the null, and (b) the distance between $\Gamma(\widehat{f}, \widehat{F})$ and $\Gamma(\widehat{f}, F)$ is asymptotically negligible. The latter is important in that it is easier to study the asymptotic behavior of $\Gamma(\widehat{f}, F)$ than that of $\Gamma(\widehat{f}, \widehat{F})$. The former is useful in that it underlies the distribution theories for a variety of tests for independence, serial independence, significance, correct specification, etc.. The usual $\sqrt{n}-$asymptotics (e.g. Robinson (1991)) do not apply and different normalization schemes need to be adopted, as in Hong and White (2000).

3 The asymptotic distribution of the test statistic

In this section we focus on the case for which conditional independence test is based on 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 deferred to Section 4.

3.1 Asymptotic null distribution

We work with the dependence notion of $\beta-$mixing. See Appendix A for its definition and other technical material. Our assumptions are as follows.

Assumption A.1 (Stochastic Process)

(i) $\{W_t \equiv (X_t', Y_t', Z_t')',\ t \geq 0\}$ is a strictly stationary absolutely regular process on $R^{d_1+d_2+d_3} \equiv R^d$ with mixing coefficients β_m that satisfy $\beta_m = O(\rho^m)$ for some $0 < \rho < 1$.

(ii) For some even integer $r \geq 2$, $W_t \equiv (X_t', Y_t', Z_t')'$ has a joint distribution F and joint density f such that $f \in \mathcal{W}^d(r+1)$, i.e., f has continuous partial derivatives up to order $r+1$ which are bounded and integrable on R^d. Furthermore, f satisfies a Lipschitz condition: $|f(w+u) - f(w)| \leq D(w)||u||$ where D has finite $(2+\eta)th$ moment for some $\eta > 0$ and $||\cdot||$ is the usual Euclidean norm.

(iii) The joint probability density function ($p.d.f.$) $f_{t_1,\dots,t_l}$ of $(W_0, W_{t_1}, \dots, W_{t_l})$ $(1 \leq l \leq 5)$ is bounded and satisfies a Lipschitz condition: $|f_{t_1,\dots,t_l}(w_0+u_0, \cdots, w_l+u_l) - f_{t_1,\dots,t_l}(w_0, \cdots, w_l)| \leq D_{t_1,\dots,t_l}(w_0, \dots, w_l)||u||$, where $u \equiv (u_0, \dots, u_l)$ and $D_{t_1,\dots,t_l}$ is integrable and satisfies the conditions that $\int D_{t_1,\dots,t_l}(w_0, \dots, w_l)||w||^{2\xi} dw < \overline{M} < \infty$ and $\int D_{t_1,\dots,t_l}(w_0, \dots, w_l) f_{t_1,\dots,t_l}(w_0, \cdots, w_l) dw < \overline{M} < \infty$ for some $\xi > 1$.

[9] One can choose the compact subset of the support of (X, Z) so that it expands as the sample size increases. In this sense, our test is fairly general and may lose power agaist deviations in the extreme tails.

Assumption A.2 (Kernel and bandwidth)

(i) The kernel K is a product kernel of the same univariate symmetric kernel $k : R \to R$ satisfying

$\int_R u^i k(u)du = \delta_{i0}$ $(i = 0, 1, ..., r-1)$,

$C_0 \equiv \int_R u^r k(u)du < \infty$, $\int_R u^2 k(u)^2 du < \infty$,and

$k(u) = O((1 + |u|^{r+1+\delta})^{-1})$ for some $\delta > 0$,

where δ_{ij} is Kronecker's delta. (For example, for $u \in R^{d_j}$, $K(u) \equiv \Pi_{i=1}^{d_j} k(u_i)$ for $j = 1, 2, 3$ or 4 with $d_4 \equiv d$.)

(ii) The bandwidth sequences $h_1 = O(n^{-1/\delta_1})$ and $h_2 = O(n^{-1/\delta_2})$ are such that $2(d_1 + d_3) < \delta_1 < 2r + (d_1 + d_3)/2$, $d_1 < \delta_2 \leq 2r + d_1$ and $\delta_1 d_1/(d_1 + d_3) \leq \delta_2 < \delta_1$.

Assumption A.3 (Weight functions)

(i) The nonnegative weight function a is chosen such that $f(x, z)$ is bounded away from zero on the compact support A of $a(x, z)$, i.e., $\inf_{(x,z)\in A} f(x, z) \equiv b > 0$.

(ii) The weight function g_0 has full support on 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.

(iii) The weight function g is uniformly bounded, integrable and nonnegative everywhere on its support S.

Remarks.

Assumption $A.1(i)$ requires that $\{W_t\}$ be a stationary absolutely regular process with geometric decay rate. This is standard for application of a central limit theorem for U-statistics for weakly dependent data (e.g. Fan and Li 1999a). This condition is not stringent because it is weaker than $\phi-$mixing, and many well-known processes are absolutely regular with geometric decay rate.[10] For example, linear stationary ARMA processes satisfy this condition provided the innovation process $\{\varepsilon_t\}$ satisfies certain conditions (e.g. one sufficient condition is that $\{\varepsilon_t\}$ has absolutely continuous distribution with respect to Lebesgue measure). Moreover, under certain conditions, a large class of processes implied by numerous nonlinear models such as bilinear models, NLAR models and ARCH models satisfy absolute regularity with geometric decay rate (see Fan and Li 1999b)). $A.1(ii)$ and (iii) are primarily smoothness conditions, which are weak in the sense that they are similar to those for the case of independent data. Similar conditions are used in Li (1999).

Assumption $A.2(i)$ requires that the kernel be of second order or higher. Unless $d_1 + d_3 = 2$, a high order kernel has to be used, which is nevertheless common in the literature. See Robinson (1988) and Li (1999) among many others. Assumption $A.2(ii)$ specifies conditions on the choice of bandwidth sequences. Under the assumptions made on the bandwidth sequences, we have in particular that $nh_1^{(d_1+d_2)/2+2r} \to 0$, $nh_1^{(d_1+d_3)} \to \infty$, $nh_2^{d_1} \to \infty$ and $nh_2^{d_1+2r} \to C$ for some $C \in [0, \infty)$, $h_2/h_1 \to 0$ and $h_1^{(d_1+d_3)}/h_2^{d_1} \to 0$. Thus asymptotically we have that $h_1^{d_1} \gg h_2^{d_1} \gg h_1^{(d_1+d_3)}$.

Assumption $A.3(i)$ is discussed above. $A.3(ii)$ 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 bound-

[10] It is well known that the $\phi-$mixing condition has limited applications; for example, an ARMA process is never $\phi-$mixing but generally geometrically absolutely regular (Harel and Puri, 1996).

edly $(r+1)$- differentiable is also easily met in practice by choosing g_0 appropriately.[11] For example, one can choose g_0 as a standard normal density function on R^{d_2}. Alternatively, one can choose it to be a double exponential density function with mean zero and arbitrary finite variance. The main aspect of Assumption $A.3$ that is of potential concern in applications is how to choose g_0 and g so that any numerical integration can be done quickly or one can work out the integration analytically. We return to this point in Section 5.

Our first main result is that the test statistic is asymptotically normally distributed with asymptotic bias terms that can be consistently estimated. To state the result, we define the following notation:

$\alpha(y; x, z, \tau) \equiv [H(y+\tau) - m(x, z; \tau)]/f(x, z)$,
$\beta(y; x, \tau) \equiv [H(y+\tau) - m(x; \tau)]/f(x)$,
$\sigma^2(x, z; \tau) \equiv E\{[H(Y+\tau) - m(X, Z; \tau)]^2 | X = x, Z = z\}$,
$\sigma^2(x; \tau) \equiv E\{[H(Y+\tau) - m(X; \tau)]^2 | X = x\}$,
$B_1 \equiv C_1{}^{(d_1+d_3)} \int_S \int_A \sigma^2(x, z; \tau)\ a(x, z)\ d(x, z) dG(\tau)$,
$B_2 \equiv -2C_2{}^{d_1} \int_S \int_A \sigma^2(x, z; \tau)\ [f(x, z)/f(x)]\ a(x, z)\ d(x, z) dG(\tau)$, and
$B_3 \equiv C_1{}^{d_1} \int_S \int_A \sigma^2(x; \tau)\ [f(x, z)/f(x)]\ a(x, z) d(x, z) dG(\tau)$,

where

$C_1 \equiv \int_R k(u)^2 du$ and $C_2 \equiv k(0)$.

For simplicity, we will often omit the integration ranges. Further, define

$\sigma_1^2 \equiv 2C_3{}^{(d_1+d_3)} \int\int\int [\sigma^2(x, z; \tau, \tau')]^2 a(x, z)^2 d(x, z) dG(\tau) dG(\tau')$,

where

$C_3 \equiv \int_R (\int_R k(u+v) k(u) du)^2 dv$ and
$\sigma^2(x, z; \tau, \tau') \equiv \mathrm{cov}(H(Y+\tau), H(Y+\tau') \mid X = x, Z = z)$.

Assumptions $A.1 - A.3$ guarantee that the $C_i's$ and σ_1^2 are well defined and bounded away from zero and infinity. For any given univariate kernel satisfying Assumption $A.2(i)$, the C_i's can be calculated explicitly. Take $r = 4$ as an example. If the fourth order kernel $k(\cdot)$ is constructed from $\varphi(\cdot)$, the $p.d.f.$ of the standard normal distribution, then $k(u) = (3 - u^2)\varphi(u)/2$, and the $C_i's$ can be obtained as follows: $C_1 = 27/(32\sqrt{\pi})$, $C_2 = 3/(2\sqrt{2\pi})$, and $C_3 = 7881/(8192\sqrt{2\pi})$.

We can now sate our first result.

Theorem 3.1 *Under Assumptions $A.1 - A.3$ and under H_0, then,*

$$nh_1^{(d_1+d_3)/2}\{\widehat{\Gamma} - n^{-1}h_1^{-(d_1+d_3)}B_1 - n^{-1}h_1^{-d_1}B_2 - n^{-1}h_2^{-d_1}B_3\} \xrightarrow{d} N(0, \sigma_1^2).$$

Remarks.

The proof of the above theorem relies on the expansion of the functional $\Gamma(\cdot, F)$ around $\Gamma(f, F)$ as done by Ait-Sahalia et al. (2001) and the use of some preliminary results in Tenreiro (1997). In studying goodness-of-fit tests for kernel regression, Ait-Sahalia et al. derive the expansion of their functional for the sum-of-squared departures between the restricted regression and the unrestricted regression. As in their case, we carry out the second order expansion, because the first order terms vanish under the null hypothesis. Tenreiro (1997) uses U-statistic theory to study the asymptotics for the integrated squared

[11] One can relax the assumption that H be real-valued using somewhat more complicated notation.

error of kernel density estimators (see also Tenreiro, 1995, and Gourieroux and Tenreiro, 2001), and his result can be adapted to our framework.

To implement the test, we require consistent estimators of the bias terms. We denote the consistent estimators of these bias terms as $\widehat{B}_j$, $j = 1, ..., 3$, which are:

$\widehat{B}_1 \equiv C_1{}^{d_1+d_3} n^{-1} \sum_{t=1}^n \int \{\widehat{\sigma}^2_{h_1}(X_t, Z_t; \tau) a(X_t, Z_t)/\widehat{f}_{h_1}(X_t, Z_t)\} dG(\tau)$,

$\widehat{B}_2 \equiv -2C_2{}^{d_1} n^{-1} \sum_{t=1}^n \int \{\widehat{\sigma}^2_{h_1}(X_t, Z_t; \tau) a(X_t, Z_t)/\widehat{f}_{h_2}(X_t)\} dG(\tau)$,

$\widehat{B}_3 \equiv C_1{}^{d_1} n^{-1} \sum_{t=1}^n \int \{\widehat{\sigma}^2_{h_2}(X_t; \tau) a(X_t, Z_t)/\widehat{f}_{h_2}(X_t)\} dG(\tau)$,

$\widehat{\sigma}^2_1 \equiv 2C_3{}^{(d_1+d_3)} n^{-1} \sum_{t=1}^n \int \int [\widehat{\sigma}^2_{h_1}(X_t, Z_t; \tau, \tau')]^2 a(X_t, Z_\tau)^2/\widehat{f}_{h_1}(X_\tau, Z_\tau) dG(\tau) dG(\tau')$,

where

$$\widehat{\sigma}^2_{h_2}(x;\tau) = \frac{\sum_{t=1}^n K_{h_2}(x - X_t) H^2(Y_t + \tau)}{\sum_{t=1}^n K_{h_2}(x - X_t)} - \left[\frac{\sum_{t=1}^n K_{h_2}(x - X_t) H(Y_t + \tau)}{\sum_{t=1}^n K_{h_2}(x - X_t)}\right]^2,$$

$$\widehat{\sigma}^2_{h_1}(x, z;\tau) = \frac{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) H^2(Y_t + \tau)}{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)} - \left[\frac{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) H(Y_t + \tau)}{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)}\right]^2,$$

and

$$\begin{aligned} \widehat{\sigma}^2_{h_1}(x, z;\tau, \tau') \quad = \quad & \frac{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) H(Y_t + \tau) H(Y_t + \tau')}{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)} \\ & - \frac{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) H(Y_t + \tau)}{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)} \frac{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) H(Y_t + \tau')}{\sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)}. \end{aligned}$$

It is easy to show that $\widehat{\sigma}^2_1$ and $\widehat{B}_j$, j =1, 2, 3 can be substituted for σ^2_1 and the B_j's in Theorem 3.1 with no effect on the asymptotic distribution.

We then compare

$$T_{1,n} \equiv n h_1^{(d_1+d_3)/2} \{\widehat{\Gamma} - n^{-1} h_1^{-(d_1+d_3)} \widehat{B}_1 - n^{-1} h_1^{-d_1} \widehat{B}_2 - n^{-1} h_2^{-d_1} \widehat{B}_3\} / \sqrt{\widehat{\sigma}^2_1} \tag{3.1}$$

with the one-sided critical value z_α from the $N(0, 1)$ distribution, i.e., $z_{0.01} = 2.327$, $z_{0.05} = 1.645$ and $z_{0.10} = 1.282$, and reject the null when $T_{1,n} > z_\alpha$.

3.2 Consistency and local power properties

In this section, we start by studying the consistency of the test, i.e., its ability to reject a false null hypothesis with probability approaching 1 as $n \to \infty$. We then examine its local power, i.e., the probability of rejecting a false hypothesis, against sequences of alternatives that get closer to the null as $n \to \infty$. This is given more precisely as follows.

Proposition 3.2 *Under Assumptions A.1 − A.3, the test based on the statistics (3.1) is consistent for all F such that* $\Gamma(f, F) \geq \varepsilon > 0$.

Remarks.

Note that the above proposition is equivalent to saying that $\{E[H(Y + \tau)|x, z] - E[H(Y + \tau)|x]\} \times a(x, z) \neq 0$ in a region of positive density -$F \times G$. In theory, we would like the support A of $a(\cdot)$ to be as large as possible. In practice, it is usually taken that $A = A_1 \times A_2 \subset R^{d_1} \times R^{d_3}$, $A_1 = \{x = (x_1, ..., x_{d_1}) \in R^{d_1} : x_i \in [\overline{X}_i - c\ \widehat{S}_{X_i}, \overline{X_i} + c\ \widehat{S}_{X_i}]$ for some positive constant c, with $\overline{X}_i$ and $\widehat{S}_{X_i}$ being the sample average and standard deviation of X_i, respectively; and A_2 is defined analogously.[12] Note that here the support A is data-dependent, but this has no asymptotic impact on the distribution of our test statistic.

We now examine the power of our test against the sequence of local alternatives defined by a sequence of densities $f^{[n]}(x,y,z)$ such that, for $f^{[n]}(x,y) \equiv \int f^{[n]}(x,y,z)dz$, $f^{[n]}(x,z) \equiv \int f^{[n]}(x,y,z)dy$, $f^{[n]}(x) \equiv \int f^{[n]}(x,y,z)dydz$, we have

$$||f^{[n]}(x,y,z) - f(x,y,z)||_\infty = o(n^{-1/2}h_1^{-(d_1+d_3)/4}).$$

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, the local alternative can be specified as[13]

$$H_1(\alpha_n) : \sup\{|m^{[n]}(x,z;\tau) - m^{[n]}(x;\tau) - \alpha_n \Delta(x,z;\tau)| : (x,z) \in R^{d_1+d_3}, \tau \in R^{d_2}\} = o(\alpha_n) \qquad (3.2)$$

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

$$\delta \equiv \lim_{n\to\infty} \int \int_A \Delta^2(x,z;\tau)a(x,z)dF^{[n]}(x,z)dG(\tau) < \infty,$$

where $F^{[n]}(x,z)$ is the *c.d.f.* associated with $f^{[n]}(x,z)$.

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 3.3 *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*

$$\Pr(T_{1,n} \geq z_\alpha | H_1(\alpha_n)) \to 1 - \Phi(z_\alpha - \delta/\sigma_1).$$

Remarks.

Proposition 3.3 says that our test statistic $T_{1,n}$ has nontrivial power against $H_1(\alpha_n)$ with $\alpha_n = n^{-1/2}h_1^{-(d_1+d_2)/4}$ whenever $\delta \neq 0$. The rate $\alpha_n = n^{-1/2}h_1^{-(d_1+d_3)/4}$ is slower than the parametric rate $n^{-1/2}$, as $h_1 \to 0$, but is faster than $n^{-1/4}$. For example, when $d_1 = d_3 = 1$, one can choose $h_1 \propto n^{-1/6}$, $h_2 \propto n^{-1/5}$, and have $n^{-1/2}h_1^{-(d_1+d_3)/4} \propto n^{-5/12}$, which converges to zero faster than $n^{-1/3}$. The rate α_n could be made even closer to $n^{-1/2}$ but is always slower than $n^{-1/2}$. In practice, we need to choose h_1 and h_2 to balance the level and power in finite samples, and data-driven methods will be desirable in choosing the $h's$ for simulation and empirical applications.

4 Extensions

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

[12] Alternatively, one can use the Bartlett kernel function (or other density-form function) as the weighting function a. For example, if the i'th element in $U \equiv (X,Z)$, U_i, has mean zero and standard deviation one (perhaps after being recentered and rescaled), for $i = 1, ..., d$, one can use $a(u) = \Pi_{i=1}^d[(1/2 + 1/4u_i)1\{-2 \leq u_i \leq 0\} + (1/2 - 1/4u_i)1\{0 < u_i \leq 2\}]$. In this case, a has compact support $[-2,2]^d$.

[13] Alternatively, one can specify local alternatives in terms of densities as in Su and White (2002): $f^{[n]}(y|x,z) = f^{[n]}(y|x)[1 + a_n\widetilde{\Delta}(x,y,z) + o(a_n)\widetilde{\Delta}_n(x,y,z)]$. Then $\Delta(x,z;t) = \lim_{n\to\infty} \int H(y+t)\widetilde{\Delta}(x,y,z)f^{[n]}(y|x)dy$ in (3.2).

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 cases that generalize the basic result in the last section but focus on the case of testing for conditional independence with parameters estimated.

4.1 Conditional independence test with unobservables

In this subsection, we consider the case for which $W = (X', Y', Z')'$ is not observed but can be estimated using a finite-dimensional parameter estimator.[14] Asymptotic results for this case are useful when the conditional independence test is conducted using residuals or other estimated random variables. An example is given in the introduction of Su and White (2002), in which a sample selection model is parametrically specified.

Now the observed process is written $\{M_t \in R^k,\ t \geq 0\}$. Of interest are certain residual or index functions calculated from M, that is, $W(M,\theta) \equiv (X(M,\theta), Y(M,\theta), Z(M,\theta)) \in R^{d_1+d_2+d_3} \equiv R^d$, where the parameter θ belongs to $\Theta \subset R^p$. The null hypothesis to be tested is that

$$H_0 : Y(M,\theta_0) \perp Z(M,\theta_0)|X(M,\theta_0) \tag{4.1}$$

for some particular $\theta_0 \in \Theta$, whose value is unknown to us.

Denote the probability density function of $W \equiv (X,Y,Z), (X,Y), (X,Z)$ and X by $f(w;\theta)$, $f(x,y;\theta)$, $f(x,z;\theta)$ and $f(x;\theta)$, respectively. Let $F(w;\theta)$ be the *c.d.f.* of W. Also, let $f(w) = f(w;\theta_0)$, $f(x,y) = f(x,y;\theta_0)$, $f(x,z) = f(x,z;\theta_0)$, $f(x) = f(x;\theta_0)$ and $F(x,z) = F(x,z;\theta_0)$. Let E_θ be the expectation under the law associated with $f(\cdot;\theta)$. Define $m(x,z;\tau,\theta) \equiv E_\theta(H(Y+\tau)|X = x, Z = z)$ and $m(x;\tau,\theta) \equiv E_\theta(H(Y+\tau)|X = x)$. Under the null, we have

$$\Gamma_2(f,F;\theta) \equiv \int\int [m(x,z;\tau,\theta) - m(x;\tau,\theta)]^2 a(x,z;\theta)dF(x,z;\theta)dG(\tau) = 0, \tag{4.2}$$

for $\theta = \theta_0$, the pseudo-true underlying parameter, where $a(x,z;\theta) \equiv a(x(\theta), z(\theta))$ is a nonnegative weighting function which depends on θ only through (x,z) and is otherwise the same as $a(x,z)$ used in Section 3.

We suppose that there is an estimator $\widehat{\theta}$ of θ_0 that is $\sqrt{n}-$consistent under the null hypothesis.[15] In some cases, there are many candidate estimators. In particular, Linton and Gozalo (1997) show how one can obtain $\sqrt{n}-$consistent estimates with the null hypothesis imposed. To implement the test, we replace $\Gamma_2(f,F;\theta)$ by its sample analogue

$$\begin{aligned}\widehat{\Gamma}_2(\widehat{\theta}) &\equiv \Gamma_2(\widehat{f},\widehat{F};\widehat{\theta}) \equiv \int\int_A \left[\widehat{m}_{h_1}(x,z;\tau,\widehat{\theta}) - \widehat{m}_{h_2}(x;\tau,\widehat{\theta})\right]^2 a(x,z;\widehat{\theta})d\widehat{F}(x,z;\widehat{\theta})dG(\tau) \\ &= \frac{1}{n}\sum_{\tau=1}^{n}\int \left[\widehat{m}_{h_1}(X_t(\widehat{\theta}), Z_t(\widehat{\theta});\tau) - \widehat{m}_{h_2}(X_t(\widehat{\theta});\tau)\right]^2 a(X_t(\widehat{\theta}), Z_t(\widehat{\theta}))dG(\tau) \end{aligned} \tag{4.3}$$

[14] Extension to the case in which an infinite-dimensional parameter estimator is allowed is underway. This is very important in some semi- or non-parametrically specified simutaneous models. The major difficulty lies in the establishment of stochastic equicontinuity (SE) for the underlying empirical process with a sharp convergence rate. Results in Andrews (1995) need to be strengthened by using some primitive conditions for SE in Andrews (1994).

[15] This assumption can be relaxed. We only need to requre that the estimator $\widehat{\theta}$ converge to θ_0 at a sufficiently fast rate.

where, for example, $\widehat{m}_{h_1}(x,z;\tau,\widehat{\theta}) \equiv \widehat{m}_{h_1}(x(\widehat{\theta}),z(\widehat{\theta});\tau)$ is the standard N-W kernel estimator of $m(x,z;\tau,\widehat{\theta})$ by using "observations" $\{W_t(\widehat{\theta}) \equiv (X_t(\widehat{\theta}),Y_t(\widehat{\theta}),Z_t(\widehat{\theta})), 1 \le t \le n\}$.

Under some regularity conditions, we will show that

$$\widehat{\Gamma}_2(\widehat{\theta}) = \widehat{\Gamma}_2(\theta_0) + o_p(n^{-1}h_1^{-(d_1+d_3)/2}), \tag{4.4}$$

in which case we can say that our test for conditional independence is nuisance-parameter free.

Specifically, we make use of the following additional assumptions.

Assumption A.4

(i) $\sqrt{n}(\widehat{\theta}-\theta_0) = O_p(1)$.

(ii) $\{M_t \in R^k,\ t \ge 0\}$ is a strictly stationary absolutely regular process with mixing coefficients β_m that satisfy $\beta_m = O(\rho^m)$ for some $0<\rho<1$.

(iii) $W_t(\theta_0)$ satisfies Assumptions $A.1(ii)$ and (iii).

(iv) The joint density $f(w;\theta)$ of $W_t(\theta)$ is $(r+1)$ times continuously differentiable in θ in a neighborhood $\Theta_0 \subset \Theta$ of θ_0.

(v) The weighting function $a(u) \equiv a(x,z)$ is twice continuously differentiable a.e.-u on its support A and $W_t(\theta) \equiv W(M_t,\theta)$ is twice continuously differentiable in θ in a neighborhood $\Theta_0 \subset \Theta$ of θ_0 *a.s.*. Moreover,

$$E\sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\theta}W(M_t,\theta)\right\| < \infty, \quad E\sup_{\theta\in\Theta_0}\left\|\frac{\partial a}{\partial W_i}\frac{\partial^2}{\partial\theta\partial\theta'}W_i(M_t,\theta)\right\| < \infty,$$

and

$$E\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2 a}{\partial u_i\partial u_j}\frac{\partial}{\partial\theta}W_i(M_t,\theta)\frac{\partial}{\partial\theta'}W_j(M_t,\theta)\right\| < \infty,$$

where $W_i(M_t,\theta)$ is the ith element of the random vector $W(M_t,\theta)$.

Assumption $A.4(iv)$ implies θ_0 is an interior point of Θ. Assumption $A.4(v)$ is standard in the literature of nonparametric kernel estimation with estimated random variables; see Assumption $NP9$ in Andrews (1995). Like Linton and Gozalo (1997), we require that $W(m;\theta)$ be second order continuously differentiable in θ. The difference is that we don't assume that $W(m;\theta)$ is uniformly continuous in m.

The large sample behavior of $\widehat{\Gamma}_2(\widehat{\theta})$ is given in the following corollary, which is proved in Appendix C.

Corollary 4.1 *Under Assumptions* $A.1$ – $A.4$ *and under* H_0, *if* $nh_1^{3(d_1+d_3)/2+2} \to \infty$, *then* $nh_1^{(d_1+d_3)/2}\{\Gamma_2(\widehat{f},\widehat{F};\widehat{\theta}) - \Gamma_2(\widehat{f},\widehat{F};\theta_0)\} = o_p(1)$.

Note that the above corollary implies that the variance of the $\sqrt{n}$−consistent estimator $\widehat{\theta}$ does not affect the limiting distribution of $\Gamma_2(\widehat{f},\widehat{F};\widehat{\theta})$ and that we can ignore the fact that $\widehat{\theta}$ is estimated for the purposes of testing. The proof of the above corollary relies on the second order Taylor expansion of $\widehat{\Gamma}_2(\widehat{\theta}) = \Gamma_2(\widehat{f},\widehat{F};\widehat{\theta})$ around θ_0. Unsurprisingly, due to the finite dimensionality of θ and the smooth nature of the functional Γ_2, high level tools like stochastic equicontinuity are not required for our purposes. We remark that our problem is quite different from the situations discussed in White and Hong (1999) where stochastic equicontinuity plays an indispensable role. In our case, $\widehat{\theta}$ is not separable from $\widehat{m}_{h_j}(\cdot;\widehat{\theta})$. In other words, when $\theta=\theta_0$, $\widehat{m}_{h_j}(\cdot;\tau,\theta_0)$, $j=1,2$, are the estimators for the true conditional expectations, as desired. In contrast, White and Hong (1999) treat the nonparametric kernel estimators ($\widehat{m}'s$ in our notation) themselves as an infinite-dimensional parameter, and what makes the difference is that their finite-dimensional parameter estimator ($\widehat{\pi}_n$ in their notation) is separable from the infinite-dimensional parameter estimator ($\widehat{\theta}_n$ in their notation), whereas this is not satisfied in our case.

4.2 Further Extensions

4.2.1 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 functional expansion in a natural way, 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 limited dependent variables. For example, Y may be a discrete response, or a more complicated censored or truncated version of a continuous (latent) variable. For a different approach, see Su and White (2002). Also, one can allow a mixture of continuous and discrete conditioning variables. The modification can be done by following the approach of Racine and Li (2000).

4.2.2 Nuisance parameter

Our results produce a testing procedure based on a direct comparison of two sequences of nonparametric regression estimators $\widehat{m}_{h_1}(x, z; \tau)$ and $\widehat{m}_{h_2}(x; \tau)$ indexed by the nuisance parameter[16] τ. Instead of taking finitely many gridpoints over the support of $dG(\tau)$, we integrate out τ over a compact support as in Hong (1999). This ensures a reasonable ominibus test which has a nice limiting null distribution. As Bierens (1990) shows, if (X, Z) is bounded on its support, we can limit τ to a neighborhood of zero. Let Φ be an arbitrary Borel measurable bounded one-to-one mapping from $R^{d_1+d_3}$ into $R^{d_1+d_3}$; then conditioning on (X, Z) is equivalent to conditioning on the bounded random vector $\Phi(X, Z)$, for $\Phi(X, Z)$ and (X, Z) generate the same Borel field. Thus it is not restrictive to limit one's attention to the case where (X, Z) is a bounded random vector.

4.2.3 Bias correction

There are three bias terms to be corrected in our test statistic. However, if the data are *i.i.d.*, it can be shown easily that the second and third terms of the bias can be removed in our setting by appealing to the clever centering device of Härdle and Mammen (1993). Given τ, we can first compute the conditional expectation estimator $\widehat{m}_{h_2}(x; \tau)$, then compute the kernel regression of $\widehat{m}_{h_2}(x; \tau)$ on (x, z), say $\widehat{sm}_{h_1}(x, z; \tau)$, and base the test on the difference between $\widehat{m}_{h_1}(x, z; \tau)$ and $\widehat{sm}_{h_1}(x, z; \tau)$. Note

[16] Strictly speaking, (x, z) are also nuisance parameters.

that for $U_1^n \equiv (X_1, ..., X_n; Z_1, ..., Z_n)$,

$$\begin{aligned} E[\widehat{m}_{h_1}(x,z;\tau)|U_1^n] &= \frac{1}{n}\sum_{t=1}^{n} K_{h_1}(x-X_t, z-Z_t)E[H(Y_t+\tau)|U_1^n]/\widehat{f}_{h_1}(x,z) \\ &= \frac{1}{n}\sum_{t=1}^{n} K_{h_1}(x-X_t, z-Z_t)m(X_t,Z_t;\tau)]/\widehat{f}_{h_1}(x,z), \end{aligned}$$

where the last equality follows from the *i.i.d.* assumption (in fact only independence is needed). Under the null, $m(X_t, Z_t; \tau) = m(X_t; \tau)$ a.s. for every $\tau \in R^{d_2}$ and we are thus led to replace $\widehat{m}_{h_2}(x;\tau)$ in our statistic by

$$\widehat{sm}_{h_1}(x,z;\tau) \equiv \frac{1}{n}\sum_{t=1}^{n} K_{h_1}(x-X_t, z-Z_t)\widehat{m}_{h_2}(X_t;\tau)/\widehat{f}_{h_1}(x,z)$$

to form $\widetilde{\Gamma} \equiv \frac{1}{n}\sum_{t=1}^{n}\int[\widehat{m}_{h_1}(X_t,Z_t;\tau) - \widehat{sm}_{h_1}(X_t,Z_t;\tau)]^2 a(X_t,Z_t)dG(\tau)$.

This procedure differs from our previous approach in that the restricted regression $\widehat{m}_{h_2}(x;\tau)$ is replaced by a kernel-smoothed version of it. As a result, the last two of the three bias terms in Theorem 3.1 become asymptotically negligible. In practical terms, this would reduce the need to do bias correction from three terms to one term. When the *i.i.d* or independence assumption fails, the above intuitive argument does not go through but the procedure still works. The following corollary is formally established in Appendix C.

Corollary 4.2 *Under Assumptions A.1-3, we have that under* H_0,

$$T_{2,n} \equiv nh_1^{(d_1+d_3)/2}\{\widetilde{\Gamma} - n^{-1}h_1^{-(d_1+d_3)}\widehat{B}_1\}/\sqrt{\widehat{\sigma}_1^2} \to N(0,1). \tag{4.5}$$

4.2.4 Testing for independence

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

$$H_0^*: \Pr\{f(Y|Z) = f(Y)\} = 1.$$

The alternative is that $f(y|z) \neq f(y)$ over a non-negligible range of the support of the joint density $f(y,z)$. To test H_0^*, we suggest the following test statistic

$$\widehat{\Gamma}_3 \equiv \Gamma(\widehat{f}, \widehat{F}) = \frac{1}{n}\sum_{t=1}^{n}\int \left|\widehat{m}_{h_1}(Z_t;\tau) - \overline{H}(\tau)\right|^2 a(X_t,Z_t)dG(\tau), \tag{4.6}$$

where $\overline{H}(\tau) = n^{-1}\sum_{t=1}^{n} H(Y_t+\tau)$, and $\widehat{m}_{h_1}(z;\tau)$ is the NW kernel estimator of $E(H(Y_t+\tau)|Z_t = z)$ using bandwidth h_1 and a product kernel of k, say. Note that we still use the bandwidth h_1, meaning that it satisfies Assumption A.2 (ii) where $d_1 = 0$ and the condition on h_2 is redundant. Let $\widetilde{B} \equiv {C_1}^{d_3}n^{-1}\sum_{t=1}^{n}\int\{\widehat{\sigma}_{h_1}^2(Z_t;\tau)a(Z_t)/\widehat{f}_{h_1}(Z_t)\}dG(\tau)$, $\widetilde{\sigma}_1^2 \equiv 2{C_3}^{d_3}n^{-1}\sum_{t=1}^{n}\int\int[\widehat{\sigma}_{h_1}^2(Z_t;\tau,\tau')]^2 a(Z_\tau)^2/\widehat{f}_{h_1}(Z_\tau)$ $dG(\tau)dG(\tau')$, where $\widehat{\sigma}_{h_1}^2(z;\tau)$ and $\widehat{\sigma}_{h_1}^2(z;\tau,\tau')$ are defined as $\widehat{\sigma}_{h_1}^2(x,z;\tau)$ and $\widehat{\sigma}_{h_1}^2(x,z;\tau,\tau')$. One can readily modify the other assumptions in Section 3 and show that

$$T_{3,n} \equiv nh_1^{d_3/2}\{\widehat{\Gamma}_3 - n^{-1}h_1^{-d_3}\ \widetilde{B}\}/\sqrt{\widetilde{\sigma}_1^2} \tag{4.7}$$

is asymptotically distributed as $N(0,1)$ under the null. For brevity, we don't repeat the argument.

4.2.5 Relation to the bootstrap

One can develop suitable versions of the bootstrap or other resampling methods which may improve the small sample performance of our test. In a similar but *i.i.d.* context, Härdle and Mammen (1993) show the validity of the wild bootstrap for obtaining the critical values for their goodness-of-fit test statistic. It is routine to justify that subsampling works in our context (Politis et al. 1999). Simulation results suggest that subsampling produces correct critical values but does not result in significant improvement despite its high computational cost.

5 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. We conduct simulation experiments extensively on testing for Granger causality and choosing the right order of nonlinear autoregressive (NAR) 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.

5.1 Testing for Granger noncausality

For Granger noncausality, our simulation covers three cases. We set $d_1 = d_2 = d_3 = 1$ in the first case, $d_1 = 2$ and $d_2 = d_3 = 1$ in the second case, and $d_1 = 3$ and $d_2 = d_3 = 1$ in the third case.

We use the following data generating processes (DGPs) for the first case:

DGP1: $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$, where $\{\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}\}$ are *i.i.d.* $N(0, I_3)$.

For DGP2 through DGP7, $W_t = (Y_{t-1}, Y_t, Z_{t-1})$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

DGP2: $Y_t = 0.5Y_{t-1} + \varepsilon_{1,t}$;

DGP3: $Y_t = 0.5Y_{t-1} + \alpha Z_{t-1} + \varepsilon_{1,t}$;

DGP4: $Y_t = 0.5Y_{t-1} + \alpha Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5: $Y_t = \alpha Y_{t-1} Z_{t-1} + \varepsilon_{1,t}$;

DGP6: $Y_t = 0.5Y_{t-1} + (0.5 + 0.5\alpha) Z_{t-1} \varepsilon_{1,t}$;

DGP7: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.5\alpha Z_{t-1}^2$; and

DGP8: $W_t = (Y_{t-1}, Y_t, Z_{t-1})$, where $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}$, $Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}$, $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + \alpha Z_{t-1}^2$, $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$, where $\{\varepsilon_{1,t},\ \varepsilon_{2,t}\}$ are *i.i.d.* $N(0, I_2)$ in DGPs 2-8 and $\alpha = 0.5$ in DGPs 3-8.

DGP1 and DGP2 allow us to examine the level of the test. DGPs 3-8 cover a variety of linear and nonlinear time series processes commonly used in time series analysis. Of these, DGPs 3-5 (resp. DGPs 6-8) are alternatives that allow us to study the power properties of our test for Granger-causality in the mean (resp. variance). DGP3 studies Granger linear causality in the mean whereas DGPs 4-5 study Granger nonlinear causality in the mean. In DGPs 6-8, $\{Z_t\}$ Granger-causes $\{Y_t\}$ only through the variance. A conditional mean-based Granger causality test, linear or nonlinear, may fail to detect such causality. Note that DGP7 is an ARCH-type specification and DGP8 specifies a bivariate GARCH process. Consequently, the study of such processes indicates whether our test may be applicable to financial time series. These DGPs are identical to those used in Su and White (2002).

We use a fourth order kernel in estimating all required quantities: $k(u) = (3 - u^2)\varphi(u)/2$, where $\varphi(u)$ is the *p.d.f.* of the standard normal distribution. We choose the weighting function $a(x, z)$ to be an indicator function on the compact set $A \equiv \{u = (u_1, ..., u_{d_1+d_3}) : |u_i| \leq 1.5,\ i = 1, ..., d_1 + d_3\}$ and choose both $g_0(\cdot)$ and $g(\cdot)$ (see Assumption $A.3$) to be a standard normal *p.d.f.*, which is in accordance with Assumption $A.3$. 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 parameters to choose, namely, the bandwidth sequences, h_1 and h_2, and it is difficult to pin down the optimal bandwidth sequences, we choose $h_1 = n^{-\frac{1}{6}}$ and $h_2 = cn^{-\frac{1}{5}}$ and let c vary over an interval to be specified later. Note that that h_1 (resp. h_2) is proportional to the optimal bandwidth for estimating the joint density $f(x, z)$ (resp. the marginal density $f(x)$) if a second order kernel were used instead. Both choices are allowed by Theorem 3.1 or Corollary 4.2. In some preliminary simulations, we find that the averages of c chosen by leave-one-out least squares cross validation for the marginal density $f(x)$ range between 0.7 to 1.1 across different DGPs. Also, we find through preliminary simulations that finite sample improvement can result if we ignore the second and the third bias correction terms in calculating $T_{1,n}$. One possible explanation is that these two terms are of smaller order than the first bias term in (3.1), and their estimation error is relatively large in small samples. For this reason, in all the following simulations, we only correct the first bias terms in $T_{1,n}$.

For DGPs 1 and 2, we first conduct 1000 repetitions for each sample size and each value of c under study. Specifically, we choose n to be 100 and 200 with c equally spaced on [0.5, 2], which includes the range of the averages of cross-validated values for c. Figure 1 and Figure 2 plot the empirical rejection frequency of our tests $T_{1,n}$ and $T_{2,n}$ as a function of c where the sample sizes are 100 and 200, respectively.

First, we comment on the test $T_{1,n}$. From the figures it appears that the level of $T_{1,n}$ is well behaved over a large range of values for c and the test is not sensitive to the choice of c . For both sample sizes, the 1% test is well behaved and the 5% and 10% tests are a little undersized. Note that the level curves are not smooth for either DGP; we think this is because our number of repetitions is somewhat small. Second, we comment on the test $T_{2,n}$. Surprisingly, for both sample sizes under study, the level curves are trending for all of the 1%, 5% and 10% tests. For the 1% test, it tends to be undersized for small values of c and oversized for large values of c. For the two other levels, the tests are undersized for the range of c under our study.

Third, when we increase the sample size to 500 or 1000, we find that the level of the test $T_{1,n}$ tends to decrease for fixed values of c and the level of the test $T_{2,n}$ diminishes even faster. So we propose to use a tuning parameter $a_n = \log(n/25)$ to prevent the level from diminishing too fast asymptotically. Namely, we use $h_1 = (na_n)^{-\frac{1}{6}}$ and $h_2 = c(na_n)^{-\frac{1}{5}}$ where we adjust c when the sample size doubles. For example, for $n \leq 200$, $c = 1.5$ is a good choice for a variety of DGPs that we study but don't report here. For $n = 500$ or so, $c = 2$ is a good choice. For $n = 1000$ or so (resp. 2000 or so), $c = 2.5$ (resp. 3.5) works well in simulations. This device is used in obtaining the empirical rejection frequency in the following study.

We now compare our tests with two tests proposed by Linton and Gozalo (1997) and a test by Su and White (2002). Linton and Gozalo (1997) base their nonparametric tests of conditional independence on the functional $A_n(w) = \left\{n^{-1} \sum_{t=1}^n 1(W_t \leq w)\right\} \left\{n^{-1} \sum_{t=1}^n 1(X_t \leq x)\right\} - \left\{n^{-1} \sum_{t=1}^n 1(X_t \leq x) 1(Y_t \leq y)\right\}$ $\times \left\{n^{-1} \sum_{t=1}^n 1(X_t \leq x) 1(Z_t \leq z)\right\}$, where $w = (x, y, z)$. Specifically, their test statistics are of the

Cramér von-Mises and Kolmogorov-Smirnov types: $CM_n = n^{-1}\sum_{t=1}^{n} A_n^2(W_t)$, $KS_n = \max_{1\leq t\leq n}|A_n(W_t)|$. The asymptotic null distribution of both test statistics is non-standard so that a local bootstrap procedure is needed to obtain the critical values.[17] To implement the tests, we set the number of bootstrap resamples as 100, use the product kernel of k as before and choose the bandwidth parameter, b_n, for the local bootstrap procedure according to $b_n = n^{-1/5}$.

Su and White (2002) base a test for conditional independence on the Hellinger distance between the two conditional densities $f(y|x,z)$ and $f(y|x)$, using the same bandwidth sequence h in estimating all required densities, namely, $f(x,y,z)$, $f(x,y)$, $f(x,z)$, and $f(x)$. Let $\widetilde{a} : R^{d_1} \times R^{d_2} \times R^{\ d_3} \to R$ be a nonnegative weighting function with compact support $\widetilde{A} \subset R^d$, where $d = d_1 + d_2 + d_3$, such that the joint density $f(x,y,z)$ is bounded below by b on $\widetilde{A}$. Specifically their $\widetilde{a}$ is defined as in our Footnote 12. Further define $\Gamma^* \equiv \frac{1}{n}\sum_{t=1}^{n}\left[1-\sqrt{\widehat{f}_h(X_t,Y_t)\widehat{f}_h(X_t,Z_t)/\widehat{f}_h(X_t,Y_t,Z_t)\widehat{f}_h(X_t)}\right]^2 \widetilde{a}(X_t,Y_t,Z_t)$, $\widetilde{B}_1 \equiv C_1^d \int_{\widetilde{A}} \widetilde{a}(x,y,z)d(x,y,z)$, $\widetilde{B}_2 \equiv C_1^{d_1+d_2} n^{-1}\sum_{t=1}^{n}\left[\widetilde{a}(X_t,Y_t,Z_t)/\widehat{f}_h(X_t,Y_t)\right]$, $\widetilde{B}_3 \equiv C_1{}^{d_1+d_3} n^{-1}\sum_{t=1}^{n}\left[\widetilde{a}(X_t,Y_t,Z_t)/\widehat{f}_h(X_t,Z_t)\right]$, and $\sigma_2^2 \equiv 2C_3^d\int_{\widetilde{A}}\widetilde{a}(x,y,z)^2 d(x,y,z)$. For $d_1 = d_2 = d_3 = 1$, the test statistic of Su and White (2002) can be written as

$$H_n \equiv nh^{d/2}\{4\Gamma^* - n^{-1}h^{-d}\widetilde{B}_1 + n^{-1}h^{-(d_1+d_2)}\widetilde{B}_2 + n^{-1}h^{-(d_1+d_3)}\widetilde{B}_3\}/\sqrt{\sigma_2^2},$$

which is asymptotically distributed as $N(0,1)$ under the null.

The simulations in Su and White (2002) are conducted for a variety of bandwidth sequences: $h = n^{-1/\delta}$, where $\delta = 8$, 8.5 and 9, but here we only report the case $\delta = 8.5$ because the resulting level and power tend to behave better than the other two cases. Table 1 reports the empirical rejection frequency of the five tests, namely, CM_n, KS_n, $H_n, T_{1,n}$ and $T_{2,n}$, for nominal sizes 5% and 10%. We set $c = 1.5$, 2, 2.5 for our tests $T_{1,n}$ and $T_{2,n}$ for sample size $n <= 200$, $= 500$ and $=1000$, respectively. Given the computational burden of our experiments, for all tests but H_n, there are 1000 Monte Carlo replications in the experiments for $n = 100$, 200 and 500, and 500 repetitions for $n = 1000$ when the null is true. The number of repetitions is 250 when the null is false. For the test H_n, the number of repetitions is 1000 for all sample sizes no matter whether the null is true or not.

From Table 1, we see that all three tests have reasonably good size properties for small to moderate sample sizes. For large samples, say $n = 1000$, the levels of all five tests are still well behaved. As far as power is concerned, both $T_{1,n}$ and $T_{2,n}$ exhibit significantly greater empirical power in detecting conditional dependence (Granger-causality) implied by DGPs 3 through 6 than H_n. In terms of empirical power, CM_n and KS_n are dominated by both $T_{1,n}$ and $T_{2,n}$ in all DGPs but DGP 3 and by H_n in DGPs 4 and 6-8. Even though there are some differences between $T_{1,n}$ and $T_{2,n}$ across different DGPs, we think that this is largely due to finite sample variation and the limited number of repetitions we use in our study. Note that both $T_{1,n}$ and $T_{2,n}$ have slightly worse power in detecting deviations from conditional independence that result from GARCH-type processes than H_n. For our small sample size (say $n = 100$), we find that the power of H_n from DGPs 7 and 8 is extremely sensitive to the choice of bandwidth. We emphasize that Su and White's test can work as well as our test in large samples ($n \geq 1000$), but this

[17] The setup of Linton and Gozalo (1997) is with i.i.d. data. Their bootstrap procedure can be easily modified to account for data dependence. The resulting bootstrap is a version of local bootstrap. See Paparoditis and Politis (2000) for more about local bootstrap.

does not necessarily hold for cases other than $d_1 = d_2 = d_3 = 1$ because of the disadvantage of using the same bandwidth sequence h in estimating all required densities in their approach.[18]

To see how the above tests are sensitive to the pseudo-true parameter α that controls the degree of dependence in DGPs 3-8, we choose 40 different $\alpha's$, equally spaced values on the compact interval [0 0.7], in all the above DGPs. For each value of α, we conduct 2000 repetitions and calculate the empirical rejection frequency for the tests CM_n, KS_n, $H_n, T_{1,n}$ and $T_{2,n}$.

Figure 3 reports the results for the above six sets of DGPs, with cases (a) through (f) corresponding to DGPs 3-8 when α varies over [0 0.7]. Also reported in Figure 3 is the empirical power function for the conventional linear Granger causality test with one lagged term. In the graphs, Lin stands for the linear causality test, and CM, KS, Chf1, Chf2 and Hel stand for the tests CM_n, KS_n, H_n, $T_{1,n}$and $T_{2,n}$, respectively. When the processes are truly linear, one expects the linear Granger test to be most powerful. This is verified in Figure 3 (a). Except in this case, one can see that the linear Causality test performs worst. From Figure 3, we see that $T_{1,n}$ and $T_{2,n}$ perform equally well (almost indistinguishable in most cases) and both outperform H_n for non-GARCH type processes and CM_n and KS_n for all DGPs but the linear one (DGP 3). For GARCH-type processes, H_n outperforms all other tests.

Since our test is fairly general, one can choose the weighting functions $g_0(\cdot)$ and $g(\cdot)$ (see Assumption $A.3$) differently. One set of weighting functions may have better power properties than another set in certain directions of deviations from the null hypothesis. Thus one may expect that for some weighting functions, our test can also outperform H_n for GRACH-type processes. We leave this investigation for future research.

Next, we test for Granger non-causality with $d_1 = 2$ and $d_2 = d_3 = 1$. We consider the following DGPs:

DGP1': $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$, where both $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}, \varepsilon_{3,t}\}$ are *i.i.d.* $N(0, I_2)$.

For DGP2' through DGP7', $W_t = ((Y_{t-1}, Y_{t-2}), Y_t, Z_{t-1})$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

DGP2': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \varepsilon_{1,t}$;

DGP3': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \alpha Z_{t-1} + \varepsilon_{1,t}$;

DGP4': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \alpha Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5': $Y_t = \alpha Y_{t-1}Z_{t-1} + 0.25Y_{t-2} + \varepsilon_{1,t}$;

DGP6': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + (0.3 + \alpha Z_{t-1})\varepsilon_{1,t}$;

DGP7': $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.5\alpha Z_{t-1}^2$; where $\alpha = 0.5$, $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is *i.i.d.* $N(0, I_2)$.

DGP8': same as DGP8.

Since the implementation of H_n becomes difficult here because of the previously mentioned bandwidth selection problem, we only study the finite sample behavior of the tests CM_n, KS_n, $T_{1,n}$and $T_{2,n}$. We use the same kernel and weighting functions and number of bootstrap resamples as in the first case. The only difference is that now we choose the bandwidth sequences differently. Specifically, we set $h_1 = 1.1n^{-\frac{1}{7}}$ and $h_2 = n^{-\frac{1}{6}}$ for the test $T_{1,n}$, $h_1 = 1.15(a_n n)^{-\frac{1}{7}}$ and $h_2 = (a_n n)^{-\frac{1}{6}}$ for the test $T_{2,n}$ and $b_n = n^{-1/6}$ for the CM_n and KS_n tests. Sample sizes n = 100, 200, 500 and 1000 are studied. When the null is true,

[18]The choice of bandwidth for these other cases becomes an extremely difficult task for the test of Su and White (2002).

there are 1000 Monte Carlo replications in the experiments for $n = 100$, 200 and 500, and 500 repetitions for $n = 1000$. The number of repetitions is 250 when the null is false.

Table 2 reports the empirical size and power properties of the four tests. As in the first case, both $T_{1,n}$ and $T_{2,n}$ dominate CM_n and KS_n for all nonlinear DGPs under investigation. The test $T_{1,n}$ tends to behave a little better than $T_{2,n}$ for small sample sizes; they perform almost identically well for moderate to large sample sizes. As the dimension of the conditioning variable increases, one might expect that the power of the tests is adversely affected. Table 2 suggests this conjecture is valid for small sample sizes but the effect of dimensionality is not severe. For the tests $T_{1,n}$ and $T_{2,n}$, the power is 1 or close to 1 for all DGPs under study when n is 500 for the 10% test, whereas for the tests CM_n and KS_n the power is less than 1 for GARCH-type processes even with $n = 1000$.

In the third case ($d_1 = 3$, $d_2 = d_3 = 1$), we use the following DGPs:

DGP1": $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$,where $\{\varepsilon_{1,t}\}$ is *i.i.d.* $N(0, I_3)$ and $\{\varepsilon_{2,t}, \varepsilon_{3,t}\}$ is *i.i.d.* $N(0, I_2)$.

For DGP2" through DGP7", $W_t = ((Y_{t-1}, Y_{t-2}, Y_{t-3}), Y_t, Z_{t-1})$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

DGP2": $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \varepsilon_{1,t}$;

DGP3": $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \alpha Z_{t-1} + \varepsilon_{1,t}$;

DGP4": $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \alpha Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5": $Y_t = \alpha Y_{t-1}Z_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \varepsilon_{1,t}$;

DGP6": $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + (0.3 + \alpha Z_{t-1})\varepsilon_{1,t}$;

DGP7": $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.125Y_{t-3}^2 + \alpha Z_{t-1}^2$; where $\alpha = 0.5$ and $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is *i.i.d.* $N(0, I_2)$.

DGP8": same as DGP8.

We use the same kernel and weighting functions as in the first case. The only difference is that we choose the bandwidth sequences differently and only consider sample sizes[19] $n = 200$, 500 and 1000. Specifically, we set $h_1 = 1.15n^{-\frac{1}{8}}$, $h_2 = n^{-\frac{1}{7}}$ for both $T_{1,n}$ and $T_{2,n}$ and $b_n = n^{-1/7}$ for the CM_n and KS_n tests. The number of repetitions is set as in the above two cases.

Table 3 reports the empirical size and power behavior of our tests. As we can see, the results are similar to the second case above. The tests $T_{1,n}$ and $T_{2,n}$ outperform the CM_n and KS_n tests for all nonlinear DGPs in terms of empirical power. The curse of dimensionality also exerts its expected effect.

5.2 Testing for the order of nonlinear time series

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 (1994a, 1994b), Tschernig and Yang (2000), Finkenstädt et al. (2001), Lobato (2003), among many others. In this subsection, we apply our test to determine the order d of a strictly stationary β-mixing univariate autoregressive time series model of the form

$$Y_t = g(Y_{t-1}, Y_{t-2}, ..., Y_{t-d}, \varepsilon_t), \tag{5.1}$$

[19] We don't consider the $n < 200$ case because we need to estimate nonparametrically a 4-dimensional density ($d_1 + d_3 = 4$) and this cannot be done with desirable accuracy with less than 200 observations.

where the function g is unknown and $\{\varepsilon_t\}$ is a noise process. If the model (5.1) is of the linear autoregressive form, then the Akaike (1972) information criterion and its variations provide an appropriate toolkit for determining the order of the process. Nevertheless, these methods are often not suitable if the process is nonlinear. Auestad and Tjostheim (1990) and Cheng and Tong (1992), among others, have proposed nonparametric order selection methods for the more general model

$$Y_t = E(Y_t|Y_{t-1}, Y_{t-2}, ..., Y_{t-d}) + \varepsilon_t \tag{5.2}$$

where the order of the process can be determined from estimating the conditional mean of Y_t given its past observations. Tjostheim and Auestad (1994a, 1994b) and Tschernig and Yang (2000) generalize the order determination of (5.2) to the case

$$Y_t = g_1(Y_{t-1}, Y_{t-2}, ..., Y_{t-p_1}) + g_2(Y_{t-1}, Y_{t-2}, ..., Y_{t-p_2})\varepsilon_t. \tag{5.3}$$

Using a conditional density approach, Finkenstädt et al. (2001) consider order selection under the general setting (5.1) in which the noise term ε_t may not be additive, which includes (5.3) as one special case.

Our work here is closely linked to this latter approach. The difference is that we offer a theory that pertains to the conditional distribution, not just the conditional location or multiplicative conditional standard deviation. As before, let $f(\cdot|\cdot)$ be the conditional density of one random variable given another. The null of interest is

$$H_0(d) : f(Y_t|Y_{t-1}, ..., Y_{t-d-1}) = f(Y_t|Y_{t-1}, ..., Y_{t-d}), \tag{5.4}$$

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 (5.4) is true, we say the nonlinear time series is of order d^*. In the following, we write $H_0(d) : d^* = d$ to represent (5.4). In the special case when $d = 0$, the test reduces to test of serial independence of first order: $H_0(0) : f(Y_t|Y_{t-1}) = f(Y_t)$ and the test statistic $T_{3,n}$ (see Eq. 4.7 in Section 4) is used.

We consider the following DGPs in our Monte Carlo study.

DGP9: $Y_t = \varepsilon_t$;

DGP10: $Y_t = 0.3Y_{t-1} + \varepsilon_t$;

DGP11: $Y_t = \sqrt{1 + 0.5Y_{t-1}^2}\varepsilon_t$;

DGP12: $Y_t = (-0.5Y_{t-1} + \varepsilon_t)1(Y_{t-1} \leq 1) + (0.4Y_{t-1} + \varepsilon_t)1(Y_{t-1} > 1)$;

DGP13: $Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$;

DGP14: $Y_t = 0.6\Phi(Y_{t-1})Y_{t-1} + \varepsilon_t$, where Φ represents the cumulative distribution of a standard normal distribution;

DGP15: $Y_t = 0.6Y_{t-1} + 0.35Y_{t-2} + \varepsilon_t$;

DGP16: $Y_t = -0.5Y_{t-1} + 0.5Y_{t-2}\left\{1 + \exp(-0.5Y_{t-1})\right\}^{-1} + \varepsilon_t$;

DGP17: $Y_t = 0.1\log(Y_{t-1}^2) + \sqrt{0.1 + 0.9Y_{t-2}^2}\varepsilon_t$;

DGPs 18-20: $Y_t = \exp(-Y_{t-1}^2) + |bY_{t-2}(16 - Y_{t-2})|\varepsilon_t$ with $\varepsilon_t = 0.05, 0.1$ and 0.2 respectively.

In DGPs 9-16, $\{\varepsilon_t\}$ are i.i.d. $N(0, 1)$. They are the i.i.d. sum of 30 uniformly independently distributed random variables each over the range [-0.1, 0.1] in DGP 17, and the i.i.d. sum of 10 uniformly independently distributed random variables each over the range [-1/7, 1/7] in DGP 18-20.

DGPs 9 through 13 are studied in Hong and White (2000) in testing for serial independence. DGPs 14 and 16 are studied in Lobato (2003) in testing for nonlinear autoregression. DGPs 17-20 are used in

Finkenstädt et al. (2001) in determining the order of nonlinear time series. DGP 9 is of order 0, DGPs 10-14 are of order one, and DGPS15-20 are of order 2. Note that all DGPs but DGPs 9, 10 and 15 are nonlinear in the mean or in the variance or in both.

We test for $H_0(d) : d^* = d$, where $d = 0, 1$ and 2 sequentially. For each case, we use sample sizes $n = 100$, 200 and 500 with 500 replications for each experiment to obtain the empirical rejection frequency of the 5% test in Table 4. In testing for $H_0(0)$: $d^* = 0$, only DGP1 satisfies the null. We choose the same kernel functions and weighting functions as in the previous subsection. The bandwidth is chosen to be $h_1 = 0.8n^{-1/5}$. From the top panel in Table 4, we see that our test can correctly reveal that the processes are not of order 0 for DGPs 10-20. The power is excellent for second order stochastic processes (DGPs 15-20) for sample sizes as small as 100.

DGPs 9-14 satisfy the null $H_0(1)$: $d^* = 1$ and all DGPs satisfy the null $H_0(2) : d^* = 2$. Nevertheless, in testing for the former null, DGP 9 is not among the most informative nulls in the sense the true order (0 here) of the underlying process is smaller than the tested order (1 here). Similarly, in testing for the latter null, DGPs 9-14 are not among the most informative null in the sense the true orders (0 or 1 here) of the underlying process are smaller than the tested order (2 here). To check the robustness of our testing procedure in the previous subsection, we use the same kernels, weighting functions and bandwidth sequences corresponding to the right dimension. For example, in testing the null $H_0(1)$: $d^* = 1$, $d_1 = d_2 = d_3$, and thus the bandwidth is chosen according to the first case above. For this case, we see from the middle panel in Table 4 that the test behaves reasonably well for both level and power for all DGPs under study. When testing for $H_0(2)$: $d^* = 2$, the level of our tests is well behaved for small sample sizes but tends to be inflated for most DGPs with large sample sizes. Finding a suitable explanation for this behavior is an interesting topic for future research.

6 Applications to financial time series

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, 1991; Sheedy 1998), researchers often neglect this when they test for causal relationships. As documented by Hiemstra and Jones (1994), all prior studies of causal relationship rely exclusively on the traditional linear Granger causality test, which unfortunately has little power in detecting nonlinear relationships as revealed in our simulation studies.

In this section, we first study the dynamic linkage between pairwise daily exchange rates across five industrialized countries, namely, Canada, France, Germany, Italy and the UK by using both our test $T_{1,n}$ and the traditional linear Granger causality test. Then with the same technique, we study the dynamic linkage between three US stock market price indices (Dow Jones 65 components, Nasdaq, and S&P 500) and the trading volumes in the New York Stock Exchange (NYSE), Nasdaq, and NYSE markets, respectively.

6.1 Application 1: exchange rates

The data for the daily exchange rates in the six industrialized countries are obtained from Datastream with the sample period from January 3rd, 1995 to December 17th, 2002 with 2077 observations total. The exchange rates are the local currency against the US dollar. Nevertheless, due to national holidays and certain other reasons, some observations of exchange rates are missing but entered in Datastream with the realizations from the previous trading day. Moreover, different nations have different national holidays and thus different missing observations. Because we do causality tests with exchange rates from pairwise countries, if the observation for one country is missing, we also delete that for the other country of the pair. This results in varying numbers of observations for each pairwise test. Following the literature, we let E_t stand for the natural logarithm of exchange rates multiplied by 100.

Since both the linear Granger causality test and our nonparametric test require that all time series involved be stationary and we are interested in the relation between the changes in the exchange rates, we first employ the augmented Dickey-Fuller test to check for stationarity of exchange rates (E_t) for all five countries under investigation. 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 ΔE_t in the following text. Next, since the appropriate formulation of a linear Granger causality analysis may need to incorporate an error correction term into the test if the underlying variables (pairwise E_t here) are cointegrated, we employ Johansen's likelihood ratio method to examine whether or not exchange rates for pairwise countries are cointegrated. The conclusion is that there is no cointegration between any pair of exchange rates. Consequently, no error correction terms need to be included in the linear Granger causality test.

6.1.1 Linear Granger causality test results

Let DX be the first differenced exchange rate in Country X and DY the first differenced exchange rate in Country Y. Loosely speaking, the time series $\{DX_t\}$ does not (linearly) Granger cause the time series $\{DY_t\}$ if the null hypothesis

$$H_{0,L} : \beta_1 = ... = \beta_{L_x} = 0 \tag{6.1}$$

holds in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + ... + \alpha_{L_y} DY_{t-L_y} + \beta_1 DX_{t-1} + ...\beta_{L_x} DX_{t-L_x} + \epsilon_t \tag{6.2}$$

where $\epsilon_t \,\tilde{}\, i.i.d.(0, \sigma^2)$ under $H_{0,L}$. An $F-$statistic can be constructed to check whether the null $H_{0,L}$ is true or not.

Nevertheless, in order to make a direct comparison with our nonparametric test for nonlinear Granger causality in the next subsection, we focus on testing for a variant[20] of $H_{0,L}$:

$$H^*_{0,L} : \beta = 0 \tag{6.3}$$

in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + ... + \alpha_{L_y} DY_{t-L_y} + \beta DX_{t-i} + \epsilon_t,\ i = 1, ..., L_x. \tag{6.4}$$

The results of linear Granger causality tests between pairwise exchange rates are given in Table 5, where we choose L_y to be 1, 2 or 3. When it is 1, we also choose L_x to be 1 so that we only check whether DX_{t-1} should enter (6.4) or not. This corresponds to the first row in each panel of Table 5.

[20]Clearly, the null $H^*_{0,L}$ is nested in the null $H_{0,NL}$. The rejection of $H^*_{0,NL}$ indicates the rejection of $H_{0,NL}$ but not the other way around.

When L_y is 2, we choose L_x to be 2. In this case, we check whether DX_{t-1} or DX_{t-2} (but not both) should enter (6.4) or not, which corresponds to the second and third rows in each panel of Table 5. The case for $L_y = 3$ is done analogously, corresponding to the fourth to sixth rows in each panel of Table 5.

To summarize the results in Table 5, we focus on the case of 5% significance level only. First of all, most causal links are associated with the Italian and French exchange rates. The Italian exchange rate is led by both the exchange rate in Canada and that in Germany at a variety of lags, and the exchange rate in France links closely with that in Germany and the UK. Secondly, no other causal links are detected by the linear Granger causality test at the 5% significance level. In particular, at one to two days' lag, no bidirectional causality is found and at three days' lag, there is one bidirectional link between the exchange rate in Germany and that in France. The question is whether there are some causal links that the linear causality test fails to detect and others that cannot be detected by our nonparametric test for nonlinear Granger causality.

6.1.2 Nonlinear Granger causality test results

To implement our test, we set all smoothing parameters according to those used in the simulations of Tables 2-4. The null of interest is now

$$H_{0,NL} : \Pr(f(DY_t|DY_{t-1}, ..., DY_{t-L_y}; DX_{t-1}, ..., DX_{t-L_x}) = f(DY_t|DY_{t-1}, ..., DY_{t-L_y})) = 1. \quad (6.5)$$

Due to the curse of dimensionality, we must choose L_y to be small. Specifically, we study the cases in which L_y =1, 2 and 3, respectively. Further, for each test we only include one lagged DX_t in the conditioning set. So we actually test for a variant[21] of $H_{0,NL}$:

$$H^*_{0,NL} : \Pr(f(DY_t|DY_{t-1}, ..., DY_{t-L_y}; DX_{t-i},) = f(DY_t|DY_{t-1}, ..., DY_{t-L_y})) = 1, \ i = 1, ..., L_x. \quad (6.6)$$

When L_y is 1, we also choose L_x to be 1 so that we only check whether DX_{t-1} should enter (6.6) or not. This corresponds to the first row in each panel of Table 6. When L_y is 2, we choose L_x to be 2. In this case, we check whether DX_{t-1} or DX_{t-2} (but not both) should enter (6.6) or not, which corresponds to the second and third rows in each panel of Table 6. The case for $L_y = 3$ is done analogously, corresponding to the fourth to sixth rows in each panel of Table 6.

The results in Table 6 are interesting. First, the first row in each panel suggests that if we set $L_y = 1$ and $L_x = 1$ in (6.6), our nonparametric test yields results different than the linear Granger causality test. At the 5% significance level, the test reveals causal links between pairwise exchange rates in three pairs of countries, each in both directions, although some links are stronger than others, according to the magnitudes of the test statistics. They are (France, Germany), (France, Italy) and (Germany, Italy). This suggests that at a one day lag, the exchange rates across the European countries (other than the UK) interact with each other strongly and in most cases the dependence between exchange rates is nonlinear, which the linear causality test fails to reveal.

One may argue that the above bidirectional causal relations are potentially spurious in that we have not included enough lags of the dependent variable in the test. To check this, we increase L_y to 2. The results in Table 6 reveal that the causal relations are robust. Further, more causal links are detected

[21] The null $H^*_{0,NL}$ is nested in the null $H_{0,NL}$. The rejection of $H^*_{0,NL}$ indicates the rejection of $H_{0,NL}$ but not the other way around. In this sense, our test is conservative.

at the two day lag. For example, the Canadian exchange rate is led by that in both Italy and Japan and the exchange rate in Italy is led also by that in Canada. The UK exchange rates tend to lead all other four exchanges by 1 or 2 days. When we increase L_y further to 3, the results suggest that further causal linkages between pairwise exchange rates may exist at various lags. One obvious reason for the failure of the linear Granger causality test in detecting such causal linkages is that exchange rates exhibit unambiguously nonlinear dependence. Hong (2001) studies the causal link between German and Japanese weekly exchange rates and concludes that for causality in mean, there exists only strong simultaneous interaction between the two exchange rates, whereas for causality in variance there are both simultaneous and asynchronous interactions between them.

To facilitate further comparison between the linear and nonlinear test results, we use boldfaced numbers in Table 6 to denote the causal relations which are revealed by our nonparametric causality test but not by the linear causality test at the 5% significance level. Similarly, the boldfaced numbers in Table 5 denote the causal relations which are revealed by the linear causality test but not by our nonparametric causality test at the 5% significance level. From Table 5, we see that our nonparametric test fails to reveal 6 linear causal relations at various lags, which suggests that our test can have less power than a test specifically designed to exploit linearity. From Table 6, we see that our nonparametric test reveals 40 linear causal relations that can't be detected by the linear Granger causality test, which is strong evidence in favor of the nonlinear dependence between exchange rates.

6.2 Application 2: stock prices and trading volumes

Daily data for the three major stock market price indices and trading volumes have been obtained from Yahoo Finance with the sample period from January 2nd, 1995 to January 10th, 2003. After excluding weekends and holidays, the total numbers of observations are 2022 for the Dow Jones 65 composite and Nasdaq series and 2020 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 in the following text. Next, we employ Johansen's likelihood ratio method to examine whether P_t and V_t are cointegrated or not. The conclusion is that there is no cointegration between them. Consequently, no error correction term needs to be included in the linear Granger causality test.

6.2.1 Linear Granger causality test results

We first let ΔP_t and ΔV_t play the roles of DX_t and DY_t in (6.4) and test the null that stock price does not linearly Granger cause trading volume. Then we reverse their roles to test for the null that trading volume does not linearly Granger cause stock price. The results of the linear causality test between stock prices and volumes are given in Panel A of Table 7. At all levels of L_y, we find causal links from stock prices to trading volumes for the Nasdaq and S&P 500 data but not for the Dow Jones data at the 5% significance level. Unambiguously, no causality from trading volume to stock price is revealed by the linear causality test.[22]

6.2.2 Nonlinear Granger causality test results

We first let ΔP_t and ΔV_t play the roles of DX_t and DY_t in (6.6) and test the null that stock price does not Granger cause trading volume. Then we reverse their roles to test the null that trading volume does not Granger cause stock price. The results for our nonparametric test are reported in Panel B of Table 7. From Panel B, we see that our nonparametric test reveals bidirectional causal relations between stock prices and trading volumes at various levels of L_y for all stock series. For $L_y = 1$ and 2, bidirectional causal links also exist between the Dow Jones price index and the NYSE trading volume.

To facilitate the comparison between the linear and nonlinear test results, we use boldfaced numbers in Panel A of Table 7 to denote the causal relations which are revealed by the linear Granger causality test but not by our nonparametric test at the 5% significance level. Similarly, the boldfaced numbers in Panel B of Table 7 denote the causal relations that are revealed by our nonparametric test but not by the linear Granger causality test at the 5% significance level. From Panel A, we see that our nonparametric test fails to detect two causal relations revealed by the linear causality test. From Panel B, one can tell that our nonparametric test can reveal 10 extra causal relations at various lags between stock prices and trading volumes besides those revealed by the linear causality test, strong evidence in favor of the nonlinear dependence between the two economic variables. One obvious reason for the failure of the linear Granger causality test in detecting such causal links is that trading volumes may only have nonlinear predictive power for stock returns.

7 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 Su and White (2002), this resolves the long standing need in econometrics for an asymptotic theory for a practical and powerful nonparametric test for conditional independence. It is directly applicable to testing for Granger non-causality for strictly stationary absolutely regular processes. It is applicable for both discretely and continuously valued random variables in the time series context. Moreover, it is nuisance-parameter free in the situation where an unknown finite-dimensional parameter can be consistently estimated at certain rates.

We also conduct some Monte Carlo experiments for our test. We find that the size of our test behaves well and its power is considerable for a variety of DGPs under study, including both Granger causality in the mean (linear or nonlinear) and Granger causality in the variance. Moreover, we compare our test to

[22] As is done for the case of exchange rates, we also conduct the linear Granger causality test for the null (6.1) by using the BIC and AIC to choose the numbers of lags, L_x and L_y, the maxima of which are set to be 10. Accoding to the BIC, only one linear causal relation is found at the 5% significance level. That is, the S&P 500 stock price index tends to lead the NYSE volume. According to the AIC, all the three stock price indices tend to lead the corresponding trading volumes at the 5% significance level. No other causal links are found.

that of Su and White (2002) and that of Linton and Gozalo (1997). We find that our test outperforms the former significantly in all cases but the GARCH-type processes and dominates the latter in all nonlinear DGPs in terms of empirical power.

To improve the asymptotic approximation to the finite sample distribution of the test statistic, one could consider higher order refinements. However, it is well known that estimation of higher order refinements is tedious, and this may not necessarily provide a sufficiently good approximation in finite samples. An alternative procedure is to use some resampling techniques. The bootstrap may offer a finite sample improvement. If the distributions of our test statistic and its bootstrap analogue admit Edgeworth expansion (Hall 1992), we conjecture that the bootstrap distribution approximates the null distribution of the test statistic with an error rate that can be arbitrarily close to $O(n^{-1/2})$, and this will significantly improve the normal approximation rate. Recently, Nishiyama and Robinson (2000) and Linton (2002) establish the validity of Edgeworth expansion for a degenerate U-statistic with variable kernel. This suggests that a rigorous proof for establishing the validity of Edgeworth expansion in our context should be possible. Also, such an expansion will offer a solution to the choice of optimal bandwidth and we hope to tackle this technical issue in the future.

There are certain issues that remain. First of all, our theory requires stationarity of the underlying process, which may not be satisfied in some applications in economics. In particular, structural breaks can invalidate our test procedure. Secondly, our test does not allow heterogeneity which may be prevalent in cross-section data. Extension to this case can be accomplished by using a CLT for degenerate U-statistics for independent but not identically distributed data (e.g. de Jong 1987). Third, as remarked earlier, there is also room to extend our test to the case in which some of the random variables are nonparametrically estimated. For this, stochastic equicontinuity (Andrews 1994, 1995) is an indispensable tool. Fourth, there is a lot of latitude concerning the choice of weighting functions, and further research is required to determine what weighting functions for our test can improve power against deviations from conditional independence in GARCH-type processes. Finally, we have not addressed the optimality of the test proposed here. The work of Andrews and Ploberger (1994), Bierens and Ploberger (1997) and Boning and Sowell (1999) should provide useful tools in answering this issue.

Appendix

A Some Useful Definitions, Lemmas and Theorems

In this appendix, we introduce a definition, two lemmas and one theorem which are used in the proof of the main theorems and propositions in the text.

Definition A.1 *Let $\{U_t,\, t \geq 0\}$ be a $d-$dimensional strictly stationary stochastic process and $\mathcal{F}_s^t$ denote the sigma algebra generated by $(U_s, ..., U_t)$ for $s \leq t$. The process is called $\beta-$mixing or absolutely regular, if as $m \to \infty$,*

$$\beta_m \equiv \sup_{s \in N} E[\sup_{A \epsilon \mathcal{F}_{s+m}^{\infty}} \{|P(A|\mathcal{F}_{-\infty}^{s}) - P(A)|\} \to 0.$$

The following Lemma is due to Yoshihara (1976); see also Li (1999).

Lemma A.2 *Let $\{U_t,\, t \geq 0\}$ be a $d-$dimensional stochastic process satisfying Assumption A.1(i) in the text. Let $h(v_1, ..., v_k)$ define a Borel measurable function on R^{kd} such that for some $\delta > 0$ and given j,*

$M \equiv \max\{\int_{R^{kd}} |h(v_1, ..., v_k)|^{1+\delta} dF(v_1, ..., v_k), \int\int_{R^{kd}} |h(v_1, ..., v_k)|^{1+\delta} dF^{(1)}(v_1, ..., v_j) dF^{(2)}(v_{j+1}, ..., v_k)\}$

exists. Then

$|\int_{R^{kd}} h(v_1, ..., v_k) dF(v_1, ..., v_k) - \int\int_{R^{kd}} h(v_1, ..., v_k) dF^{(1)}(v_1, ..., v_j) dF^{(2)}(v_{j+1}, ..., v_k)|$
$\leq 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)},$

where $m \equiv i_{j+1} - i_j$, F, $F^{(1)}$ and $F^{(2)}$ are distributions of random vectors $(U_{i_1}, ..., U_{i_k}), V_1 \equiv (U_{i_1}, ..., U_{i_j})$ and $V_2 \equiv (U_{i_{j+1}}, ..., U_{i_k})$, respectively; and $i_1 < i_2 < ... < i_k$.

The next lemma is due to Yoshihara (1989).

Lemma A.3 *Let h be defined as above; then*

$E|E[h(V_1, V_2)|V_1] - E_{V_1} h(V_1, V_2)| \leq 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)},$

where $E_{V_1} h(V_1, V_2) \equiv H(V_1)$ with $H(v_1) \equiv E[h(v_1, V_2)]$.

Now, let $g_n(\cdot)$ and $h_n(\cdot, \cdot)$ be Borel measurable functions on R^d and $R^d \times R^d$, respectively. Suppose $E[g_n(U_0)] = 0$, $E[h_n(U_{0,} v)] = 0$ and $h_n(u, v) = h_n(v, u)$ for all $(u, v) \in R^d \times R^d$. Define

$\mathcal{G}_n \equiv n^{-1/2} \sum_{i=1}^{n} g_n(U_i)$, and

$\mathcal{H}_n \equiv n^{-1} \sum_{1 \leq i < j \leq n} [h_n(U_i, U_j) - Eh_n(U_i, U_j)]$.

Clearly, $\mathcal{G}_n$ and $\mathcal{H}_n$ are degenerate $U-$statistics of respective orders 1 and 2. Let $p > 0$ and $\{\overline{U}_t,\, t \geq 0\}$ be an $i.i.d.$ sequence where $\overline{U}_0$ is an independent copy of U_0. Define

$u_n(p) \equiv \max\{\max_{1 \leq i \leq n} ||h_n(U_i, U_0)||_p, ||h_n(U_0, \overline{U}_0)||_p\},$

$v_n(p) \equiv \max\{\max_{1 \leq i \leq n} ||G_{n0}(U_i, U_0)||_p, ||G_{n0}(U_0, \overline{U}_0)||_p\},$

$w_n(p) \equiv ||G_{n0}(U_0, U_0)||_p,$

$z_n(p) \equiv \max_{0 \leq i \leq n} \max_{1 \leq j \leq n} \{||G_{nj}(U_i, U_0)||_p, ||G_{nj}(U_0, U_i)||_p, ||G_{nj}(U_0, \overline{U}_0)||_p\},$

where $G_{n,i}(u, v) \equiv E[h_n(U_i, u) h_n(U_0, v)]$, and $|| \cdot ||_p \equiv \{E| \cdot |^p\}^{1/p}$.

Theorem A.4 ***(Tenreiro 1997).*** *Given the above notation, suppose there exists* $\delta_0 > 0$, $\gamma_0 < 1/2$ *and* $\gamma_1 > 0$ *such that*

(i) $||g_n(U_0)||_4 = O(1)$;

(ii) $E[g_n(U_i)g_n(U_0)] = c_i + o(1),\ i = 0, 1, 2, ...$;

(iii) $u_n(4 + \delta_0) = O(n^{\gamma_0})$;

(iv) $v_n(2) = o(1)$;

(v) $w_n(2 + \delta_0/2) = o(n^{1/2})$;

(vi) $z_n(2)n^{\gamma_1} = O(1)$;

(vii) $E[h_n(U_0, \overline{U}_0)]^2 = 2\widetilde{\sigma}_2^2 + o(1)$.

Then $(\mathcal{G}_n, \mathcal{H}_n)$ *is asymptotically normally distributed with mean zero and covariance matrix* $\begin{bmatrix} \widetilde{\sigma}_1^2 & 0 \\ 0 & \widetilde{\sigma}_2^2 \end{bmatrix}$, *where* $\widetilde{\sigma}_1^2 \equiv c_0 + 2\sum_{i=1}^{\infty} c_i$.

B Proof Theorem 3.1

We begin by studying the asymptotic properties of the functional Γ evaluated at $(\widehat{f}, F)$, using the functional delta method. The only difference between $\Gamma(\widehat{f}, F)$ and $\widehat{\Gamma} \equiv \Gamma(\widehat{f}, \widehat{F})$ is that the latter is an average over the empirical distribution function $\widehat{F}$ instead of F. We will show in Lemma B.6 that this difference is inconsequential for the asymptotic distribution of the test statistic. To bound the remainder term in the functional expansion of $\Gamma(\widehat{f}, F)$, we define the seminorms

$$||\mathrm{g}_1||_s \equiv \max\left(\sup_{(x,z)\in A,\ \tau\in S} |\int H(y+\tau)g_1(x,y,z)dy|, \sup_{(x,z)\in A} |g_1(x,z)| \right),$$

$$||\mathrm{g}_2||_s \equiv \max\left(\sup_{x\in A\cap R^{d_1}, \tau\in R^{d_2}} |\int H(y+\tau)g_2(x,y)dy|, \sup_{x\in A\cap R^{d_1}} |g_2(x)| \right)$$

where by convention $g_1(x,z) \equiv \int g_1(x,y,z)dy$ and $g_2(x) \equiv \int g_2(x,y)dy$.

Define $\Omega_1 \equiv \{g_1 : R^{d_1+d_2+d_3} \to R,\ g_1$ is bounded, $\int g_1 = 0$, and $||g_1||_s < b/2\}$, and $\Omega_2 \equiv \{g_2 : R^{d_1+d_2} \to R,\ g_2$ is bounded, $\int g_2 = 0$, and $||g_2||_s < b/2\}$. Throughout this appendix, C denotes a generic constant which may vary from one place to another. The bar notation denotes 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 $\{W_t, t \geq 0\}$. See Lemmas $B.5$ and $B.7$ for details.

One of the main ingredients in the proof of Theorem 3.1 is the functional expansion of Γ, summarized as follows.

Lemma B.1 *Let* F *be a c.d.f. on* R^d. *Let* g_1 *and* g_2 *belong to* Ω_i, $i = 1, 2$, *respectively. Then under Assumption* $A.1(ii)$, *A.3 and* H_0, $\Gamma(\cdot, F)$ *has the following expansion:*

$$\Gamma(f+g;F) = \int\int\left\{\int \alpha(y;x,z,\tau)g_1(x,y,z)dy - \int \beta(y;x,\tau)g_2(x,y)dy\right\}^2 a(x,z)dF(x,z)dG(\tau) + R(g,F)$$

where

$\sup\{\ |R(g,F)|/(||g_1||_s^3 + ||g_2||_s^3) : (g_1, g_2) \in \Omega_1 \times \Omega_2\} < \infty.$

Proof. Define

$$\varphi(\lambda;x,z,\tau)\equiv\frac{\int H(y+\tau)[f(x,y,z)+\tau g_1(x,y,z)]dy}{f(x,z)+\tau g_1(x,y)}-\frac{\int H(y+\tau)[f(x,y)+\tau g_2(x,y))]dy}{f(x)+\tau g_2(x)},$$

and

$$\Psi(\lambda)=\int\int\varphi(\lambda;x,z,\tau)^2a(x,y)dF(x,y)dG(\tau),$$

where (g_1,g_2) are such that $(\lambda g_1,\lambda g_2)\in\Omega_1\times\Omega_2$ for all $0\leq\lambda\leq 1$. From the explicit expression for $\Psi(\lambda)$ and the properties of f and g's, it follows that Ψ is three times continuously differentiable in λ on [0,1]. Applying Taylor's formula with Lagrange remainder to Ψ, we get:

$$\Psi(\lambda)=\Psi(0)+\lambda\Psi'(0)+\lambda^2\Psi''(0)/2+\lambda^3\Psi^{'''}(\lambda^*)/6,$$

where $0\leq\lambda^*\leq\lambda$. Note that $\Psi(0)=0$ under H_0 and it is immediate to compute

$$\Psi'(\lambda)=\int\int 2\varphi(\lambda;x,z,\tau)\frac{\partial\varphi(\lambda;x,z,\tau)}{\partial\lambda}a(x,y)dF(x,y)dG(\tau),$$

and

$$\Psi''(\lambda)=\int\int 2\left\{\left[\frac{\partial\varphi(\lambda;x,z,\tau)}{\partial\lambda}\right]^2+\varphi(\lambda;x,z,\tau)\frac{\partial^2\varphi(\lambda;x,z,\tau)}{\partial\lambda^2}\right\}a(x,y)dF(x,y)dG(\tau).$$

Under the null, $\Psi'(0)=0$ and

$$\Psi''(0)=2\int\int\left\{\int\alpha(y;x,z,\tau)g_1(x,y,z)dy-\int\beta(y;x,\tau)g_2(x,y)dy\right\}^2\times a(x,z)dF(x,z)dG(\tau).$$

One can characterize the remainder term by first computing $\Psi^{'''}(\lambda)$. The explicit formula for $\Psi^{'''}(\lambda)$ is lengthy. By the Cauchy-Schwartz inequality and Assumption A.2 (ii) and A.3, we can bound this remainder by a factor of $(||g_1||_s^3+||g_2||_s^3)$.

Consequently, for $\lambda=1$, we have obtained that under the null

$$\begin{aligned}\Psi(1) \quad &= \quad \int\int\left\{\int\alpha(y;x,z,\tau)g_1(x,y,z)dy-\int\beta(y;x,\tau)g_2(x,y)dy\right\}^2 a(x,z)dF(x,z)dG(\tau)\\ &\quad +O(||g_1||_s^3+||g_2||_s^3),\end{aligned}$$

and the lemma follows. ■

Lemma B.2 *Under Assumptions A.1 – A.3,*

$$\|\widehat{m}(x,z;\tau)-m(x,z;\tau)\|_\infty\equiv\sup_{(x,z)\in A,\tau\in S}|\widehat{m}(x,z;\tau)-m(x,z;\tau)|=O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}+h_1^r),\tag{B.1}$$

and

$$\|\widehat{m}(x;\tau)-m(x,\tau)\|_\infty\equiv\sup_{x\in A\cap R^{d_1},\tau\in S}|\widehat{m}(x;\tau)-m(x;\tau)|=O_p(n^{-1/2}h_2^{-d_1/2}\sqrt{\ln n}+h_2^r).\tag{B.2}$$

Proof. We only prove (B.1). Let $M(x,z;\tau) \equiv \int H(y+\tau)f(x,y,z)dy$ and $\widehat{M}_{h_1}(x,z;\tau) \equiv \widehat{m}(x,z;\tau)\widehat{f}_{h_1}(x,z)$. Note that $m(x,z;\tau) = M(x,z;\tau)/f(x,z)$, so by the triangle inequality,

$$\begin{aligned}\|\widehat{m}(x,z;\tau) - m(x,z;\tau)\|_\infty \quad &\leq \quad \frac{\sup\limits_{(x,z)\in A,\tau\in S} \widehat{m}(x,z;\tau)}{\inf\limits_{(x,z)\in A} f(x,z)} \sup_{(x,z)\in A} \left| f(x,z) - \widehat{f}_{h_1}(x,z)] \right| \\ &+ \frac{1}{\inf\limits_{(x,z)\in A} f(x,z)} \sup_{(x,z)\in A,\tau\in S} \left| \widehat{M}_{h_1}(x,z;\tau) - M(x,z;\tau) \right|.\end{aligned}$$

Note that $\inf\limits_{(x,z)\in A} f(x,z) \geq b > 0$ by assumption and that $\beta-$mixing implies $\alpha-$mixing. The bound on the L_∞ deviations of the Nadaraya-Watson kernel density estimator of $f(x,z)$ over a compact subset A of its support is standard (e.g., Theorem 4.3 in Liebscher 1996) and given by

$$\left\| \widehat{f}_{h_1}(x,z) - f(x,z) \right\|_\infty \equiv \sup_{(x,z)\in A} \left| \widehat{f}_{h_1}(x,z) - f(x,z) \right| = O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n} + h_1^r). \tag{B.3}$$

So it suffices to show

$$\sup_{(x,z)\in A,\tau\in S} \widehat{m}(x,z;\tau) = O_p(1), \tag{B.4}$$

and

$$\sup_{(x,z)\in A,\tau\in S} \left| \widehat{M}_{h_1}(x,z;\tau) - M(x,z;\tau) \right| = O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n} + h_1^r). \tag{B.5}$$

Let $C_0 = \int |k(u)|du < \infty$ (by assumption A.2(ii)), and $C = 2(C_0^{d_1+d_3}/b) \max\limits_{(x,z)\in R^{d_1+d_3}} f(x,z)$. Noticing that H is uniformly bounded by 1, we have

$$P(\sup_{(x,z)\in A,\tau\in S} |\widehat{m}(x,z;\tau)| > C) \leq P(\sup_{(x,z)\in A} \frac{n^{-1}\sum_{t=1}^n |K_{h_1}(x-X_t, z-Z_t)|}{\widehat{f}_{h_1}(x,z)} > C).$$

The right side converges to 0 by the LLN, the fact that $f(x,z)$ is bounded on $R^{d_1+d_3}$ (implied by Assumption A.1 (ii)) and $\widehat{f}_{h_1}(x,z) > b/2$ for sufficiently large n. To show (B.5), write $\sup\limits_{(x,z)\in A,\tau\in S} |\widehat{M}_{h_1}(x,z;\tau) - M(x,z;\tau)| \leq \sup\limits_{(x,z)\in A,\tau\in S} \left| \widehat{M}_{h_1}(x,z;\tau) - E\widehat{M}_{h_1}(x,z;\tau) \right| + \sup\limits_{(x,z)\in A,\tau\in S} \left| E\widehat{M}_{h_1}(x,z;\tau) - M(x,z;\tau) \right|$. By standard argument and Assumptions $A.1 - A.3$, the second term of the latter expression is $O(h_1^r)$. For the first term, using the uniform boundedness of H, one can modify the proof of Theorem 3.2 in Bosq (1996) (see also Collomb and Härdle 1986 for $\phi-$mixing processes) and show it is $O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n})$. Thus we have shown (B.1). ■

Lemma B.3 *Under Assumptions $A.1 - A.3$ and H_0, we have for any c.d.f. F,*

$$\begin{aligned}\Gamma(\widehat{f},F) \quad &= \quad \int\int \left\{ \int \alpha(y;x,z,\tau)\widehat{f}_{h_1}(x,y,z)dy - \int \beta(y;x,\tau)\widehat{f}_{h_2}(x,y)dy \right\}^2 a(x,z)dF(x,z)dG(\tau) \\ &\quad + O_p(\left\|\widehat{f}_{h_1} - f\right\|_s^3 + \left\|\widehat{f}_{h_2} - f\right\|_s^3).\end{aligned}$$

Proof. We are going to apply Lemma $B.1$ with $g_1(x,y,z) = \widehat{f}_{h_1}(x,y,z) - f(x,y,z)$, $g_2(x,y) = \widehat{f}_{h_2}(x,y) - f(x,y)$. First note that by the triangle inequality,

$$\begin{aligned}
&\sup_{(x,z)\in A,\tau\in S}\left|\int H(y+\tau)\left[\widehat{f}_{h_1}(x,y,z)-f(x,y,z)\right]dy\right| \\
\leq\; &\sup_{(x,z)\in A,\tau\in S}\left|\frac{1}{n}\sum_{j=1}^{n}K_{h_1}(x-X_j,z-Z_j)\int[H(y+\tau)-H(Y_j+\tau)]K_{h_1}(y-Y_j)dy\right| \\
&+\sup_{(x,z)\in A,\tau\in S}\left|\frac{1}{n}\sum_{j=1}^{n}K_{h_1}(x-X_j,z-Z_j)H(Y_j+\tau)]\int K_{h_1}(y-Y_j)dy-\int H(y+\tau)f(x,y,z)]dy\right| \\
\equiv\; & A_{n,1}+A_{n,2},
\end{aligned}$$

where $A_{n,1} = O_p(h_1^r)$ because $\int[H(y+\tau)-H(Y_j+\tau)]K_{h_1}(y-Y_j)dy = \int[H(Y_j+\tau+h_1u)-H(Y_j+\tau)]K(u)du = h_1^r C_0^{d_2}\frac{\partial H^{|r|}(Y_j+\tau)}{\partial Y^{|r|}} + o_p(h_1^r)$ uniformly in τ on S, and $\frac{1}{n}\sum_{j=1}^{n}K_{h_1}(x-X_j,z-Z_j)\frac{\partial H^{|r|}(Y_j+\tau)}{\partial Y^{|r|}} = O_p(1)$ for $(x,z)\in A$ uniformly in τ on S. For the second term, we have, $A_{n,2} = \sup_{(x,z)\in A,\tau\in S}|\frac{1}{n}\sum_{j=1}^{n}K_{h_1}(x-X_j,z-Z_j)H(Y_j+\tau)]\int K_{h_1}(y-Y_j)dy - m(x,z;\tau)f(x,z)| = \sup_{(x,z)\in A,\ \tau\in S}|m(x,z;\tau)\widehat{f}_{h_1}(x,z) - m(x,z;\tau)$ $\times f(x,z)| = \left\|\widehat{f}_{h_1}(x,z)-f(x,z)\right\|_\infty \|m(x,z;\tau)\|_\infty = O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}+h_1^r)$ by (B.3). Consequently $\left\|\widehat{f}_{h_1}-f\right\|_s = O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}+h_1^r) = o_p(1)$ under our assumptions.

Note that under Assumption 2.(ii), we have $h_1^{d_1} \gg h_2^{d_1} \gg h_1^{(d_1+d_3)}$ so both the asymptotic bias and variance of $\widehat{f}_{h_2}(x)$ are smaller than those of $\widehat{f}_{h_1}(x,z)$ and thus all the norms involving $\widehat{f}_{h_2}(x)-f(x)$ are strictly smaller than those involving $\widehat{f}_{h_1}(x,z)$ $-f(x,z)$.

Lastly let $F \equiv \left\{\left\|\widehat{f}_{h_1}(x,z)-f(x,z)\right\|_s \geq b/2,\ \left\|\widehat{f}_{h_2}(x)-f(x)\right\|_s \geq b/2\right\}$. Then $\Pr[F]\to 0$ so that $\Pr[(g_1,g_2)\in\Omega_1\times\Omega_2]\to 1$ and the result follows. ■

To state and prove the next four lemmas, we introduce some new notation to facilitate the presentation. Denote

$$\begin{aligned}
I_n &\equiv \int\int\left\{\int\alpha(y;x,z,\tau)\widehat{f}_{h_1}(x,y,z)dy - \int\beta(y;x,\tau)\widehat{f}_{h_2}(x,y)dy\right\}^2 a(x,z)dF(x,z)dG(\tau) \\
&\equiv \int\int r_n(x,z;\tau)^2 a(x,z)dF(x,z)dG(\tau).
\end{aligned}$$

We can decompose I_n as follows:

$$\begin{aligned}
I_n = &\int\int[r_n(x,z;\tau)-Er_n(x,z;\tau)]^2 a(x,z)dF(x,z)dG(\tau) \\
&+2\int\int[r_n(x,z;\tau)-Er_n(x,z;\tau)]Er_n(x,z;\tau)a(x,z)dF(x,z)dG(\tau) \\
&+\int\int[Er_n(x,z;\tau)]^2 a(x,z)dF(x,z)dG(\tau)
\end{aligned}$$

so

$$\begin{aligned}
I_n - E[I_n] = &2\int\int[r_n(x,z;\tau)-Er_n(x,z;\tau)]Er_n(x,z;\tau)a(x,z)dF(x,z)dG(\tau) \\
&+\int\int\{[r_n(x,z;\tau)-Er_n(x,z;\tau)]^2 - E[r_n(x,z;\tau)-Er_n(x,z;\tau)]^2\}a(x,z)dF(x,z)dG(\tau).
\end{aligned}$$

For $w=(x',y',z')\in R^{d_1}\times R^{d_2}\times R^{d_3}$ and $v=(\widetilde{x},\widetilde{y},\widetilde{z})\in R^{d_1}\times R^{d_2}\times R^{d_3}$, we define $R(w;x,z,\tau)\equiv \int\alpha(y;x,z,\tau)K_{h_1}(x-x',y-y',z-z')dy - \int\beta(y;x,\tau)K_{h_2}(x-x',y-y')dy$,

$\widetilde{R}(w;x,z,\tau)\equiv R(w;x,z,\tau)-ER(W_1;x,z,\tau)$,

$G_n(w) \equiv \int\int \widetilde{R}(w;x,z,\tau)h_1^{-r}Er_n(x,z;\tau)a(x,z)dF(x,z)dG(\tau)$,

and

$H_n(w,v) \equiv h_1^{(d_1+d_3)/2}\int\int \widetilde{R}(w;x,z,\tau)\widetilde{R}(v;x,z,\tau)a(x,z)dF(x,z)dG(\tau)$.

Note that we have suppressed the dependence of R and $\widetilde{R}$ on n.

Then we can write

$$\begin{aligned} I_n - E[I_n] &= 2n^{-1/2}h_1^r\left\{n^{-1/2}\textstyle\sum_{i=1}^n G_n(W_i)\right\} \\ &\quad +2n^{-1}h_1^{-(d_1+d_3)/2}\left\{n^{-1}\textstyle\sum_{1\le i<j\le n}[H_n(W_i,W_j)-EH_n(W_i,W_j)]\right\} \\ &\quad + n^{-1}h_1^{-(d_1+d_3)/2}\left\{n^{-1}\textstyle\sum_{i=1}^n[H_n(W_i,W_i)-EH_n(W_i,W_i)]\right\} \\ &\equiv 2n^{-1/2}h_1^rU_{n,1}+2n^{-1}h_1^{-(d_1+d_3)/2}U_{n,2}+n^{-1}h_1^{-(d_1+d_3)/2}U_{n,3}. \end{aligned}$$

We have $U_{n,3}=o_p(1)$ by the WLLN under Assumptions $A.1-A.3$ (e.g. White, 2000). We are going to use Theorem $A.4$ to study the asymptotic normality of $U_{n,1}$ and $U_{n,2}$ with $G_n(\cdot)$ and $H_n(\cdot,\cdot)$ in place of $g_n(\cdot)$ and $h_n(\cdot,\cdot)$ in the theorem, respectively. Moreover, the term involving $U_{n,1}$ is asymptotically negligible given our restriction on the choice of bandwidth and the order of the kernel (Lemma $B.4$) . To get the asymptotic distribution of our test statistic, we need to calculate both asymptotic variance (Lemma $B.5$) and bias correction terms (Lemma $B.6$).

Lemma B.4 *Under Assumptions $A.1-A.3$ and H_0, $U_{n,1}\xrightarrow{d}N(0,\widetilde{\sigma}^2)$, where $\widetilde{\sigma}^2\equiv Var(\gamma(W_0))\ +2\sum_{j=1}^{\infty}Cov(\gamma(W_j),\gamma(W_0))$ with $\gamma(\cdot)$ defined by Equation (B.6).*

Proof. First, we find the limit of $h_1^{-r}Er_n(x,z;\tau)$:

$$\begin{aligned} h_1^{-r}Er_n(x,z;\tau) &= h_1^{-r}ER(W_i;x,z,\tau)] \\ &= h_1^{-r}\int\{\int\alpha(y;x,z,\tau)K_{h_1}(x-x')K_{h_1}(y-y')K_{h_1}(z-z')dy \\ &\quad -\int\beta(y;x,\tau)K_{h_2}(x-x')K_{h_2}(y-y')dy\}f(x',y',z')d(x',y',z') \\ &= \{\int\alpha(y;x,z,\tau)\left[C_0{}^{d_1}\frac{\partial^{|r|}f(x,y,z)}{\partial x^{|r|}}+C_0{}^{d_3}\frac{\partial^{|r|}f(x,y,z)}{\partial z^{|r|}}\right]dy \\ &\quad +C_0{}^{d_2}\int\frac{\partial^{|r|}H(y+\tau)}{\partial y^{|r|}}\frac{f(x,y,z)}{f(x,z)}dy\}+o(1) \\ &\equiv \widetilde{\gamma}(x,z;\tau)+o(1), \end{aligned}$$

where $C_0\equiv\int_R u^rk(u)du$ is defined in Assumption $A.2$.

Notice that $EG_n(W)=0$, and

$$\begin{aligned} &E[G_n(W_i)G_n(W_0)] \\ &= E\{\int\int\widetilde{R}(W_i;x,z,\tau)h_1^{-r}Er_n(x,z;\tau)a(x,z)dF(x,z)dG(\tau) \\ &\qquad \times\int\int\widetilde{R}(W_0;x,z,\tau)h^{-r}Er_n(x,z;\tau)a(x,z)dF(x,z)dG(\tau)\} \\ &= \int\int\int\int\int\int[R(w_i;x,z,\tau)R(w_0;x',z',\tau')\widetilde{\gamma}(x,z;\tau)a(x,z)\widetilde{\gamma}(x',z';\tau')a(x',z')] \\ &\qquad \times dF(x,z)dF(x',z')\ f_i(w_0,w_i)dw_idw_0dG(\tau)dG(\tau') \\ &\quad -\left\{\int\int\int\int R(w;x,z,\tau)\widetilde{\gamma}(x,z;\tau)a(x,z)dF(x,z)f(w)dwdG(\tau)\right\}^2+o(1). \\ &= Cov(\gamma(W_i),\gamma(W_0))+o(1), \end{aligned}$$

where for $u=(x,y,z)$,

$$\gamma(u)\equiv\int\left[\alpha(y;x,z,\tau)-\beta(y;x,\tau)\frac{f(x,y)}{f(x,y,z)}\right]\widetilde{\gamma}(x,z;\tau)a(x,z)f(x,z)dG(\tau). \tag{B.6}$$

Consequently, Conditions (i) and (ii) in Theorem A.4 are satisfied and thus $U_{n,1} \xrightarrow{d} N(0, \widetilde{\sigma}^2)$. ■

Lemma B.5 *Under Assumptions A.1 – A.3 and H_0, $2U_{n,2} \xrightarrow{d} N(0, \sigma_1^2)$, where σ_1^2 is defined in the text.*

Proof. Note that $2U_{n,2} = 2n^{-1} \sum_{1 \le i < j \le n} [H_n(W_i, W_j) - EH_n(W_i, W_j)]$. By construction, $H_n(w, v) = H_n(v, w)$, and $EH_n(W_0, v) = 0$. We are going to verify conditions $(iii) - (vii)$ in Theorem $A.4$.

First, $H_n(W_i, W_0) = h_1^{(d_1+d_3)/2} \int\int \widetilde{R}(W_i; x, z, \tau)\widetilde{R}(W_0; x, z, \tau) a(x, z) dF(x, z) dG(\tau)$. Notice that for $w_0 = (x_0, y_0, z_0)$ and $w_i = (x_i, y_i, z_i)$,

$$\begin{aligned} H_n(u, v) &= h_1^{(d_1+d_3)/2} \int\int_A \{\alpha(y'; x, z, \tau) K_{h_1}(x - x',\ z - z') - \beta(y'; x, \tau) K_{h_2}(x - x')\} \\ &\quad \times \{\alpha(\widetilde{y}; x, z, \tau) K_{h_1}(x - \widetilde{x},\ z - \widetilde{z}) - \beta(\widetilde{y}; x, \tau) K_{h_2}(x - \widetilde{x})\}\, a(x, z) dF(x, z) dG(\tau)\, (1 + O(h_1^r)), \end{aligned}$$

and

$$\begin{aligned} & E|\int\int \{\alpha(Y_i; x, z, \tau) K_{h_1}(x - X_i) K_h(z - Z_i) - \beta(Y_i; x, \tau) K_{h_2}(x - X_i)\} \\ & \times \{\alpha(Y_0; x, z, \tau) K_{h_1}(x - X_0) K_h(z - Z_0) - \beta(Y_0; x, \tau) K_{h_2}(x - X_0)\} a(x, z) dF(x, z) dG(\tau)|^p \\ \le\ & C \int\int |\int\int \alpha(y_i; x, z, \tau)\alpha(y_0; x, z, \tau) K_{h_1}(x - x_i) K_{h_1}(z - z_i) K_{h_1}(x - x_0) K_{h_1}(z - z_0) \\ & \times a(x, z) dF(x, z) dG(\tau)|^p\ f_i(w_0, w_i) dw_0 dw_i \\ \le\ & C h_1^{-(d_1+d_3)(p-1)} \int_{R^{d_1+d_3}} \int_{R^{d_1+d_3}} |K(u_1) K(u_1 + u_2)|^p du_1 du_2, \end{aligned}$$

so we have

$||H_n(W_i, W_0)||_p \le C h_1^{(d_1+d_3)/2} h^{-(d_1+d_3)(p-1)/p} = C(h_1^{(d_1+d_3)})^{(1/p-1/2)}$.

Let $\overline{W}_0$ be an independent copy of W_0; one can show by similar argument that

$||H_n(W_0, \overline{W}_0)||_p \le C(h_1^{d_1+d_3})^{(1/p-1/2)}$.

Consequently, one obtains $u_n(p) \le C(h^{d_1+d_3})^{(1/p-1/2)}$ for some $C > 0$.

Now we show $v_n(p) \le C(h_1^{d_1+d_3})^{1/p}$. Note that $G_{n0}(w, v) \equiv E[H_n(W_0, w) H_n(W_0, v)]$

$= h^{d_1+d_3} E\{\int\int \widetilde{R}(W_0; x, z, \tau)\widetilde{R}(w; x, z, \tau)\widetilde{R}(W_0; x', z', \tau')\widetilde{R}(v; x', z', \tau') a(x, z) a(x', z')$

$\times dF(x, z) dF(x', z') dG(\tau) dG(\tau')\}$

$\le C \int_{R^{d_1+d_3}} \int_{R^{d_1+d_3}} \int_{R^{d_1+d_3}} K(w) K(w + w') K(\widetilde{w}) K(\widetilde{w} + w' + (u - v)/h_1) dw dw' d\widetilde{w} + O(h_1^{d_1+d_3})$,

so $||G_{n0}(W_i, W_0)||_p \le C((h_1^{d_1+d_3})^{1/p} + h_1^{d_1+d_3})$ and $||G_{n0}(W_i, W_0)||_p \le C(h_1^{d_1+d_3})^{1/p}$.

Similarly, one can show $||G_{n0}(W_0, \overline{W}_0)||_p \le C(h_1^{d_1+d_3})^{1/p}$.and thus $v_n(p) \le C(h_1^{d_1+d_3})^{1/p}$. By the same argument, we have, $w_n(p) \equiv ||G_{n0}(W_0, W_0)||_p \le C$ and $z_n(p) \le C h_1^{d_1+d_3}$.

For some fixed $\delta_0 > 0$, Conditions (iv) and (v) in Theorem $A.4$ are satisfied. Take $\gamma_0 = (2 + \delta_0)/(8 + 2\delta_0) \in (0, 1/2)$ and $\gamma_1 \in (0, \gamma_0]$ by Assumption $A.2(ii)$. Conditions (iii) and (vi) in Theorem $A.4$ are also satisfied.

Finally, $E[H_n(W_0, \overline{W}_0)^2]$

$= h_1^{d_1+d_3} E\{\int\int\int\int \widetilde{R}(W_0; x, z, \tau)\widetilde{R}(\overline{W}_0; x.z, \tau)\widetilde{R}(W_0; x', z', \tau')\widetilde{R}(\overline{W}_0; x', z', \tau') a(x, z) a(x', z')$

$\times dF(x, z) dF(x', z') dG(\tau) dG(\tau')\}$

$= C_3{}^{(d_1+d_3)} \int\int\int \left[\sigma^2(x, z; \tau, \tau')\right]^2 a(x, z)^2 d(x, z) dG(\tau) dG(\tau') + o(1)$,

where C_3 and $\sigma^2(x, z; \tau, \tau')$ are defined in the main text (Section 3.1).

It follows that $U_{n,2} \xrightarrow{d} N(0, \sigma_1^2)$. ■

Lemma B.6 *Under Assumptions A.1 − A.3 and* H_0,
$nh_1^{(d_1+d_3)/2}EI_n = h_1^{-(d_1+d_3)/2}B_1 + h_1^{(d_3-d_1)/2}B_2 + h_1^{(d_1+d_3)/2}h_2^{-d_1}B_3 + o(1)$.

Proof. Write

$$\begin{aligned} EI_n &= \int\int [Er_n(x,z;\tau)]^2 a(x,z)dF(x,z)dG(\tau) \\ &\quad + E\left\{\int\int [r_n(x,z;\tau) - Er_n(x,z;\tau)]^2 a(x,z)dF(x,z)dG(\tau)\right\} \\ &\equiv A_{n,1} + A_{n,2}. \end{aligned}$$

From the proof of Lemma B.3, we obtain

$$\begin{aligned} A_{n,1} &= h_1^{2r}\int\int \widetilde{\gamma}(x,z;\tau)^2 a(x,z)dF(x,z)dG(\tau) + o(h_1^{2r}) \\ &= n^{-1}h_1^{-(d_1+d_3)/2}\{nh_1^{(d_1+d_3)/2+2r}\int\int \widetilde{\gamma}(x,z;\tau)^2 a(x,z)dF(x,z)dG(\tau) + o(h_1^{2r}) \qquad \text{(B.7)} \\ &= o(n^{-1}h_1^{-(d_1+d_3)/2}). \end{aligned}$$

Now write
$A_{n,2} = n^{-2}E\left\{\int_A \left[\sum_{t=1}^n \widetilde{R}(W_t;x,z,\tau)\right]^2 a(x,z)dF(x,z)dG(\tau)\right\}$
$= n^{-2}\sum_{i=1}^n E\left\{\int_A \widetilde{R}(W_i;x,z,\tau)]^2 a(x,z)dF(x,z)dG(\tau)\right\}$
$+2n^{-2}\sum_{1\le i<j\le n} E\{\int_A \widetilde{R}(W_i;x,z,\tau)\widetilde{R}(W_j;x,z,\tau)a(x,z)dF(x,z)dG(\tau)$
$= n^{-1}h_1^{-(d_1+d_3)/2}\left\{EH_n(W_0,W_0) + 2n^{-1}\sum_{1\le i<j\le n} EH_n(W_i,W_j)\right\}.$
We want to show

$$EH_n(W_0,W_0) = \left\{h_1^{-(d_1+d_3)/2}B_1 + h_1^{(d_3-d_1)/2}B_2 + h_1^{(d_1+d_3)/2}h_2^{-d_1}B_3\right\}\{1+O(h_1^{2r})\} \qquad \text{(B.8)}$$

and

$$D_n \equiv 2n^{-1}\sum_{1\le i<j\le n} EH_n(W_i,W_j) = o_p(1). \qquad \text{(B.9)}$$

Now
$EH_n(W_0,W_0) = E\left[h_1^{(d_1+d_3)/2}\int\int \widetilde{R}(W_0;x,z,\tau)^2 a(x,z)dF(x,z)dG(\tau)\right]$
$= h_1^{(d_1+d_3)/2}E\left\{\int\int R(W_0;x,z,\tau)^2 a(x,z)dF(x,z)dG(\tau)\right\}\{1+O(h_1^{2r})\}$
$= h_1^{(d_1+d_3)/2}\int\int\int[\alpha(y_0;x,z,\tau)K_{h_1}(x-x_0,z-z_0) - \beta(y_0;x,\tau)K_{h_2}(x-x_0)]^2$
$\times a(x,z)dF(x,z)dG(\tau)f(x_0,y_0,z_0)d(x_0,y_0,z_0)\{1+O(h_1^{2r})\}$
$= \{h_1^{(d_1+d_3)/2}\int\int\int \alpha^2(y_0;x,z,\tau)K_{h_1}^2(x-x_0,z-z_0)a(x,z)dF(x,z)dG(\tau)dF(x_0,y_0,z_0)$
$-2h_1^{(d_1+d_3)/2}\int\int\int \alpha(y_0;x,z,\tau)\beta(y_0;x,\tau)K_{h_1}(x-x_0,z-z_0)K_{h_2}(x-x_0)a(x,z)dF(x,z)dG(\tau)dF(x_0,y_0,z_0)$
$+h_1^{(d_1+d_3)/2}\int\int\int \beta^2(y_0;x,\tau)K_{h_2}^2(x-x_0)a(x,z)dF(x,z)dG(\tau)dF(x_0,y_0,z_0)\}\{1+O(h_1^{2r})\}$
$\equiv \{B_{n,1}+B_{n,2}+B_{n,3}\}\{1+O(h_1^{2r})\}.$

Simple algebra shows that

$$B_{n,1} = h_1^{-(d_1+d_3)/2}C_1^{(d_1+d_3)}\int\int_A \sigma^2(x,z;\tau)a(x,z)d(x,z)dG(\tau)\{1+O(h_1^{2r})\},$$

$$B_{n,2} = -2h_1^{(d_3-d_1)/2}C_2^{d_1}\int\int_A \sigma^2(x,z;\tau)a(x,z)\frac{f(x,z)}{f(x)}d(x,z)dG(\tau)\{1+O(h_2^{2r})\}, \text{ and}$$

$$B_{n,3} = h_1^{(d_1+d_3)/2}h_2^{-d_1}\int\int \sigma^2(x;\tau)a(x)dxdG(\tau)\{1+O(h_2^{2r})\}.$$

Combining the last three terms, we have (B.8).

We next show $D_n = o_p(1)$. Let $m = [B\log n]$ (the integer part of $B\log n$), where B is a large positive constant so that $n^4\beta_m^{\delta/(1+\delta)} = o(1)$ for some $\delta > 0$ by Assumption $A.1(i)$[23]. We consider two different cases for D_n : (a) $j-i > m$ and (b) $0 < j-i \le m$. We use $D_{n,a}$ and $D_{n,b}$ to denote these two cases. For case (a), we use Lemma $A.2$ and the bound $u_n(p) \le C(h_1^{(d_1+d_3)})^{1/p-1/2}$ with $p = 1+\delta$ (see the proof of Lemma $B.5$) to obtain

$D_{n,a} = n^{-1}\sum_{j-i>m} EH_n(W_i,W_j) \le Cn^{-1}n^2(h_1^{(d_1+d_3)})^{1/(1+\delta)-1/2}\beta_m^{\delta/(1+\delta)} = o(nh_1^{-(d_1+d_3)/2}\beta_m^{\delta/(1+\delta)})$
$= o(1)$.

For case (b), using the bound $u_n(1) \le Ch_1^{(d_1+d_3)/2}$, we have

$D_{n,b} = n^{-1}\sum_{j-i\le m} EH_n(W_i,W_j) \le Cn^{-1}nmh_1^{(d_1+d_3)/2} = O(mh_1^{(d_1+d_3)/2}) = o(1)$.

Consequently, we have (B.9). Combining (B.7), (B.8) and (B.9), the conclusion thus follows. ■

Lemma B.7 *Let $\widetilde{\Delta}_n = \Gamma(\widehat{f},\widehat{F}) - \Gamma(\widehat{f},F)$, then under Assumptions $A.1 - A.3$ and H_0, $nh_1^{(d_1+d_3)/2}$ $\widetilde{\Delta}_n = o_p(1)$.*

Proof. By the same argument as used to obtain the expansion of $\Gamma(\widehat{f},F)$, one can obtain under H_0,

$$\Gamma(\widehat{f},\widehat{F}) = \int\int\left\{\int \alpha(x,y,z;\tau)\widehat{f}_{h_1}(x,y,z)dy - \int \beta(x,y;\tau)\widehat{f}_{h_2}(x,y)dy\right\}^2 a(x,z)d\widehat{F}(x,z)dG(\tau)$$
$$+O_p(\left\|\widehat{f}_{h_1}-f\right\|_s^3).$$

So it suffices to show that $\Delta_n = o_p(n^{-1}h_1^{-(d_1+d_3)/2})$, where

$$\Delta_n \equiv \int\int\left\{\int \alpha(x,y,z;\tau)\widehat{f}_{h_1}(x,y,z)dy - \int \beta(x,y;\tau)\widehat{f}_{h_2}(x,y)dy\right\}^2 a(x,z)dG(\tau)d[\widehat{F}(x,z)-F(x,z)].$$

$\Delta_n = \int\int_A r_n(x,z;\tau)^2 a(x,z)dG(\tau)d[\widehat{F}(x,z)-F(x,z)]$
$= n^{-2}\sum_{j,k=1}^n \int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dG(\tau)d[\widehat{F}(x,z)-F(x,z)]$
$= n^{-3}\sum_{j,k,l=1}^n \{\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)$
$- \int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dF(x,z)dG(\tau)\}$
$= \sum_{i=1}^4 \Delta_{n,i}$,

where

$\Delta_{n,1} \equiv n^{-3}\sum_{l\ne j,k}^n \{\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)$
$- \int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dF(x,z)dG(\tau)\}$

is the summation of the centered terms with $l \ne j, l \ne k$;

$\Delta_{n,2} \equiv 2n^{-3}\sum_{j\ne k}^n \int R(W_j;X_j,Z_j,\tau)R(W_k;X_j,Z_j)a(X_j,Z_j)dG(\tau)$

is the summation of the terms with $l = j \ne k$;

$\Delta_{n,3} \equiv n^{-3}\sum_{j=1}^n \int R(W_j;X_j,Z_j,\tau)^2 a(X_j,Z_j)dG(\tau)$

[23]For example, for fixed $\delta > 0$, if $\rho < 1/2.71828$ in Assumption $A.1(i)$, $B = 4(1+\delta)/\delta$ would suffice.

is the summation of the terms with $l = j = k$; and

$\Delta_{n,4} \equiv -n^{-3}\sum_{j,k=1}^{n}\int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dF(x,z)dG(\tau)$

is summation of the centering terms for $\Delta_{n,2}$ and $\Delta_{n,3}$.

Dispensing with the simpler terms first, we have

$-\Delta_{n,4} = n^{-1}I_n = n^{-1}(I_n - EI_n) + n^{-1}EI_n$

$= n^{-2}h_1^{-(d_1+d_3)/2}\left[nh_1^{(d_1+d_3)/2}(I_n - EI_n)\right] + n^{-1}O(h_1^{2r} + n^{-1}h_1^{-(d_1+d_3)})$

$= O_p(n^{-2}h_1^{-(d_1+d_3)/2}) + O(n^{-1}h_1^{-(d_1+d_3)/2}h_1^{(d_1+d_3)/2+2r}) + O(n^{-2}h_1^{-(d_1+d_3)}) = o_p(n^{-1}h_1^{-(d_1+d_3)/2})$.

$E|\Delta_{n,3}| = n^{-3}\sum_{j=1}^{n}\int R(W_j;X_j,Z_j,\tau)^2 a(X_j,Z_j)dG(\tau) = n^{-2}E\left[\int R(W_0;X_0,Z_0,\tau)^2 a(X_0,Z_0)dG(\tau)\right]$

$= O(n^{-2}h_1^{-2(d_1+d_3)}) = o(n^{-1}h_1^{-(d_1+d_3)/2})$.

Consequently, by the Markov inequality, $\Delta_{n,3} = o_p(n^{-1}h_1^{-(d_1+d_3)/2})$.

It is difficult to show that the other two terms are small. Our strategy is to use Lemmas $A.2 - A.3$ repeatedly and show asymptotically negligibility in that $\Delta_{n,i} = o_p(n^{-1}h_1^{-(d_1+d_3)/2})$, for $i = 1$ and 2.

For $j \neq k$, we can show that (recall that the bar notation denotes an $i.i.d.$ sequence)

$$E\left[\int R(W_j;X_j,Z_j,\tau)R(W_k;X_j,Z_j)a(X_j,Z_j)dG(\tau)\right] = O(h_1^{-(d_1+d_3)}).$$

and

$$E\left[\int R(\overline{W}_j;\overline{X}_j,\overline{Z}_j,\tau)R(\overline{W}_k;\overline{X}_j,\overline{Z}_j)a(\overline{X}_j,\overline{Z}_j)dG(\tau)\right] = O(h_1^{r-(d_1+d_3)}).$$

To bound $D_n \equiv E(\Delta_{n,2}) = 2n^{-3}\sum_{j\neq k}^{n}E\left[\int R(W_j;X_j,Z_j,\tau)R(W_k;X_j,Z_j)a(X_j,Z_j)dG(\tau)\right]$, we consider two different cases for D_n : (a) $|j-k| > m$ and (b) $|j-k| \leq m$. We use $D_{n,a}$ and $D_{n,b}$ to denote these two cases. For case (a), we use Lemma $A.2$ to obtain

$D_{n,a} = 2n^{-3}\sum_{|j-k|>m}E\left[\int R(W_j;X_j,Z_j,\tau)R(W_k;X_j,Z_j)a(X_j,Z_j)dG(\tau)\right]$

$\leq C\{n^{-1}h_1^{r-(d_1+d_3)} + n^{-3}n^2(h_1^{-(d_1+d_3)})^{(1+2\delta)/(1+\delta)}\beta_m^{\delta/(1+\delta)}\}$

$= O(n^{-1}h_1^{-(d_1+d_3)/2}h_1^{r-(d_1+d_3)/2}) + o(n^{-1}h_1^{-(d_1+d_3)}\beta_m^{\delta/(1+\delta)}) = o(n^{-1}h_1^{-(d_1+d_3)/2})$.

For case (b),

$D_{n,b} = 2n^{-3}\sum_{|j-k|\leq m}E\left[\int R(W_j;X_j,Z_j,\tau)R(W_k;X_j,Z_j)a(X_j,Z_j)dG(\tau)\right]$

$\leq Cn^{-3}nmh_1^{-(d_1+d_3)} = O(n^{-2}mh_1^{-(d_1+d_3)}) = o(n^{-1}h_1^{-(d_1+d_3)/2})$.

In consequence,

$$E(\Delta_{n,2}) = o(n^{-1}h_1^{-(d_1+d_3)/2}).$$

Now, we want to show

$$E_n \equiv E(\Delta_{n,2})^2 = o(n^{-2}h_1^{-(d_1+d_3)}),$$

where

$E_n = 4n^{-6}\sum_{t_1\neq t_2}\sum_{t_3\neq t_4}E\{\int\int R(W_{t_1};X_{t_1},Z_{t_1},\tau)R(W_{t_2};X_{t_1},Z_{t_1},\tau)a(X_{t_1},Z_{t_1})R(W_{t_3};X_{t_3},Z_{t_3},\tau')$

$\times R(W_{t_4};X_{t_3},Z_{t_3},\tau')a(X_{t_3},Z_{t_3})dG(\tau)dG(\tau')\}$.

We consider two cases: (a) for all $i \in \{1,2,3,4\}, |t_i - t_j| > m$ for all $j \neq i$; and (b) all the other remaining cases. We will use $E_{n,s}$ to denote these cases $(s = a, b)$.

Using Lemma $A.2$ three times, we have:

$$\begin{aligned} E_{n,a} &\leq 4n^{-6}\sum_{t_1\neq t_2}\sum_{t_3\neq t_4} E\left\{\int R(\overline{W}_{t_1};\overline{X}_{t_1},\overline{Z}_{t_1},\tau)R(\overline{W}_{t_2};\overline{X}_{t_1},\overline{Z}_{t_1},\tau)a(\overline{X}_{t_1},\overline{Z}_{t_1})dG(\tau)\right\} \\ &\quad \times E\left\{\int R(\overline{W}_{t_3};\overline{X}_{t_3},\overline{Z}_{t_3},\tau')R(\overline{W}_{t_3};\overline{X}_{t_3},\overline{Z}_{t_3},\tau')a(\overline{X}_{t_3},\overline{Z}_{t_3})dG(\tau')\right\} \\ &\quad +C(n^{-2}(h_1^{-(d_1+d_3)})^{4\delta/(1+\delta)}\beta_m^{\delta/(1+\delta)}) \\ &= O(n^{-2}h_1^{2(r-d_1-d_3)})+o(\beta_m^{\delta/(1+\delta)})=o(n^{-2}h_1^{-(d_1+d_3)}). \end{aligned}$$

For all the other remaining cases, there exists at least one $i\in\{1,2,3,4\}$, such that $|t_i-t_j|\leq m$ for some $j\neq i$. The number of such terms is of the order $O(n^3m)$. For $t_1\neq t_2$ and $t_3\neq t_4$, one can bound $E|\int\int R(W_{t_1};X_{t_1},Z_{t_1},\tau)R(W_{t_2};X_{t_1},Z_{t_1},\tau)a(X_{t_1},Z_{t_1})R(W_{t_3};X_{t_3},Z_{t_3},\tau')\ R(W_{t_4};X_{t_3},Z_{t_3},\tau')a(X_{t_3},Z_{t_3})$

$\times dG(\tau)dG(\tau')|$ by $Ch_1^{-2(d_1+d_3)}$ if $\{t_1,t_2\}\cap\{t_3,t_4\}\neq\{t_1,t_2\}$ and by $Ch_1^{-3(d_1+d_3)}$ otherwise. Consequently,

$$E_{n,b}\leq C(n^{-6}n^3mh_1^{-2(d_1+d_3)}+n^{-6}n^2h_1^{-3(d_1+d_3)})=o(n^{-2}h_1^{-(d_1+d_3)}).$$

In sum, $E(\Delta_{n,2})^2=o(n^{-2}h_1^{-(d_1+d_3)})$, and by Chebyshev's inequality, we have $\Delta_{n,2}=o_p(n^{-1}h_1^{-(d_1+d_3)/2})$.

Lastly, we want to show

$$\Delta_{n,1}=o_p(n^{-1}h_1^{-(d_1+d_3)/2}),$$

where

$$\begin{aligned} \Delta_{n,1} &= n^{-3}\textstyle\sum_{l\neq j,k}^n\{\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau) \\ &\qquad -\int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dF(x,z)dG(\tau)\} \\ &= n^{-3}\textstyle\sum_{l\neq j,k}^n\{\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau) \\ &\qquad -E\left[\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)|W_j,W_k)\right]\} \\ &\quad +n^{-3}\textstyle\sum_{l\neq j,k}^n\{E\left[\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)|W_j,W_k)\right] \\ &\qquad -\int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)a(x,z)dF(x,z)dG(\tau)\} \\ &\equiv \Delta_{n,11}+\Delta_{n,12}. \end{aligned}$$

Using Lemma $A.3$, we can show

$$\begin{aligned} E|\Delta_{n,12}| &\leq n^{-3}\sum_{l\neq j,k}^n E|E\left[\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)|W_j,W_k)\right] \\ &\quad -\int\int R(W_j;x,z,\tau)R(W_k;x,z,\tau)\ dG(\tau)a(x,z)dF(x,z)| \\ &= o(n^{-1}h_1^{-(d_1+d_3)/2}), \end{aligned}$$

implying $\Delta_{n,12}=o_p(n^{-1}h_1^{-(d_1+d_3)/2})$ by the Markov inequality.

Now let

$$\begin{aligned} S_{j,k,l} &\equiv \int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau) \\ &\quad -E\left[\int R(W_j;X_l,Z_l,\tau)R(W_k;X_l,Z_l,\tau)a(X_l,Z_l)dG(\tau)|W_j,W_k)\right]; \end{aligned}$$

then

$\Delta_{n,11} = n^{-3}\sum_{l\neq j,k}^{n} S_{j,k,l}$ with $E(\Delta_{n,11}) = 0$ because $E(S_{j,k,l}) = 0$ for all $l \neq j$ and $l \neq k$. We shall show

$$F_n \equiv E(\Delta_{n,11})^2 = n^{-6} \sum_{t_1\neq t_3, t_2 \neq t_3, t_3} \sum_{t_4 \neq t_6, t_5 \neq t_6, t_6} E\{S_{t_1,t_2,t_3} S_{t_4,t_5,t_6}\} = o(n^{-2}h_1^{-(d_1+d_3)}).$$

We consider four different cases:(a) for all $i's$, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly four different $i's$, $|t_i - t_j| > m$ for all $j \neq i$; (c) for exactly three different $i's$, $|t_i - t_j| > m$ for all $j \neq i$; (d) all the other remaining cases. We will use $F_{n,s}$ to denote these cases $(s = a, b, c, d)$. For each case, one can use Lemma $A.2$ and Lemma $A.3$ to show $F_{n,s} = o(n^{-2}h_1^{-(d_1+d_3)})$. For example, for case (a), noting that $E(S_{t_1,t_2,t_3}) = E(S_{t_4,t_5,t_6}) = 0$, we have by Lemma $A.2$

$$|F_{n,a}| \leq C(h_1^{-(d_1+d_3)})^{4\delta/(1+\delta)}\beta_m^{\delta/(1+\delta)} = o(n^2\beta_m^{\delta/(1+\delta)}) = o(n^{-2}h_1^{-(d_1+d_3)}).$$

For case (d), the number of terms in the summation is of order $O(n^3m^3)$ and each term can be bounded by $Ch_1^{-(d_1+d_3)}$ for some finite positive constant C if there are at least three distinct elements in $\{t_1, t_2, t_3, t_4, t_5, t_6\}$ and $Ch_1^{-2(d_1+d_3)}$ otherwise. So

$$F_{n,d} = O(n^{-6}n^3m^3h_1^{-(d_1+d_3)} + n^{-6}n^2h_1^{-2(d_1+d_3)}) = o(n^{-2}h_1^{-(d_1+d_3)}).$$

In sum, $F_n = o(n^{-2}h_1^{-(d_1+d_3)})$ and thus $\Delta_{n,11} = o_p(n^{-1}h_1^{-(d_1+d_3)/2})$ by the Chebyshev inequality. The conclusion thus follows. ■

Lemma B.8 *Under Assumptions $A.1 - A.3$* , $nh_1^{(d_1+d_3)/2}\left\|\widehat{f}_{h_1} - f\right\|_s^3 = o_p(1)$.

Proof. This follows from the proof of Lemma B.3 and the fact that
$nh_1^{(d_1+d_3)/2}O_p(n^{-3/2}h_1^{-3(d_1+d_3)/2}(\ln n)^{3/2} + h_1^{3r})$
$= O_p(n^{-1/2}h_1^{-(d_1+d_3)}(\ln n)^{3/2} + nh_1^{d/2+3r}) = o_p(1)$. ■
Putting Lemmas B.3-B.8 together, we have proved Theorem 3.1 in the main text.

C Proof of Lemmas, Propositions and Corollaries

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 A for every $\tau \in R^{d_2}$; we must show that $\psi(u;x,z) = 0$ a.e.$-G_0 \times F$ on $R^{d_2} \times A$.

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 R^d satisfying $\mathrm{Re}(\psi) = \psi_1 - \psi_2$ and $\mathrm{Im}(\psi) = \psi_3 - \psi_4$.

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

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

Then[24]

$\int\ \psi(u;x,z)e^{i\tau' u}dG_0(u)$

$= \left[\int \psi_1(u;x,z)e^{i\tau' u}dG_0(u) - \int \psi_2(u;x,z)e^{i\tau' u}dG_0(u)\right] + i\ [\int \psi_3(u;x,z)e^{i\tau' u}dG_0(u) - \int \psi_4(u;x,z)$
$\times e^{i\tau' u}dG_0(u)]$

$= \left[c_1 \int e^{i\tau' u}dv_1(u;x,z) - c_2 \int e^{i\tau' u}dv_2(u;x,z)\right] + i\ \left[c_3 \int e^{i\tau' u}dv_3(u;x,z) - c_4 \int e^{i\tau' u}dv_4(u;x,z)\right]$

$= [c_1\eta_1(\tau;x,z) - c_2\eta_2(\tau;x,z)] + i\ [c_3\eta_3(\tau;x,z) - c_4\eta_4(\tau;x,z)],$

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 A for every $\tau \in R^{d_2}$, $c_1\eta_1(\tau;x,z) = c_2\eta_2(\tau;x,z)$ and $c_3\eta_3(\tau;x,z) = c_4\eta_4(\tau;x,z)$ a.e.$-(x,z)$ for every $\tau \in R^{d_2}$. Note that $\eta_1(0;x,z) = \eta_2(0;x,z) = \eta_3(0;x,z) = \eta_4(0;x,z) = 1$, so

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

and

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

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

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

From (C.1), (C.2) and (C.3), we obtain that for every Borel set B on 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 $B_1 \equiv \{u \in R^{d_2} : \text{Re}(\ \psi(u;x,z)) > 0\}$ is a Borel set, $\int_{B_1} \text{Re}(\ \psi(u;x,z))dG_0(u) = 0$, which is only possible if B_1 is a null set with respect to $dG_0(u)$ a.e.$-F$ on A. Similarly, one concludes that the Borel sets $B_2 \equiv \{u \in R^{d_2} : \text{Re}(\psi(u;x,z)) < 0\}$, $B_3 \equiv \{u \in R^{d_2} : \text{Im}(\psi(u;x,z)) > 0\}$ and $B_4 \equiv \{u \in R^{d_2} : \text{Im}(\psi(u;x,z)) < 0\}$ are all null sets with respect to $dG_0(u)$ a.e.$-F$ on A. Hence, $\cup_{i=1}^4 B_i = \{u \in R^{d_2} : \psi(u;x,z) \neq 0\}$ is a null set with respect to $dG_0(u)$ a.e.$-F$ on A. This means $\psi(u;x,z) = 0$ a.e.$-G_0 \times F$ on $R^{d_2} \times A$. 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 Proposition 3.2. The analysis is similar to that of Lemma B.1 and Lemmas B.2-B.7, keeping now the additional terms that were not present under the null. That is, as before,

$\Psi(\lambda) = \Psi(0) + \tau\Psi'(0) + \lambda^2\Psi''(0)/2 + \lambda^3\Psi'''(\lambda^*)/6$

for some $\lambda^* \in [0,\lambda]$, where $\Psi(0) = \Gamma(f,F)$,

[24] Here, $dv_j(u;x,z) \equiv v_j(du;x,z)$ for $j = 1,2,3$ and 4.

$$\Psi'(0)=\int\int 2\varphi(0;x,z,\tau)\frac{\partial\varphi(0;x,z,\tau)}{\partial\lambda}a(x,y)dF(x,y)dG(\tau),$$

and

$$\Psi''(0)=\int\int 2\left\{\left[\frac{\partial\varphi(0;x,z,\tau)}{\partial\lambda}\right]^2+\varphi(0;x,z,\tau)\frac{\partial^2\varphi(0;x,z,\tau)}{\partial\lambda^2}\right\}a(x,y)dF(x,y)dG(\tau).$$

The expression for $\Psi'''(\lambda^*)$ is lengthy. However, it can be bounded as before by the same upper bound which is $o_p(n^{-1}h_1^{-(d_1+d_3)/2})$. Consequently,

$\Gamma(\widehat{f},F)=\Gamma(f,F)+\Psi'(0)+\Psi''(0)/2+o_p(n^{-1}h_1^{-(d_1+d_3)/2})$

Furthermore, $\Gamma(\widehat{f},\widehat{F})=\Gamma(\widehat{f},F)+o_p(n^{-1}h^{-d/2})$ continues to hold, and thus $\Gamma(\widehat{f},\widehat{F})=\Gamma(f,F)+\Psi'(0)+\Psi''(0)/2+o_p(n^{-1}h_1^{-(d_1+d_3)/2})$.

Notice that when $\Gamma(f,F)>0$, $n^{1/2}\Psi'(0)=O_p(1)$ and $\Psi'(0)$ dominates $\Psi''(0)$. So

$T_{1,n}=nh_1^{(d_1+d_3)/2}\Gamma(f,F)/\sqrt{\sigma_1^2}+n^{1/2}h_1^{(d_1+d_3)/2}O_p(1)\xrightarrow{p}\infty$ if $\Gamma(f,F)\geq\varepsilon>0$.

Hence, the test is consistent. ■

Proof of Proposition 3.3. First, for the double array stochastic process $\{W_{nt},0\leq t\leq n\}$, the functional expansion of $\Gamma(\widehat{f^{[n]},F^{[n]}})$ and subsequent lemmas in Appendix B continue to hold, when accommodating the additional terms arising under the local alternative. Under $H_1(\alpha_n)$, $\widehat{\sigma}_1^2\xrightarrow{p}\sigma_1^2$ and $T_{1,n}-nh^{(d_1+d_3)/2}\Gamma(f^{[n]},F^{[n]})/\sqrt{\widehat{\sigma}_1^2}\rightarrow N(0,1)$.

Moreover, under $H_1(\alpha_n)$,

$$\begin{aligned}\Gamma(f^{[n]},F^{[n]}) &= \int\int\{m^{[n]}(x,z;\tau)-m^{[n]}(x;\tau)\}^2a(x,z)dF^{[n]}(x,z)dG(\tau)\\ &= \alpha_n^2\int\int\triangle(x,z;\tau)^2a(x,z)dF^{[n]}(x,z)dG(\tau)+o(\alpha_n^2).\end{aligned}$$

For $\alpha_n=n^{-1/2}h_1^{-(d_1+d_3)/4}$, $nh_1^{(d_1+d_3)/2}\Gamma(f^{[n]},F^{[n]})=\int\int\triangle(x,z;\tau)^2a(x,z)dF^{[n]}(x,z)dG(\tau)$

$\rightarrow\int\int\triangle(x,z;\tau)^2a(x,z)dF(x,z)dG(\tau)\equiv\delta$ as $n\rightarrow\infty$.

Consequently, $\Pr(T_{1,n}\geq z_\alpha|H_1(\alpha_n))\rightarrow 1-\Phi(z_\alpha-\delta/\sigma_1))$. ■

Proof of Corollary 4.1. To ease the notational burden, let $U_j(\theta)\equiv(X_j(\theta),Z_j(\theta))$, $u\equiv(x,z)$ and $f(x,y,z;\theta)\equiv f(y,u;\theta)$.

Theorem 3.1 applies to $\widehat{\Gamma}_2(\theta_0)$ where $\widehat{\Gamma}_2(\theta)\equiv\Gamma_2(\widehat{f},\widehat{F};\theta)$ and θ_0 is the true parameter. Consider the statistic $\widehat{\Gamma}_2(\widehat{\theta})$; we apply a second order Taylor expansion to it around θ_0 :

$$\begin{aligned}\widehat{\Gamma}_2(\widehat{\theta}) &= \widehat{\Gamma}_2(\theta_0)+\triangledown'\widehat{\Gamma}_2(\theta_0)(\widehat{\theta}-\theta_0)+(\widehat{\theta}-\theta_0)'\nabla^2\widehat{\Gamma}_2(\theta^*)(\widehat{\theta}-\theta_0)/2\\ &\equiv \widehat{\Gamma}_2(\theta_0)+D_{n,1}+D_{n,2},\end{aligned}$$

where $\triangledown\widehat{\Gamma}_2$ and $\nabla^2\widehat{\Gamma}_2$ are first and second order derivatives of $\widehat{\Gamma}_2$ with respect to θ, θ^* lies on the segment connecting $\widehat{\theta}$ and θ_0, and primes denote transposition. It suffices to show that $D_{n,i}=o_p(n^{-1}h_1^{-(d_1+d_3)/2})$, $i=1$ and 2.

First, we show

$$\triangledown\widehat{\Gamma}_2(\theta_0)=o_p(n^{-1/2}h_1^{-(d_1+d_3)/2}). \tag{C.4}$$

Write

$$\begin{aligned}\bigtriangledown\Gamma_2(f,\widehat{F};\theta) &= 2\int\int A_1(u;\tau,\theta)A_2(u;\tau,\theta)a(u,\theta)d\widehat{F}(u;\theta)dG(\tau) \\ &\quad +\int\int A_1(u;\tau,\theta)^2\sum_{i=1}^{d_1+d_3}\frac{\partial a}{\partial u_i}\frac{\partial u_i}{\partial\theta}d\widehat{F}(u;\theta)dG(\tau),\end{aligned}$$

where

$$\begin{aligned}A_1(u;\tau,\theta) &\equiv m(x,z;\tau,\theta)-m(x;\tau,\theta) \\ &= \frac{\int H(y+\tau)f(y,u;\theta)dy}{f(u;\theta)}-\frac{\int H(y+\tau)f(y,x;\theta)dy}{f(x;\theta)},\end{aligned}$$

and

$$\begin{aligned}A_2(w,\theta) &\equiv \frac{\partial A_1(u;\tau,\theta)}{\partial\theta} \\ &= \frac{\int H(y+\tau)[f_u(y,u;\theta)\frac{\partial u}{\partial\theta}+f_\theta(y,u;\theta)]dy}{f(u;\theta)}-\frac{\int H(y+\tau)f(y,u;\theta)dy[f_u(u;\theta)\frac{\partial u}{\partial\theta}+f_\theta(u;\theta)]}{f^2(u;\theta)} \\ &\quad -\frac{\int H(y+\tau)[f_x(y,x;\theta)\frac{\partial x}{\partial\theta}+f_\theta(y,x;\theta)]dy}{f(x;\theta)}+\frac{\int H(y+\tau)f(y,x;\theta)dy[f_x(x;\theta)\frac{\partial x}{\partial\theta}+f_\theta(x;\theta)]}{f^2(x;\theta)}.\end{aligned}$$

Clearly, $\bigtriangledown\Gamma_2(f,\widehat{F};\theta_0)=0$ because $A_1(u;\tau,\theta_0)=0$ $a.e.-u$ for every $\tau\in R^{d_2}$. Similarly,

$$\begin{aligned}\bigtriangledown\Gamma_2(\widehat{f},\widehat{F};\theta) &= \int\int \widehat{A}_1(u;\tau,\theta)\widehat{A}_2(u;\tau,\theta)a(u,\theta)d\widehat{F}(u;\theta)dG(\tau) \\ &\quad +\int\int \widehat{A}_1(u;\tau,\theta)^2\sum_{i=1}^{d_1+d_3}\frac{\partial a}{\partial u_i}\frac{\partial u_i}{\partial\theta}d\widehat{F}(u;\theta)dG(\tau) \\ &\equiv d_{n,11}(\theta)+d_{n,12}(\theta),\end{aligned}$$

where, for $i=1,2$, $\widehat{A}_i(u;\tau,\theta)$ is obtained from $A_i(u;\tau,\theta)$ by replacing the unknown densities and their derivatives in the expression by their kernel estimators (e.g., $\widehat{f}_u(u;\theta)\equiv(nh_1)^{-1}\sum_{j=1}^n\bigtriangledown K_{h_1}(u-U_j(\theta))$, and $\widehat{f}_\theta(u;\theta)\equiv(nh_1)^{-1}\sum_{j=1}^n\frac{\partial}{\partial\theta}U_j(\theta)\bigtriangledown K_{h_1}(u-U_j(\theta))$.)

By the argument used in the proof of Theorem 3.1, we have

$$d_{n,12}(\theta_0)=O_p(n^{-1}h_1^{-(d_1+d_3)/2}+n^{-1}h_1^{-(d_1+d_3)})=O_p(n^{-1}h_1^{-(d_1+d_3)}), \tag{C.5}$$

where $n^{-1}h_1^{-(d_1+d_3)}$ is the order of the dominant bias term before correction. To bound $d_{n,11}(\theta_0)$, write

$$\begin{aligned}d_{n,11}(\theta_0) &= \int\int \widehat{A}_1(u;\tau,\theta_0)\left[\widehat{A}_2(u;\tau,\theta_0)-A_2(u;\tau,\theta_0)\right]a(u,\theta_0)d\widehat{F}(u)dG(\tau) \\ &\quad +\int \widehat{A}_1(u;\tau,\theta_0)A_2(u;\tau,\theta_0)a(u,\theta_0)d\widehat{F}(u)dG(\tau).\end{aligned}$$

Notice that $\sup_{u\in A,\tau\in S}|\widehat{A}_1(u;\tau,\theta_0)|=O_p(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}+h_1^r)$ under the null by Lemma $B.2$ and the assumptions on the bandwidth sequences. By similar argument as used in proving Lemma B.2, we can show that $\underset{u\in A,\tau\in S}{sup}\left|\widehat{A}_2(u;\tau,\theta_0)-A_2(u;\tau,\theta_0)\right| = O_p(n^{-1/2}h_1^{-(d_1+d_3)/2-1}\sqrt{\ln n}+h_1^r)$. Further notice that $\sqrt{nh_1^{(d_1+d_3)}}\int\int\widehat{A}_1(u;\tau,\theta_0)A_2(u;\tau,\theta_0)a(u,\theta_0)d\widehat{F}(u) = o_p(1)$, so

$$d_{n,11}(\theta_0) = O_p(n^{-1}h_1^{-(d_1+d_3)-1}\ln n + h_1^{2r} + n^{-1/2}h_1^{-(d_1+d_3)/2}). \tag{C.6}$$

(C.5) and (C.6) then imply (C.4).

Next, we show that

$$\sup_{\theta\in\Theta_0}\bigtriangledown^2\widehat{\Gamma}_2(\theta) = o_p(h_1^{-(d_1+d_3)/2}). \tag{C.7}$$

$$\begin{aligned}
\bigtriangledown^2\widehat{\Gamma}_2(\theta) &= \int\int\widehat{A}_2(u;\tau,\theta)\widehat{A}_2(u;\tau,\theta)'a(u,\theta)d\widehat{F}(u,\theta)dG(\tau)\\
&\quad+\int\int\widehat{A}_1(u;\tau,\theta)\frac{\partial}{\partial\theta'}\widehat{A}_2(u;\tau,\theta)a(u;\tau,\theta)d\widehat{F}(u,\theta)dG(\tau)\\
&\quad+2\int\int\widehat{A}_1(u;\tau,\theta)\widehat{A}_2(u;\tau,\theta)\sum_{i=1}^{d_1+d_3}\frac{\partial a}{\partial u_i}\frac{\partial u_i}{\partial\theta'}d\widehat{F}(u,\theta)dG(\tau)\\
&\quad+\int\int\widehat{A}_1(u;\tau,\theta)^2\left[\sum_{i=1}^{d_1+d_3}\sum_{j=1}^{d_1+d_3}\frac{\partial^2 a}{\partial u_i\partial u_j}\frac{\partial u_i}{\partial\theta}\frac{\partial u_i}{\partial\theta'}+\sum_{i=1}^{d_1+d_3}\frac{\partial a}{\partial u_i}\frac{\partial^2 u_i}{\partial\theta\partial\theta'}\right]d\widehat{F}(u,\theta)dG(\tau)\\
&\equiv d_{n,21}(\theta)+d_{n,22}(\theta)+d_{n,23}(\theta)+d_{n,24}(\theta)
\end{aligned}$$

Noting that $\widehat{A}_i(u;\tau,\theta)$, $i=1,2$, depend on θ only through $\{W_t(\theta)\}$, we can use Theorems $2(c)$ and $2(d)$ in Andrews (1995) (with $\eta=\infty$) to obtain

$\sup_{\theta\in\Theta_0}|d_{n,21}(\theta)| = O_p((n^{-1/2}h_1^{-(d_1+d_3)-1}+h_1^{r-1})^2) = o_p(h_1^{-(d_1+d_3)/2})$,

$\sup_{\theta\in\Theta_0}|d_{n,22}(\theta)| = O_p((n^{-1}h_1^{-(d_1+d_3)-1}\ln n+h_1^r)(n^{-1/2}h_1^{-(d_1+d_3)-2}+h_1^{r-2})) = o_p(h_1^{-(d_1+d_3)/2})$,

$\sup_{\theta\in\Theta_0}|d_{n,23}(\theta)| = O_p((n^{-1}h_1^{-(d_1+d_3)-1}\ln n+h_1^r)(n^{-1/2}h_1^{-(d_1+d_3)-1}+h_1^{r-1})) = o_p(h_1^{-(d_1+d_3)/2})$, and

$\sup_{\theta\in\Theta_0}|d_{n,24}(\theta)| = O_p((n^{-1/2}h_1^{-(d_1+d_3)-1}+h_1^{r-1})^2) = o_p(h_1^{-(d_1+d_3)/2})$.

Consequently, (C.7) is satisfied. (C.4) and (C.7) together conclude the corollary. ■

Proof of Corollary 4.2. For simplicity, we only prove the case for which the kernel function K is compactly supported.[25] Define $sm_{h_1}(x,z;\tau)\equiv\widehat{f}_{h_1}^{-1}(x,z)n^{-1}\sum_{i=1}^n K_{h_1}(x-X_i,z-Z_i)m(X_i;\tau)$, and

[25] We use the compactness of the support of K to simplify the proof of (C.9) and (C.10).

write

$$\begin{aligned}
\widetilde{\Gamma} &\equiv n^{-1}\sum_{j=1}^{n}\int[\widehat{m}_{h_1}(X_j,Z_j;\tau)-\widehat{sm}_{h_1}(X_j,Z_j;\tau)]^2a(X_j,Z_j)dG(\tau)\\
&= n^{-1}\sum_{j=1}^{n}\int[\widehat{m}_{h_1}(X_j,Z_j;\tau)-sm_{h_1}(X_j,Z_j;\tau)]^2a(X_j,Z_j)dG(\tau)\\
&\quad +n^{-1}\sum_{j=1}^{n}\int[sm_{h_1}(X_j,Z_j;\tau)-\widehat{sm}_{h_1}(X_j,Z_j;\tau)]^2a(X_j,Z_j)dG(\tau)\\
&\quad +2n^{-1}\sum_{j=1}^{n}\int[\widehat{m}_{h_1}(X_j,Z_j;\tau)-sm_{h_1}(X_j,Z_j;\tau)][sm_{h_1}(X_j,Z_j;\tau)-\widehat{sm}_{h_1}(X_j,Z_j;\tau)]a(X_j,Z_j)dG(\tau)\\
&\equiv G_{n,1}+G_{n,2}+G_{n,3}.
\end{aligned}$$

We want to show that

$$nh_1^{(d_1+d_3)/2}\{G_{n,1}-n^{-1}h_1^{-(d_1+d_3)}B_1\}\xrightarrow{d}N(0,\sigma_1^2), \tag{C.8}$$

$$nh_1^{(d_1+d_3)/2}G_{n,2}=o_p(1), \text{ and} \tag{C.9}$$

$$nh_1^{(d_1+d_3)/2}G_{n,3}=o_p(1). \tag{C.10}$$

Let $K_{1ji}\equiv K_{h_1}(X_j-X_i,Z_j-Z_i)$, $K_{(x,z),j}\equiv K_{h_1}(x-X_j,z-Z_j)$, $K_{2ji}\equiv K_{h_2}(X_j-X_i)$, $K_{x,i}\equiv K_{h_2}(x-X_i)$, $f_{1j}=f(X_j,Z_j)$, $\widehat{f}_{1j}=\widehat{f}_{h_1}(X_j,Z_j)$, $f_{2j}=f(X_j)$, $\widehat{f}_{2j}=\widehat{f}_{h_2}(X_j)$, $a_j=a(X_j,Z_j)$ and $\varepsilon_j(\tau)\equiv H(Y_j+\tau)-m(X_j;\tau)$. Let $A^\epsilon\equiv\{u\in R^{d_1+d_3}:\|u-v\|\le\epsilon \text{ for some } v\in A\}$ with ϵ being an arbitrarily small positive number. Then the uniform consistency results for the kernel density estimators hold also on A^ϵ by the continuty of the density functions. We can write

$$\begin{aligned}
G_{n,1} &= n^{-1}\sum_{j=1}^{n}\int\{n^{-1}\sum_{i=1}^{n}K_{1ji}\varepsilon_i(\tau)\}^2\widehat{f}_{1j}^{-2}a_jdG(\tau)\\
&= n^{-3}\sum_{j=1}^{n}\int\{\sum_{i=1}^{n}K_{1ji}\varepsilon_i(\tau)\}^2f_{1j}^{-2}a_jdG(\tau)\{1+o_p(1)\}\\
&\equiv T_{n,1}\{1+o_p(1)\}.
\end{aligned}$$

We can decompose $T_{n,1}$ as follows: $T_{n,1}=n^{-2}\sum_{j=1}^{n}\sum_{i=1}^{n}\int\left[K_{1ij}^2\varepsilon_i(\tau)^2\right]a_jf_{1j}^{-2}dG(\tau)+2n^{-3}\sum_{j=1}^{n}\int\{\sum_{i\neq k}^{n}K_{1ij}K_{1kj}\varepsilon_i(\tau)\varepsilon_k(\tau)\}a_jf_{1j}^{-2}dG(\tau)\equiv T_{n,1a}+T_{n,1b}$. As done in Lemma B.7, it can be shown that $T_{n,1a}=\widetilde{T}_{n,1a}+o_p(n^{-1}h_1^{-(d_1+d_3)/2})$, where $\widetilde{T}_{n,1a}=n^{-2}\sum_{i=1}^{n}\int\int K_{(x,z),i}^2\varepsilon_i^2(\tau)f^{-2}(x,z)a(x,z)dF(x,z)dG(\tau)$ $=n^{-1}h_1^{-(d_1+d_3)}B_1\{1+o_p(1)\}$, with B_1 being defined in the text, the only bias term that needs to be corrected in Corollary 4.2. Similarly, it can be shown that $T_{n,1b}=\widetilde{T}_{n,1b}+o_p(n^{-1}h_1^{-(d_1+d_3)/2})$, where $\widetilde{T}_{n,1b}=n^{-2}\sum_{i=1}^{n}\sum_{k\neq i}^{n}\int\int K_{(x,z),i}K_{(x,z),k}\varepsilon_i(\tau)\varepsilon_k(\tau)f^{-2}(x,z)a(x,z)dF(x,z)dG(\tau)$. Again, using the $U-$statistic theory, it is easy to show $nh_1^{(d_1+d_3)/2}\widetilde{T}_{n,1b}\xrightarrow{d}N(0,\sigma_1^2)$. Consequently, $nh_1^{(d_1+d_3)/2}$ $\{G_{n,1}-n^{-1}h_1^{-(d_1+d_3)}B_1\}\xrightarrow{d}N(0,\sigma_1^2)$.

Next, write $\widehat{sm}_{h_1}(X_j,Z_j;\tau)-sm_{h_1}(X_j,Z_j;\tau)=\widehat{f}_{1j}^{-1}n^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n}\widehat{f}_{2i}^{-1}K_{1ji}K_{2ik}\{\varepsilon_k(\tau)+[m(X_k;\tau)-m(X_i;\tau)]\}=f_{1j}^{-1}n^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n}f_{2i}^{-1}K_{1ji}K_{2ik}\{\varepsilon_i(\tau)+[m(X_k;\tau)-m(X_i;\tau)]\}\{1+o_p(1)\}$, where the $o_p(1)$ holds uniformly for $((X_j,Z_j),\tau)\in S\times R$. Define $S_{t_0,t_1,t_2}^{(1)}(\tau)=f_{1t_0}^{-1}\ f_{2t_1}^{-1}K_{1t_0t_1}K_{2t_1t_2}\varepsilon_{t_2}(\tau)a_{t_0}^{1/2}$ and

$S^{(2)}_{t_0,t_1,t_2}(\tau) = f^{-1}_{1t_0}\ f^{-1}_{2t_1} K_{1t_0t_1} K_{2t_1t_2}[m(X_{t_2};\tau) - m(X_{t_1};\tau)]a^{1/2}_{t_0}$. Then by the Cauchy-Schwarz inequality,

$$\begin{aligned} nh_1^{(d_1+d_3)/2} G_{n,2} &= \{n^{-4} h_1^{(d_1+d_3)/2} \sum_{t_0,t_1,t_2,t_3,t_4} \int S^{(1)}_{t_0,t_1,t_2}(\tau) S^{(1)}_{t_0,t_3,t_4}(\tau)) dG(\tau) \\ &\quad + n^{-4} h_1^{(d_1+d_3)/2} \sum_{t_0,t_1,t_2,t_3,t_4} \int S^{(2)}_{t_0,t_1,t_2}(\tau) S^{(2)}_{t_0,t_3,t_4}(\tau) dG(\tau) \\ &\quad + 2n^{-4} h_1^{(d_1+d_3)/2} \sum_{t_0,t_1,t_2,t_3,t_4} \int S^{(1)}_{t_0,t_1,t_2}(\tau) S^{(2)}_{t_0,t_3,t_4}(\tau) dG(\tau)\}\{1+o_p(1)\} \\ &\equiv \{G^{(1)}_{n,2} + G^{(2)}_{n,2} + 2G^{(3)}_{n,2}\}\{1+o_p(1)\}. \end{aligned}$$

Note that $G^{(1)}_{n,2} + G^{(2)}_{n,2} + G^{(3)}_{n,2}$ is nonnegative, it suffices to show that $E\left[G^{(i)}_{n,2}\right] = o(1)$, $i = 1,\ 2$, and 3. To show $EG^{(1)}_{n,2} \equiv E[G^{(1)}_{n,2}] = o(1)$, we consider two different cases: (a) for each $i \in \{0,1,2,3,4\}$, $|t_i - t_j| > m$ for all $j \neq i$; and (b) all the other remaining cases. We will use $EG^{(1s)}_{n,2}$ to denote these cases $(s = a,b)$. For case (a), noting that $ES^{(1)}_{\overline{t_0},\overline{t_1},\overline{t_2}}(\tau) = 0$ (recall that bar notation means *i.i.d.* sequence), we have by Lemma $A.2$, $EG^{(1a)}_{n,2} \leq Cn(h_1^{-(d_1+d_3)})^{4\delta/(1+\delta)}\beta_m^{\delta/(1+\delta)} = o(n^5\beta_m^{\delta/(1+\delta)}) = o(1)$. For case (b), the number of terms in the summation is of order $O(n^4 m)$. One can use Lemma A.2 repeatedly to get $EG^{(1b)}_{n,2} \leq Cn^{-4}h_1^{(d_1+d_3)/2}(n^4m + n^3m^2 + n^3mh_1^{-(d_1+d_3)} + n^2h_1^{-2(d_1+d_3)}) + o(n^5\beta_m^{\delta/(1+\delta)}) = o(1)$. Next, let $S^{(2)}_{t_0,t_1,t_2,t_3,t_4} \equiv E\int S^{(2)}_{t_0,t_1,t_2}(\tau)S^{(2)}_{t_0,t_3,t_4}(\tau)dG(\tau)$. This term is bounded by Ch_2^{2r} if $\{t_1,t_2\}\cap\{t_3,t_4\} \neq \{t_1,t_2\}$ and $t_1 \neq t_0 \neq t_3$; by $Ch_2^{2-d_1}$ if $\{t_1,t_2\}\cap\{t_3,t_4\} = \{t_1,t_2\}$ and $t_1 \neq t_0 \neq t_3$; by $Ch_1^{-(d_1+d_3)}h_2^{2r}$ if $\{t_1,t_2\}\cap\{t_3,t_4\} \neq \{t_1,t_2\}$ and either t_1 or $t_3 = t_0$ but not both. The other cases are of smaller orders after summation. Consequently, $E\left[G^{(2)}_{n,2}\right] = h_1^{(d_1+d_3)/2}n^{-4}O(n^5h_2^{2r} + n^3h_2^{2-d_1} + n^4h_1^{-(d_1+d_3)}h_2^{2r}) = o(1)$. Similarly, one can show that $E\left[G^{(3)}_{n,2}\right] = o(1)$ and (C.9) follows.

To show (C.10), we write

$$\begin{aligned} & nh_1^{(d_1+d_3)/2} G_{n,3} \\ = {} & h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1,t_2,t_3,t_4} \int \widehat{f}^{-1}_{1t_1} \widehat{f}^{-1}_{2t_3} K_{1t_1t_2} K_{1t_1t_3} K_{2t_3t_4} \varepsilon_{t_2}(\tau)\{\varepsilon_{t_3}(\tau) + [m(X_{t_3};\tau) - m(X_{t_4};\tau)]\} a_{t_1} dG(\tau) \\ = {} & \{h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1,t_2,t_3,t_4} \int f^{-1}_{1t_1} f^{-1}_{2t_3} K_{1t_1t_2} K_{1t_1t_3} K_{2t_3t_4} \varepsilon_{t_2}(\tau)\varepsilon_{t_3}(\tau) a_{t_1} dG(\tau) \\ & + h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1,t_2,t_3,t_4} \int f^{-1}_{1t_1} f^{-1}_{2t_3} K_{1t_1t_2} K_{1t_1t_3} K_{2t_3t_4} \varepsilon_{t_2}(\tau)[m(X_{t_3};\tau) - m(X_{t_4};\tau)] a_{t_1} dG(\tau)\}\{1+o_p(1)\} \\ \equiv {} & \{G^{(1)}_{n,3} + G^{(2)}_{n,3}\}\{1+o_p(1)\}, \end{aligned}$$

Straightforward but tedious calculations show that $E(G^{(i)}_{n,3}) = o(1)$ and $E(G^{(i)}_{n,3})^2 = o(1)$ for $i = 1$ and 2, implying that (C.10) holds. The proof is complete. ■

References

[1] Agresti, A. (1990), *Categorical Data Analysis*, John Wiley, New York.

[2] Ait-Sahalia, Y., P. J. Bickel and T. M. Stoker (2001), Goodness-of-fit for Kernel Regression with an Application to Option Implied Volatilities, *Journal of Econometrics* 105, 363-412.

[3] Alba, M. V., D. Barrera and M. D. Jimenez (2001), A Homogeneity Test Based on Empirical Characteristic Functions, *Communicational Statistics* 16, 255-270.

[4] Akaike, H (1972), Information Theory and an Extension of the Maximum Likelihood Principle. In: B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, pp. 267-281.

[5] Andrews, D. W. K. (1994), Empirical Process Methods in Econometrics, In R. F. Engle & D. McFadden (eds.), *Handbook of Econometrics* v.4, 2247–2294. New York: North Holland.

[6] Andrews, D. W. K. (1995), Nonparametric Kernel Estimation for Semiparametric Models, *Econometric Theory* 11, 560-596.

[7] Andrews, D. W. K. and W. Ploberger (1994), Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative, *Econometrica* 62, 1383-1414.

[8] Auestad, B. and D. Tjostheim (1990), Identification of Nonlinear Time Series: First Order Characterization and Order Determination, *Biometrika* 77, 669-687.

[9] Baek, E., and W. Brock (1992), A General Test for Nonlinear Granger Causality: Bivariate model, *Discussion Paper*, Iowa State University and University of Wisconsin, Madison.

[10] Baringhaus, L. and N. Henze (1988), A Consistent Test for Multivariate Normality Based on the Empirical Characteristic Function, *Metrika* 35, 339-348.

[11] Bierens H. J. (1982), Consistent Model Specification Test, *Journal of Econometrics 20, 105-134.*

[12] Bierens H.J. (1984), Model Specification Testing of Time Series Regressions, *Journal of Econometrics* 26, 323-353.

[13] Bierens, H. J. (1987), ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands, *Journal of Econometrics* 35, 161-191.

[14] Bierens, H. J. (1990), A Consistent Conditional Moment Test of Functional Form, *Econometrica 58, 1443-1458.*

[15] Bierens, H. J. and J. Hartog(1988), Nonlinear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function, *Journal of Econometrics 38, 269-299.*

[16] Bierens, H. J. and W. Ploberger (1997) Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica* 65, 1129-1152.

[17] Blum, J. R., J. Kiefer, and M. Rosenblatt (1961), Distribution Free Tests of Independence Based on the Sample Distribution Function, *Annals of Mathematical Statistics* 32, 485-498.

[18] Boning, W. B. and F. Sowell (1999), Optimality for the Integrated Conditional Moment Test, *Econometric Theory* 15, 710-718.

[19] Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.

[20] Brock, W., Dechert, W. and Scheinkman, J. (1987), A Test for Independence Based on the Correlation Dimension, *Working Paper*, University of Wisconsin, Madison.

[21] Brett, C. and J. Pinkse (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.

[22] Cheng, B. and H. Tong (1992), On Consistent Nonparametric Order Determination and Chaos, *Journal of the Royal Statistical Society B* 54, 427-449.

[23] Collomb, G. and W. Härdle (1986), Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: Kernel Regression Estimation from Dependent Observations, *Stochastic Processes and Their Applications* 23, 77-89.

[24] Csörgo, S. (1985), Testing for Independence by the Empirical Characteristic Function, *Journal of Multivariate Analysis* 16, 290-299.

[25] de Jong, P. (1987), A Central Limit Theorem for Generalized Quadratic Forms, *Probability Theory and Related Fields* 75, 261-277.

[26] de Jong, R. M. (1996), The Bierens Test under Data Dependence, *Journal of Econometrics* 72, 1-32.

[27] Delgado, M. (1996), Testing Serial Independence Using the Sample Distribution Function, *Journal of Time Series Analysis* 17, 271-287.

[28] Fan, Y. (1997), Goodness-of-fit Tests for a Multivariate Distribution by the Empirical Characteristic Function, *Journal of Multivariate Analysis* 62, 36-63.

[29] Fan Y., and R. Gencay (1993), Hypothesis Testing Based on Modified Nonparametric Estimation of an Affine Measure between Two Distributions, *Journal of Nonparametric Statistics* 2, 389-403.

[30] Fan Y. and Q. Li (1999a), Central Limit Theorem for Degenerate U-statistics of Absolutely Regular Processes with Applications to Model Specification Testing, *Journal of Nonparametric Statistics* 10, 245-271.

[31] Fan Y. and Q. Li (1999b), Root-N-consistent Estimation of Partially Linear Time Series Models, *Journal of Nonparametric Statistics* 11, 251-269.

[32] Feuerverger, A. (1987), On Some ECF Procedures for Testing Independence, in *Time Series and Econometric Modelling*, eds. I. B. MacNeill and G. J. Umphrey, Boston Reidel, pp. 189-206.

[33] Feuerverger, A. and R. A. Mureika (1977), The Empirical Characteristic Function and its Applications, *Annals of Statistics* 5, 88-97.

[34] Finkenstadt, B. F., Q. Yao, and H. Tong (2001), A Conditional Density Approach to the Order Determination of Time Series, *Statistics and Computing* 11, 229-240.

[35] Florens, J. P., and D. Fougere (1996), Noncausality in Continuous Time, *Econometrica 64, 1195-1212.*

[36] Florens, J. P., and M. Mouchart (1982), A Note on Causality, *Econometrica* 50, 583-591.

[37] Ghosh, S. and F. H. Ruymgaart (1992), Applications of Empirical Characteristic Functions in Some Multivariate Problems, *Canadian Journal of Statistics* 20, 429-440.

[38] Gourieroux. C., and C. Tenreiro (2001), Local Power Properties of Kernel Based Goodness of Fit Tests, *Journal of Multivariate Analysis* 78, 161-190.

[39] Granger, C. W. J. (1969), Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica* 37, 424-438.

[40] Granger, C. W. J. (1980), Testing for Causality: A Personal Viewpoint, *Journal of Economic Dynamics and Control* 2, 1980, 329-352.

[41] Gurtler, N. and N. Henze (2000), Goodness-of-fit Tests for the Cauchy Distribution Based on the Empirical Characteristic Function, *Annals of the Institute of Statistical Mathematics* 52, 267-286.

[42] Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer-Verlag. Berlin.

[43] Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press. New York.

[44] Härdle, W., and E. Mammen (1993), Comparing Nonparametric versus Parametric Regression Fits, *Annals of Statistics* 21, 1926-1947.

[45] Harel, M. and M. L. Puri (1996), Conditional U-statistics for Dependent Random Variables, *Journal of Multivariate Analysis* 57, 84-100.

[46] Heathcote, C. R., S. T. Rachev and B. Cheng (1995), Testing Multivariate Symmetry, *Journal of Multivariate Analysis 54, 91-112.*

[47] Henze, N. and B. Zirkler (1990), A Class of Invariant Consistent Tests for Multivariate Normality, *Communications in Statistics–Theory and Methodology* 19, 3595-3617.

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

[49] Hoeffding, W. (1948), A Non-parametric Test of Independence, *Annals of Mathematical Statistics, 58, 546-557.*

[50] 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.

[51] Hong, Y. (2001), A Test for Volatility Spillover with Application to Exchange Rates, *Journal of Econometrics* 103, 183-224.

[52] Hong, Y. and H. White (2000), Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence, *Discussion Paper*, Department of Economics, Cornell University and UCSD.

[53] Hsieh, D. A. (1989), Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *Journal of Business* 62, 339-368.

[54] Hsieh, D. A. (1991), Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance* 5, 1839-1877.

[55] Inoue, A. (1998), A Conditional Goodness-of-fit Test for Time Series, Discussion Paper, Department of Agricultural and Resource Economics, North Carolina State University.

[56] Khashimov, S. A. (1992), Limiting Behavior of Generalized U-statistics of Weakly Dependent Stationary Processes, *Theory of Probability and its Applications* 37, 148-150.

[57] Koutrouvelis, I. A. (1980), A Goodness-of-fit Test of Simple Hypotheses Bases on the Empirical Characteristic Function, *Biometrika* 67, 238-240.

[58] Koutrouvelis, I. A. and J. Kellermeier (1981), A Goodness-of-fit Test Based on the Empirical Characteristic Function When Parameters Must Be Estimated, *Journal of the Royal Statistical Society* 43, 173-176.

[59] Li, Q. (1999), Consistent Model Specification Tests for Time Series Econometric Models, *Journal of Econometrics* 92, 101-147.

[60] Liebscher, E. (1996), Strong Convergence of Sums of $\alpha-$mixing Random Variables with Applications to Density Estimation, *Stochastic Process and their Applications* 65, 69-80.

[61] Linton, O. (2002), Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics, *Journal of Econometrics* 106, 325-368.

[62] Linton, O. and P. Gozalo (1997), Conditional Independence Restrictions: Testing and Estimation, *Discussion Paper*, Cowles Foundation for Research in Economics, Yale University.

[63] Lobato, I. (2003), Testing for Nonlinear Autoregression, *Journal of Business and Economic Statistics* 21, 164-173.

[64] Nishiyama, Y. and P. M. Robinson (2000), Edgeworth Expansions for Semiparametric Averaged Derivatives, *Econometrica* 68, 931-980.

[65] Naito, K. (1996a), On Weighting the Studentized Empirical Characteristic Function for Testing Normality, *Communications in Statistics–Simulations* 25, 201-213.

[66] Naito, K. (1996b), Modification of Statistics Based on the Empirical Characteristic Function to Yield Asymptotic Normality, *Communications in Statistics–Theory and Methodology* 25, 105-114.

[67] Pagan, A., and A. Ullah (1999), *Nonparametric Econometrics,* Cambridge University Press.

[68] Paparoditis, E. and D. N. Politis (2000), The Local Bootstrap for Kernel Estimators under General Dependence Conditions, *Annals of the Institute of Statistical Mathematics* 52, 139-159.

[69] Pinkse, J. (1998), A Consistent Nonparametric Test for Serial Independence, *Journal of Econometrics 84, 205-231.*

[70] Pinkse, J. (2002), Nonparametric Misspecification Testing, Discussion paper, Dept. of Economics, University of British Columbia.

[71] Racine, J. and Q. Li (2000), Nonparametric Estimation of Conditional Distributions with Mixed Categorical and Continuous Data, *Discussion Paper*, Dept. of Economics, Univ. of South Florida and Texas A & M Univ.

[72] Robinson, P. M. (1988), Root-N-consistent Semiparametric Regression, *Econometrica* 56, 931-954.

[73] Robinson P. M. (1991), Consistent Nonparametric Entropy-based Testing, *Review of Economic Studies* 58, 437-453.

[74] Rosenbaum, P. R. (1984), Testing the Conditional Independence and Monotonicity Assumptions of Item Response Theory, *Psychometrika* 49, 425-435.

[75] Rosenbaum, P. R., and D. B. Rubin (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70, 41-55.

[76] Sheedy, E. (1998), Correlation in Currency Markets: a Risk-Adjusted Perspective, *Journal of International Financial Markets, Institutions & Money* 8, 59-82.

[77] Skaug, H. J. and D. Tjostheim (1993), A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function, *Biometrika* 80, 591-602.

[78] Skaug, H. J. and D. Tjostheim (1996), Measure of Distance between Densities with Application to Testing for Serial Independence, In P. M. Robinson and M. Rosenblatt (eds.), *Time series Analysis in Memory of E. J. Hannan*, 363-377, Springer : New York.

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

[80] Su, L. and H. White (2002), A Hellinger-Metric Nonparametric Test for Conditional Independence, *Working Paper*, Dept. of Economics, UCSD.

[81] Tauchen, G. (1985), Diagnostic Testing and Evaluation of Maximum Likelihood Models, *Journal of Econometrics 30, 415-443.*

[82] Tenreiro, C (1995), Theoremes Limites pour les Erreurs Quadratiques Integrees des Estimateurs a Noyau de la Densite et de la Regression sous des Conditions de Dependance, *C. R. Acad. Sci. Paris* 320, 1535-1538.

[83] 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.

[84] Tjostheim, D (1996), Measures and Tests of Independence: a Survey, *Statistics* 28, 249-284.

[85] Tjostheim, D and B. Auestad (1994a), Nonparametric Identification of Nonlinear Time Series: Projections, *Journal of the American Statistical Association* 89, 1398-1409.

[86] Tjostheim, D and B. Auestad (1994b), Nonparametric Identification of Nonlinear Time Series: Selecting Significant Lags, *Journal of the American Statistical Association* 89, 1410-1419.

[87] Tschernig, R. and L. Yang (2000), Nonparametric Lag Selection for Time Series, *Journal of Time Series Analysis* 21, 457-487.

[88] White, H. (2000), *Asymptotic Theory for Econometricians*, Academic Press.

[89] White, H., and Y. Hong (1999), $M-$testing Using Finite and Infinite Dimensional Parameter Estimators, In R. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Oxford: Oxford University Press, 326-345.

[90] Yao, Q. and D. Tritchler (1993), An Exact Analysis of Conditional Independence in Several 2×2 Contingency Tables, *Biometrics* 49, 233-236.

[91] Yoshihara, K (1976), Limiting Behavior of U-statistics for Stationary, Absolutely Regular Processes, *Z. Wahrsch. Verw. Gebiete* 35, 237-252.

[92] Yoshihara, K (1989), Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Processes, In *Proceedings of the Fourth Prague Conference on Asymptotic Statistics*, 539-547.

[93] Zheng, J. Z. (1997), A Consistent Specification Test of Independence, *Nonparametric Statistics* 7, 297-306.

Table 1: Comparison of Tests for Causality (d1=d2=d3=1)

	5%							
	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100								
CM_n	0.054	0.058	0.920	0.548	0.504	0.412	0.384	0.188
KS_n	0.042	0.056	0.780	0.404	0.380	0.288	0.292	0.156
H_n	0.072	0.055	0.412	0.658	0.454	0.920	0.728	0.560
T_{1n}	0.078	0.079	0.792	0.844	0.592	0.944	0.656	0.420
T_{2n}	0.065	0.067	0.768	0.888	0.620	0.960	0.656	0.412
n=200								
CM_n	0.045	0.056	0.992	0.748	0.788	0.680	0.476	0.360
KS_n	0.067	0.053	0.952	0.552	0.660	0.532	0.336	0.284
H_n	0.045	0.025	0.829	0.870	0.485	1	0.984	0.852
T_{1n}	0.063	0.062	0.940	0.968	0.848	1	0.864	0.604
T_{2n}	0.050	0.076	0.954	0.99	0.882	1	0.872	0.608
n=500								
CM_n	0.025	0.046	1	0.984	0.992	0.984	0.824	0.728
KS_n	0.042	0.044	1	0.884	0.976	0.912	0.76	0.592
H_n	0.034	0.030	0.994	0.997	0.909	1	1	1
T_{1n}	0.048	0.032	1	1	1	1	1	0.876
T_{2n}	0.050	0.032	1	1	1	1	1	0.88
n=1000								
CM_n	0.048	0.042	1	1	1	1	0.976	0.904
KS_n	0.042	0.044	1	0.992	1	1	0.952	0.804
H_n	0.052	0.026	1	1	1	1	1	1
T_{1n}	0.036	0.024	1	1	1	1	1	1
T_{2n}	0.039	0.048	1	1	1	1	1	1
$ISELR_n$	0.063	0.052	1	1	1	1	1	1

Table 1: Comparison of Tests for Causality (d1=d2=d3=1, cont.)

	10%							
	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100								
CM_n	0.094	0.101	0.964	0.652	0.644	0.480	0.472	0.304
KS_n	0.100	0.111	0.868	0.492	0.496	0.428	0.408	0.232
H_n	0.106	0.085	0.468	0.718	0.529	0.928	0.772	0.636
T_{1n}	0.122	0.115	0.844	0.888	0.688	0.976	0.728	0.552
T_{2n}	0.117	0.114	0.828	0.932	0.696	0.98	0.744	0.558
n=200								
CM_n	0.095	0.100	1	0.856	0.904	0.752	0.592	0.508
KS_n	0.096	0.089	0.988	0.676	0.756	0.676	0.484	0.404
H_n	0.08	0.05	0.889	0.912	0.602	1	1	0.904
T_{1n}	0.100	0.101	0.960	0.988	0.908	1	0.924	0.708
T_{2n}	0.092	0.117	0.98	0.996	0.918	1	0.928	0.72
n=500								
CM_n	0.052	0.065	1	0.992	1	0.992	0.896	0.792
KS_n	0.066	0.068	1	0.944	0.992	0.960	0.836	0.696
H_n	0.075	0.055	0.996	0.997	0.951	0.998	1	1
T_{1n}	0.068	0.052	1	1	1	1	1	0.924
T_{2n}	0.069	0.057	1	1	1	1	1	0.932
n=1000								
CM_n	0.112	0.076	1	1	1	1	0.992	0.936
KS_n	0.088	0.074	1	1	1	1	0.968	0.872
H_n	0.078	0.070	1	1	1	1	1	1
T_{1n}	0.080	0.076	1	1	1	1	1	1
T_{2n}	0.068	0.092	1	1	1	1	1	1

Table 2: Comparison of Tests for Causality (d1=2, d2=d3=1)

	5%							
	DGP1'	DGP2'	DGP3'	DGP4'	DGP5'	DGP6'	DGP7'	DGP8'
n=100								
CM_n	0.029	0.040	0.684	0.368	0.204	0.608	0.248	0.344
KS_n	0.032	0.045	0.440	0.304	0.108	0.512	0.200	0.312
T_{1n}	0.071	0.072	0.708	0.812	0.532	0.800	0.396	0.396
T_{2n}	0.058	0.048	0.566	0.716	0.402	0.700	0.352	0.324
n=200								
CM_n	0.053	0.028	0.964	0.656	0.352	0.872	0.336	0.432
KS_n	0.040	0.025	0.792	0.480	0.196	0.780	0.292	0.388
T_{1n}	0.049	0.051	0.920	0.972	0.800	0.988	0.588	0.640
T_{2n}	0.076	0.064	0.912	0.968	0.788	0.968	0.596	0.620
n=500								
CM_n	0.068	0.028	1	0.936	0.692	0.992	0.632	0.764
KS_n	0.056	0.040	0.996	0.784	0.456	0.988	0.552	0.656
T_{1n}	0.018	0.052	1	1	0.996	1	0.872	0.944
T_{2n}	0.078	0.077	1	1	0.996	1	0.896	0.94
n=1000								
CM_n	0.076	0.033	1	1	0.952	1	0.912	0.968
KS_n	0.062	0.048	1	0.952	0.856	1	0.832	0.944
T_{1n}	0.014	0.076	1	1	1	1	1	1
T_{2n}	0.068	0.076	1	1	1	1	1	1

Table 2: Comparison of Tests for Causality (d1=2, d2=d3=1, cont.)

	10%							
	DGP1'	DGP2'	DGP3'	DGP4'	DGP5'	DGP6'	DGP7'	DGP8'
n=100								
CM_n	0.068	0.088	0.788	0.472	0.332	0.712	0.344	0.428
KS_n	0.064	0.076	0.632	0.408	0.224	0.624	0.296	0.400
T_{1n}	0.106	0.114	0.756	0.880	0.624	0.892	0.524	0.504
T_{2n}	0.099	0.074	0.668	0.776	0.496	0.832	0.448	0.432
n=200								
CM_n	0.104	0.060	0.984	0.756	0.468	0.908	0.456	0.528
KS_n	0.104	0.064	0.896	0.588	0.332	0.848	0.408	0.468
T_{1n}	0.084	0.105	0.956	0.992	0.872	1.000	0.716	0.792
T_{2n}	0.136	0.107	0.946	0.982	0.850	0.984	0.720	0.744
n=500								
CM_n	0.127	0.056	1	0.952	0.824	1	0.732	0.824
KS_n	0.104	0.061	1	0.876	0.62	1	0.644	0.772
T_{1n}	0.042	0.082	1	1	0.998	1	0.904	0.968
T_{2n}	0.133	0.132	1	1	0.996	1	0.936	0.976
n=1000								
CM_n	0.112	0.066	1	1	0.992	1	0.964	0.984
KS_n	0.092	0.065	1	0.992	0.924	1	0.920	0.960
T_{1n}	0.016	0.104	1	1	1	1	1	1
T_{2n}	0.116	0.136	1	1	1	1	1	1

Table 3: Comparison of Tests for Causality (d1=3, d2=d3=1)

	5%							
	DGP1"	DGP2"	DGP3"	DGP4"	DGP5"	DGP6"	DGP7"	DGP8"
n=200								
CM_n	0.036	0.044	0.724	0.368	0.236	0.772	0.320	0.440
KS_n	0.052	0.04	0.344	0.276	0.128	0.684	0.256	0.408
T_{1n}	0.039	0.021	0.872	0.960	0.612	0.904	0.408	0.480
T_{2n}	0.052	0.017	0.842	0.914	0.574	0.872	0.348	0.454
n=500								
CM_n	0.044	0.055	1.000	0.852	0.464	0.996	0.544	0.808
KS_n	0.047	0.042	0.872	0.656	0.280	0.976	0.472	0.740
T_{1n}	0.044	0.020	1	1	0.964	1	0.792	0.884
T_{2n}	0.044	0.015	1	1	0.96	1	0.724	0.836
n=1000								
CM_n	0.056	0.032	1	0.976	0.804	1	0.776	0.960
KS_n	0.078	0.016	1	0.864	0.576	1	0.752	0.944
T_{1n}	0.040	0.088	1	1	1	1	0.884	1
T_{2n}	0.032	0.046	1	1	1	1	0.856	0.996

	10%							
	DGP1"	DGP2"	DGP3"	DGP4"	DGP5"	DGP6"	DGP7"	DGP8"
n=200								
CM_n	0.084	0.068	0.832	0.532	0.380	0.832	0.416	0.540
KS_n	0.116	0.06	0.572	0.420	0.248	0.788	0.376	0.480
T_{1n}	0.061	0.042	0.888	0.976	0.696	0.952	0.492	0.636
T_{2n}	0.099	0.038	0.892	0.934	0.666	0.952	0.456	0.566
n=500								
CM_n	0.081	0.068	1.000	0.892	0.656	0.996	0.640	0.856
KS_n	0.096	0.057	0.944	0.784	0.444	0.988	0.588	0.824
T_{1n}	0.074	0.038	1	1	0.976	1	0.816	0.960
T_{2n}	0.09	0.035	1	1	0.972	1	0.776	0.880
n=1000								
CM_n	0.08	0.072	1	0.992	0.888	1	0.846	0.976
KS_n	0.112	0.044	1	0.920	0.704	1	0.800	0.966
T_{1n}	0.100	0.144	1	1	1	1	0.892	1
T_{2n}	0.058	0.066	1	1	1	1	0.912	0.996

Table 4: Test of the order of nonlinear AR processes (5% test)

		DGP9	DGP10	DGP11	DGP12	DGP13	DGP14	DGP15	DGP16	DGP17	DGP18	DGP19	DGP20
							$H_0(0)$: d* = 0						
n=100	T_{1n}	0.032	0.248	0.172	0.272	0.238	0.494	0.998	0.97	1	1	1	0.996
	T_{2n}	0.034	0.202	0.174	0.292	0.23	0.504	1	0.978	1	1	1	0.992
n=200	T_{1n}	0.03	0.474	0.318	0.622	0.454	0.83	1	1	1	1	1	1
	T_{2n}	0.044	0.438	0.314	0.624	0.508	0.802	1	1	1	1	1	1
n=500	T_{1n}	0.048	0.904	0.804	0.992	0.916	0.996	1	1	1	1	1	1
	T_{2n}	0.052	0.92	0.792	0.988	0.928	1	1	1	1	1	1	1
							$H_0(1)$: d* = 1						
n=100	T_{1n}	0.062	0.08	0.104	0.072	0.084	0.084	0.568	0.242	0.324	0.212	0.716	0.864
	T_{2n}	0.072	0.066	0.088	0.084	0.060	0.080	0.262	0.204	0.222	0.092	0.556	0.824
n=200	T_{1n}	0.06	0.052	0.092	0.078	0.056	0.084	0.928	0.5	0.634	0.514	0.982	0.99
	T_{2n}	0.060	0.072	0.094	0.070	0.060	0.088	0.594	0.434	0.508	0.406	0.958	0.974
n=500	T_{1n}	0.056	0.064	0.104	0.064	0.044	0.068	0.960	0.820	0.948	0.940	0.996	0.960
	T_{2n}	0.022	0.034	0.076	0.066	0.058	0.058	0.984	0.776	0.884	0.936	0.988	0.972
							$H_0(2)$: d* = 2						
n=100	T_{1n}	0.078	0.054	0.076	0.07	0.052	0.076	0.026	0.036	0.022	0.01	0.014	0.02
	T_{2n}	0.05	0.058	0.058	0.06	0.064	0.056	0.022	0.018	0.032	0.016	0.004	0.016
n=200	T_{1n}	0.052	0.036	0.08	0.066	0.056	0.058	0.07	0.036	0.106	0.038	0.026	0.088
	T_{2n}	0.084	0.064	0.096	0.07	0.082	0.086	0.036	0.042	0.052	0.024	0.028	0.062
n=500	T_{1n}	0.036	0.044	0.102	0.04	0.04	0.036	0.44	0.096	0.108	0.096	0.098	0.132
	T_{2n}	0.076	0.094	0.106	0.1	0.072	0.104	0.104	0.118	0.124	0.096	0.048	0.122

DGP 9 is of order 1, DGPs 10-14 are of order 2 and DGPs 15-20 are of order 3. Consequently, only DGP 9 satisfies $H_0(0)$: d*= 0, DGPs 9-14 satisfy $H_0(1)$: d* =1 and all DGPs here satisfy $H_0(2)$: d* = 2.

Table 5: Bivariate linear Granger-Causality test between exchange rates

H_0: Row doesn't cause column		Canada	France	Germany	Italy	UK
Canada	L_y=1, DX_{t-1} used	-	2.65	2.34	**7.36[a]**	3.47[c]
	L_y=2, DX_{t-1} used	-	2.65	2.35	7.56[a]	3.47[c]
	L_y=2, DX_{t-2} used	-	0.22	1.62	2.77	0.66
	L_y=3, DX_{t-1} used	-	2.59	2.27	7.56[a]	3.52[c]
	L_y=3, DX_{t-2} used	-	0.22	1.61	2.68	0.62
	L_y=3, DX_{t-3} used	-	0.15	0.29	0.00	0.40
France	L_y=1, DX_{t-1} used	2.24	-	0.03	3.19[c]	0.00
	L_y=2, DX_{t-1} used	2.37	-	0.04	3.57[c]	0.00
	L_y=2, DX_{t-2} used	2.67	-	0.77	3.16[c]	**3.88[b]**
	L_y=3, DX_{t-1} used	2.40	-	0.05	3.59[c]	0.00
	L_y=3, DX_{t-2} used	2.59	-	0.73	3.27[c]	**3.89[b]**
	L_y=3, DX_{t-3} used	0.00	-	5.51[b]	0.06	2.16
Germany	L_y=1, DX_{t-1} used	1.51	4.52[b]	-	3.95[b]	0.55
	L_y=2, DX_{t-1} used	1.58	**4.53[b]**	-	**4.18[b]**	0.54
	L_y=2, DX_{t-2} used	1.82	1.48	-	2.09	1.91
	L_y=3, DX_{t-1} used	1.70	**4.61[b]**	-	4.31[b]	0.52
	L_y=3, DX_{t-2} used	1.69	1.47	-	2.22	1.91
	L_y=3, DX_{t-3} used	0.12	8.79[a]	-	0.24	3.05[c]
Italy	L_y=1, DX_{t-1} used	2.18	2.75[c]	0.03	-	1.84
	L_y=2, DX_{t-1} used	2.32	2.76[c]	0.03	-	1.90
	L_y=2, DX_{t-2} used	0.16	0.52	0.07	-	0.17
	L_y=3, DX_{t-1} used	2.48	2.76[c]	0.03	-	1.80
	L_y=3, DX_{t-2} used	0.11	0.54	0.08	-	0.15
	L_y=3, DX_{t-3} used	0.00	0.49	2.00	-	0.69
UK	L_y=1, DX_{t-1} used	0.48	0.20	0.38	0.16	-
	L_y=2, DX_{t-1} used	0.50	0.19	0.38	0.12	-
	L_y=2, DX_{t-2} used	0.64	0.45	0.02	0.18	-
	L_y=3, DX_{t-1} used	0.51	0.21	0.39	0.10	-
	L_y=3, DX_{t-2} used	0.62	0.46	0.02	0.16	-
	L_y=3, DX_{t-3} used	0.00	0.70	0.05	1.12	-

The superscripts a, b and c denote rejection of the noncasusality hypothesis at 1%, 5% and 10% significance levels, respectively. The bold elements indicate thecausal links detected by the linear causality test but not by the our nonparametric causality test at the 5% significance level.

Table 6: Bivariate nonparametric Granger-Causality test between exchange rates

H0: Row doesn't cause column		Canada	France	Germany	Italy	UK
Canada	L_y=1, DX_{t-1} used	-	0.77	-0.06	-0.07	0.04
	L_y=2, DX_{t-1} used	-	0.59	0.18	2.76[a]	0.29
	L_y=2, DX_{t-2} used	-	0.93	1.19	0.22	0.87
	L_y=3, DX_{t-1} used	-	1.27	0.18	1.78[b]	-0.60
	L_y=3, DX_{t-2} used	-	0.35	1.25	0.37	0.30
	L_y=3, DX_{t-3} used	-	0.34	0.38	**1.83[a]**	-0.13
France	L_y=1, DX_{t-1} used	-0.23	-	**3.23[a]**	4.31[a]	0.64
	L_y=2, DX_{t-1} used	0.51	-	**2.53[a]**	**5.41[a]**	**3.73[a]**
	L_y=2, DX_{t-2} used	0.59	-	-1.06	-2.23	0.79
	L_y=3, DX_{t-1} used	-0.34	-	-0.58	**4.47[a]**	1.44[c]
	L_y=3, DX_{t-2} used	-0.23	-	**2.00[b]**	**1.69[b]**	-0.59
	L_y=3, DX_{t-3} used	-1.12	-	**3.29[a]**	**5.28[a]**	**3.39[a]**
Germany	L_y=1, DX_{t-1} used	0.54	3.34[a]	-	3.57[a]	1.05
	L_y=2, DX_{t-1} used	1.54[c]	-0.87	-	-3.64	**2.50[a]**
	L_y=2, DX_{t-2} used	-0.16	**3.15[a]**	-	-5.51	1.41[c]
	L_y=3, DX_{t-1} used	0.30	-0.20	-	4.81[a]	**1.98[b]**
	L_y=3, DX_{t-2} used	0.34	-2.29	-	**2.23[b]**	-0.03
	L_y=3, DX_{t-3} used	-1.08	2.10[b]	-	**6.50[a]**	2.02[b]
Italy	L_y=1, DX_{t-1} used	0.30	**3.56[a]**	**3.10[a]**	-	0.70
	L_y=2, DX_{t-1} used	**2.06[b]**	-1.00	**44.93[a]**	-	**2.02[b]**
	L_y=2, DX_{t-2} used	0.39	**3.92[a]**	**3.19[a]**	-	**1.73[b]**
	L_y=3, DX_{t-1} used	1.57[c]	**3.92[a]**	**3.54[a]**	-	**1.66[b]**
	L_y=3, DX_{t-2} used	**2.16[b]**	**3.67[a]**	**3.64[a]**	-	-0.42
	L_y=3, DX_{t-3} used	-0.81	-5.22	-1.13	-	0.82
UK	L_y=1, DX_{t-1} used	-0.28	0.80	1.06	0.25	-
	L_y=2, DX_{t-1} used	**2.05[b]**	**1.99[b]**	**3.43[a]**	**2.82[a]**	-
	L_y=2, DX_{t-2} used	**2.18[b]**	1.30[c]	**3.39[a]**	**1.93[b]**	-
	L_y=3, DX_{t-1} used	-0.42	**5.44[a]**	**5.27[a]**	**5.20[a]**	-
	L_y=3, DX_{t-2} used	0.31	-1.49	0.73	-0.21	-
	L_y=3, DX_{t-3} used	-1.44	-0.35	-0.73	-1.05	-

The superscripts a, b and c denote rejection of the noncasusality hypothesis at 1%, 5% and 10% significance levels, respectively.The bold elements indicate the causal links detected by our nonparametric causality test but not by the linear causality test at the 5% significance level.

Table 7: Granger causality tests between stock prices and trading volumes

Panel A: Linear Granger causality test between ΔP and ΔV

	H_0: ΔP doesn't cause ΔV			H_0: ΔV doesn't cause ΔP		
	Dow Jones	Nasdaq	S&P 500	Dow Jones	Nasdaq	S&P 500
L_y=1, DX_{t-1} used	0.11	8.28^{a}	**8.63^{a}**	0.66	0.32	0.02
L_y=2, DX_{t-1} used	0	6.85^{a}	9.84^{a}	0.65	0.15	0.04
L_y=2, DX_{t-2} used	1.22	3.63^{c}	3.69^{c}	0.1	0.11	0.03
L_y=3, DX_{t-1} used	0.05	**5.78^{b}**	10.82^{a}	0.65	0.15	0.05
L_y=3, DX_{t-2} used	1.85	3.21^{c}	4.88^{b}	0.1	0.11	0.07
L_y=3, DX_{t-3} used	3.83^{c}	0.02	6.49^{b}	0.07	0	0.02

Panel B: Nonlinear Granger causality test between ΔP and ΔV

	H_0: ΔP doesn't cause ΔV			H_0: ΔV doesn't cause ΔP		
	Dow Jones	Nasdaq	S&P 500	Dow Jones	Nasdaq	S&P 500
L_y=1, DX_{t-1} used	**2.17^{b}**	2.20^{b}	0.80	**1.79^{b}**	-0.04	**1.78^{b}**
L_y=2, DX_{t-1} used	1.29^{c}	4.18^{a}	2.02^{b}	**3.87^{a}**	**3.24^{a}**	1.27
L_y=2, DX_{t-2} used	**2.07^{b}**	**4.06^{a}**	**4.33^{a}**	**2.12^{b}**	1.49^{c}	0.93
L_y=3, DX_{t-1} used	-0.39	1.59^{c}	3.54^{a}	-1.90	1.38^{c}	0.81
L_y=3, DX_{t-2} used	**9.63^{a}**	-0.40	3.53^{a}	0.46	0.85	-0.05
L_y=3, DX_{t-3} used	-2.31	1.45^{c}	3.23^{a}	-0.33	-0.29	0.33

(1)The superscripts a, b and c denote rejection of the noncasusality hypothesis at 1%, 5% and 10% significance levels, respectively.

(2)The bold elements in Panel A indicate the linear causal links that our nonlinear Granger causality test fails to detect at the 5% significane level and those in Panel B indicate the nonlinear causal links where the linear Granger causality test fails at the 5% significance level.

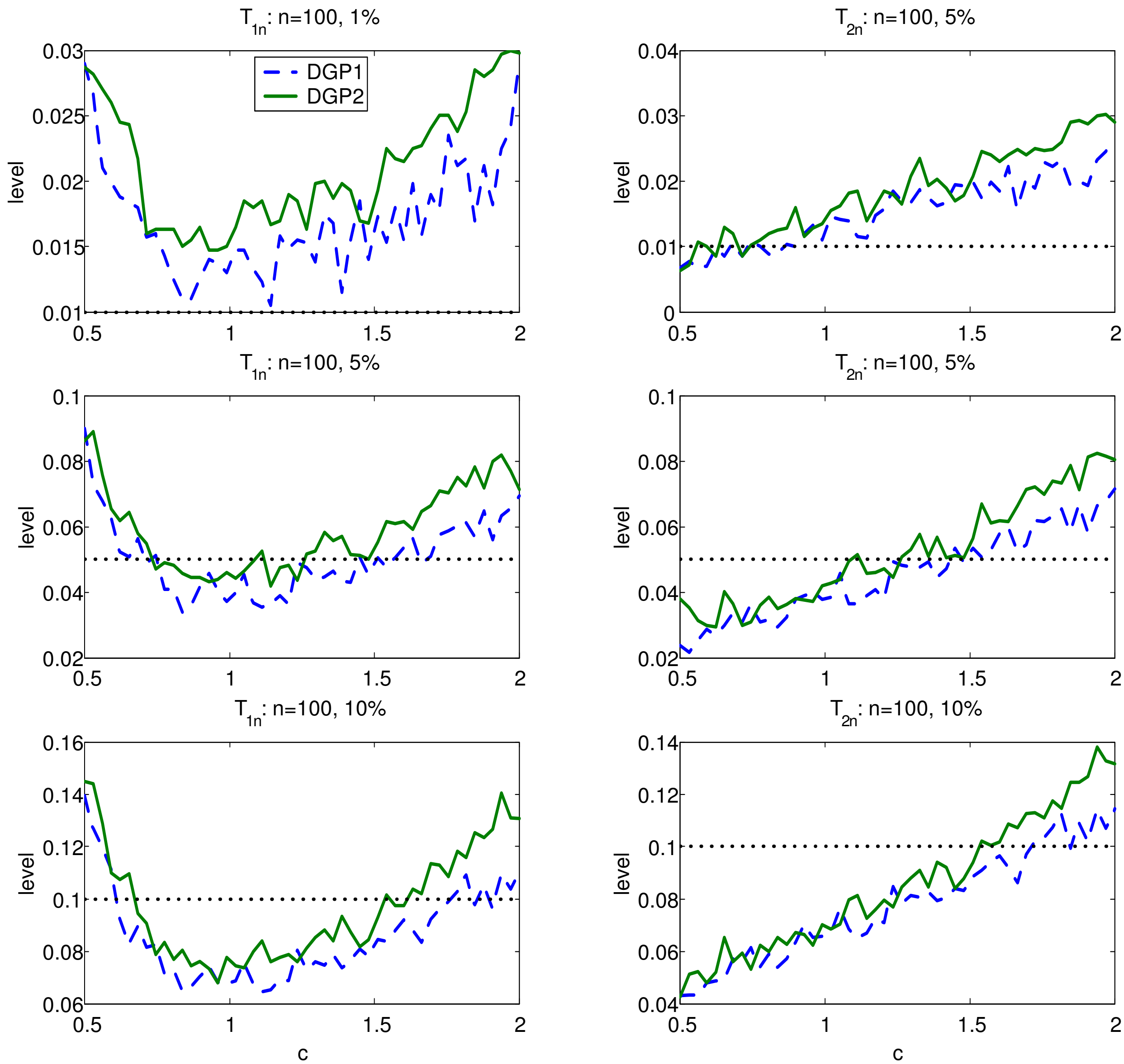

Figure 1: Level of the Tests T_{1n} and T_{2n} ($n = 100$)

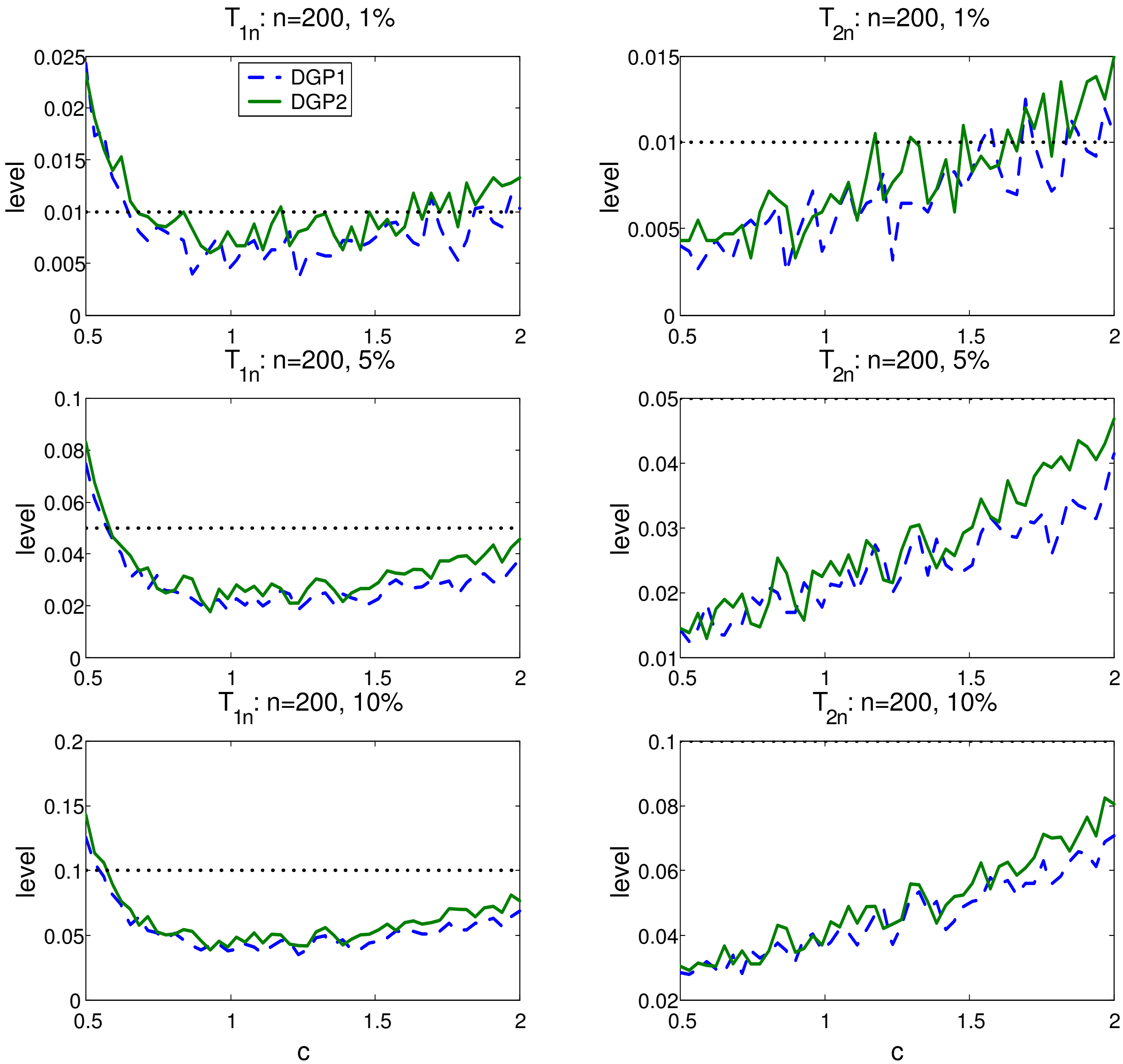

Figure 2: Level of the Tests T_{1n} and T_{2n} ($n = 200$)

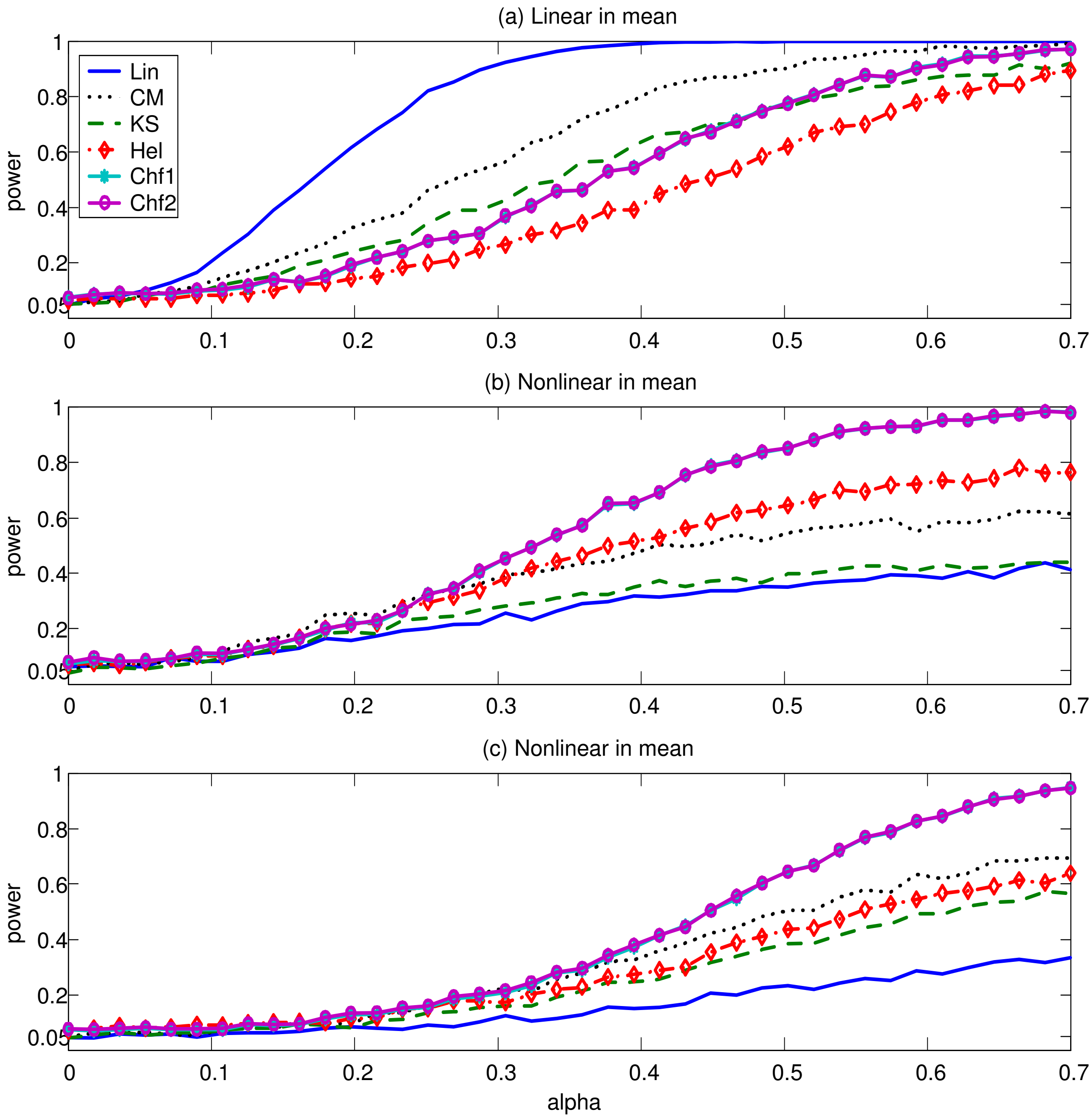

Figure 3: Power Comparison for Linear and Nonlinear Causality Tests ($d_1 = 1,\ n = 100,\ 5\%$)

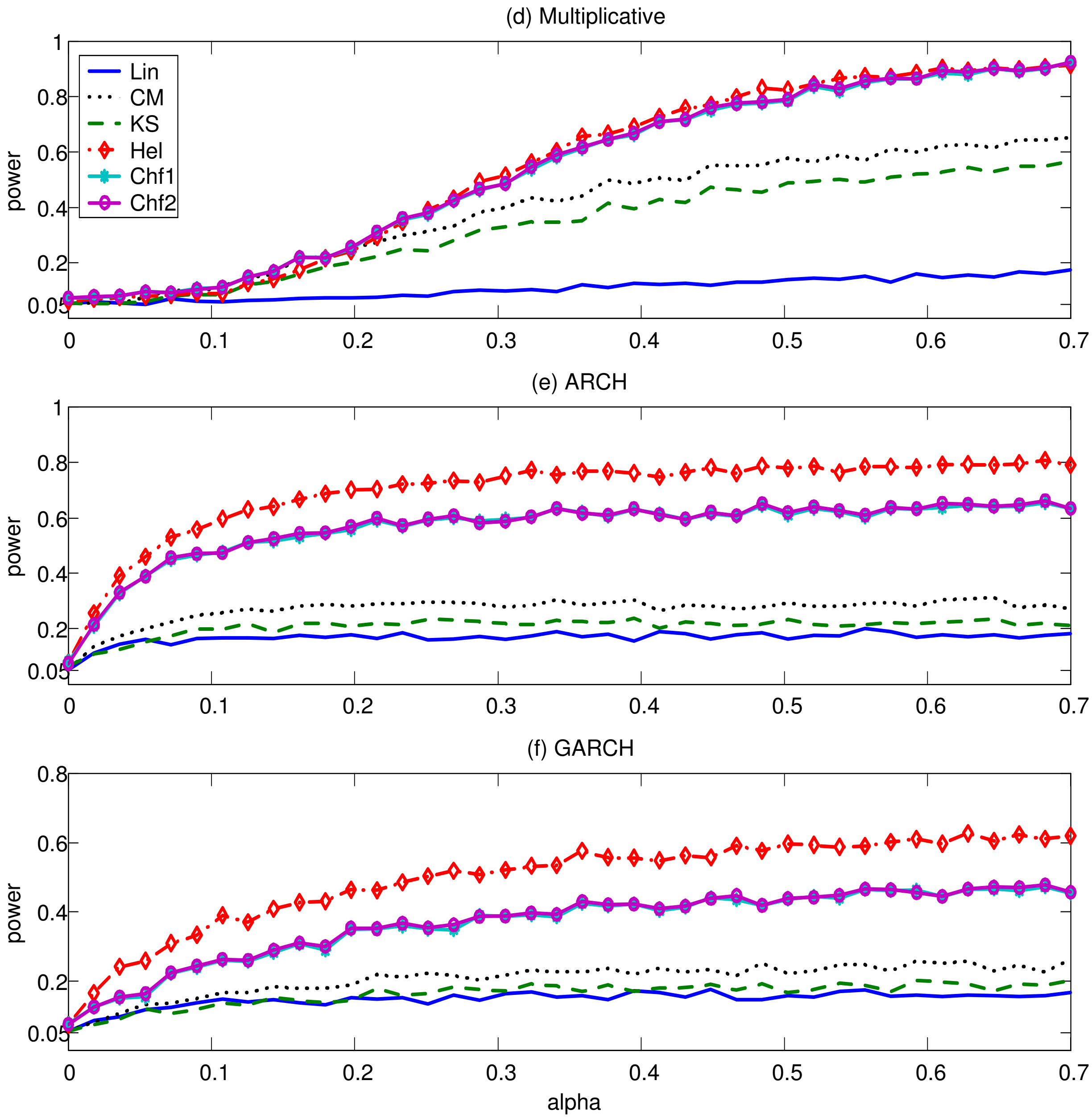

Figure 3: Power Comparison for Linear and Nonlinear Causality Tests ($d_1 = 1,\ n = 100,\ 5\%,$ cont.)