Econometric Theory, **24**, 2008, 829–864. Printed in the United States of America.
doi:10.1017/S0266466608080341

A NONPARAMETRIC HELLINGER METRIC TEST FOR CONDITIONAL INDEPENDENCE

Liangjun Su
Singapore Management University

Halbert White
University of California, San Diego

We propose a nonparametric test of conditional independence based on the weighted Hellinger distance between the two conditional densities, $f(y|x,z)$ and $f(y|x)$, which is identically zero under the null. We use the functional delta method to expand the test statistic around the population value and establish asymptotic normality under β-mixing conditions. We show that the test is consistent and has power against alternatives at distance $n^{-1/2}h^{-d/4}$. The cases for which not all random variables of interest are continuously valued or observable are also discussed. Monte Carlo simulation results indicate that the test behaves reasonably well in finite samples and significantly outperforms some earlier tests for a variety of data generating processes. We apply our procedure to test for Granger noncausality in exchange rates.

1. INTRODUCTION

We investigate a nonparametric test of the conditional independence of Y and Z given X, i.e.,

$$Y \perp Z \,|\, X. \tag{1.1}$$

This is related to the more familiar hypothesis that Y is independent of Z, but neither implies the other in general (see Phillips, 1988). Moreover, this hypothesis is important in both econometrics and statistics, in that many important concepts can be formalized using conditional independence (see Dawid, 1979).

Our first motivation is testing Granger noncausality. As Florens and Mouchart (1982) and Florens and Fougere (1996) show, Granger noncausality is a form of conditional independence. The hypothesis of distributional Granger

The authors gratefully thank Graham Elliott, Clive W.J. Granger, Peter C.B. Phillips, a co-editor, and three anonymous referees for valuable comments on an earlier version of this paper. The first author gratefully acknowledges financial support from the NSFC (Project 70501001). Address correspondence to Halbert White, Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA; e-mail: hwhite@econ.ucsd.edu.

 0266-4666/08 $15.00

(1980) noncausality for two stationary ergodic time series $\{Y_t\}$ and $\{Z_t\}$ is as follows. Given lags p and q, $\{Z_t\}$ does not Granger cause $\{Y_t\}$ if

$$Y_t \perp (Z_{t-1}, \ldots, Z_{t-q}) \mid (Y_{t-1}, \ldots, Y_{t-p}). \tag{1.2}$$

To test Granger noncausality, early studies often specified linear vector autoregressive (VAR) models. A serious problem with a linear approach is that such tests have low power in detecting nonlinear alternatives. Bell, Kay, and Malley (1996) propose a procedure using nonparametric additive models but do not provide distribution theory. In contrast, Baek and Brock (1992) use the correlation integral to detect nonlinear alternatives in independent and identically distributed (i.i.d.) data. Hiemstra and Jones (1994) modify Baek and Brock's approach to allow weak stochastic dependence.

Our second motivation concerns specifying the semiparametric binary choice model

$$Y = 1\{G(X,\beta) \geq \varepsilon\}, \tag{1.3}$$

with $1\{\cdot\}$ the indicator function, G a function known up to a parameter β (e.g., $G(X,\beta) = X'\beta$), and ε an unobservable error. The literature divides according to whether ε is assumed independent of X or only median independent. The latter condition, imposed by Manski (1975) and Horowitz (1992), accommodates conditional heteroskedasticity of unknown form but precludes estimating β at the usual $\sqrt{n}$ rate. In contrast, if ε is independent of X, one can estimate at the $\sqrt{n}$ rate; see Klein and Spady (1993). As independence between X and ε implies conditional independence of observables,

$$Y \perp X \mid G(X,\beta), \tag{1.4}$$

it suffices to assume the weaker condition (1.4) when specifying model (1.3). This permits the dispersion of ε to depend on X and still permits an $\sqrt{n}$-consistent estimator. This approach holds generally for transformation models, including binary choice, duration, and censored regression models. It also extends to panel models. (See Linton and Gozalo, 1997.)

The next example concerns sample selection. A huge literature has developed from the work of Heckman (1974) and Gronau (1974), who consider the following selection problem: each population member has a triple (X, Y, Z), with vectors X and Y and $Z = 1$ or 0 (e.g., Y is log offered wage, X is worker attributes, and $Z = 1$ if the worker has a job and $Z = 0$ otherwise). A researcher always observes (X, Z) but observes Y only when $Z = 1$. The researcher is interested in

$$P(Y|X) = P(Y|X, Z=1)P(Z=1|X) + P(Y|X, Z=0)P(Z=0|X).$$

The sample is uninformative about $P(Y|X, Z = 0)$, and so early researchers often assumed

$$Y \perp Z \mid X. \tag{1.5}$$

Given that $P(Y \mid X, Z=1)$ is identified, (1.5) identifies $P(Y|X)$. Since the 1970s, economists have used latent-variable models of the form

$$\begin{cases} Y = g_1(X) + \varepsilon_1, \\ Z = 1\{g_2(X) + \varepsilon_2 > 0\}, \end{cases}$$

with g_1 and g_2 real-valued functions and ε_1 and ε_2 unobserved errors. The early literature assumes $\varepsilon_1 \perp \varepsilon_2 | X$, implying (1.5) and the absence of selection bias. For more, see Angrist (1997).

In each of these three examples it is of interest to test whether the conditional independence hypothesis is true. This brings us to our contribution. There are many nonparametric tests of independence for continuous random variables, starting with Hoeffding (1948), including empirical distribution–based methods such as Blum, Kiefer, and Rosenblatt (1961) and Skaug and Tjøstheim (1993), smoothing-based methods such as Rosenblatt (1975), Robinson (1991), and Hong and White (2005), and others, e.g., Brock, Dechert, Scheinkman, and LeBaron (1996). Nevertheless, practical nonparametric tests for conditional independence are not as well developed.[1] Using empirical process theory, Linton and Gozalo (1997) give a nonparametric test of conditional independence using a generalized empirical distribution, and Delgado and González-Manteiga (2001) give an omnibus test of conditional independence using the weighted difference of the estimated conditional distributions under the null and the alternative. Nevertheless, both tests are for the i.i.d. case, and neither is asymptotically pivotal. In contrast, we build on the large literature on kernel-based omnibus testing of restrictions on nonparametric curves, initiated by Bickel and Rosenblatt (1973) and Rosenblatt (1975). We give a test for conditional independence based on a weighted version of Hellinger distance under weak data dependence. A main advantage is that our statistic is asymptotically pivotal. Despite its inability to detect local alternatives at rate $n^{-1/2}$ like the tests of Linton and Gozalo and Delgada and González-Manteiga, it turns out to be more efficient in the direction of certain high-frequency alternatives such as those of Rosenblatt (1975) and Horowitz and Spokoiny (2001).

Among other things, our test applies to test for Granger noncausality with no need to specify a linear or nonlinear model. Also, it applies to cases where not all variables are continuous or observable.

The paper is organized as follows. In Section 2, we give the basic framework, assuming no parameter estimation and that all random variables are continuous. Section 3 studies the asymptotic null distribution of our statistic and global and local power properties. Section 4 treats discrete variables, parameter estimation, and bootstrap approximation. We report a Monte Carlo study and an application in Section 5 and conclude in Section 6. We relegate technical details to Appendixes A–D.

2. BASIC FRAMEWORK

We wish to know if Y and Z are independent given X, where X, Y, and Z are d_1-, d_2-, and d_3-vectors, respectively. We have n identically distributed, weakly dependent observations (X_t, Y_t, Z_t), $t = 1, \ldots, n$.

The joint density (resp. cumulative distribution function) of (X_t, Y_t, Z_t) is f (resp. F). We reference marginal densities of $f(x, y, z)$ simply using the list of their arguments—e.g., $f(x, y) = \int f(x, y, z)\, dz$, $f(x, z) = \int f(x, y, z)\, dy$, and $f(x) = \int f(x, y, z)\; dy\; dz$, where $\int$ integrates on the full range of its arguments. This notation is compact and, we hope, sufficiently unambiguous.

Let $f(\cdot|\cdot)$ be the conditional density of one random vector given another. Formally, the null is

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

equivalent to $f(x, y, z)\, f(x) = f(x, y) f(x, z)$, for all (x, y, z) in the support of f. The alternative is

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

Our test statistic is based on the weighted Hellinger distance between $f(x, y, z) f(x)$ and $f(x, y) f(x, z)$:

$$\Gamma(f, F) \equiv \int \left\{ 1 - \sqrt{\frac{f(x, y) f(x, z)}{f(x, y, z) f(x)}} \right\}^2 a(x, y, z)\; dF(x, y, z), \tag{2.3}$$

with $a(\cdot)$ a specified nonnegative weighting function with compact support $A \subset \mathbb{R}^d$, $d \equiv d_1 + d_2 + d_3$.

The weighting function is crucial. It truncates integration at the extremes, where precise estimation of densities is quite hard. Thus, we only detect deviations between $f(x, y, z) f(x)$ and $f(x, y) f(x, z)$ on A. One can also assume compact support for (X, Y, Z) and use Hellinger distance ($a \equiv 1$).

Other statistics can be constructed using entropy (e.g., Robinson, 1991; Fernandes, 2000; Hong and White, 2005) or using the L^2 distance between $f(x, y, z) f(x)$ and $f(x, y) f(x, z)$. It is well known that entropy- or Hellinger-based statistics have better small-sample performance than L^2-based statistics when testing serial independence. Theoretically, Hellinger distance has some advantages over distances based on the L^q norm, e.g., $q = 1, 2$, or ∞. Let f_1 and f_2 be densities. Then: (1) The L^1 or L^2 norm of $f_1 - f_2$ equally weights identical differences between f_1 and f_2 regardless of whether the smaller of the two is large or small, whereas L^∞ only weighs the extreme distance between f_1 and f_2. (2) Like L^∞, L^1 is analytically awkward. (3) None of the L^q norms, $q = 1$, 2, or

∞, are invariant to continuous monotonic transformation. In contrast, like Shannon entropy, Hellinger distance does not have these problems. In particular, it is invariant to continuous monotonic transformation, which is important in applications. We use Hellinger distance instead of entropy as only the former yields a second-order theory à la White and Hong (1999) in the presence of the weighting function a. See Pitman (1979, Ch. 2) for more on distances between probability measures.

To define our test statistic, we first introduce kernel estimators for the unknown densities. For a kernel function[2] K and bandwidth $h \equiv h(n)$, we define

$$K_h(u) \equiv h^{-d} K(u/h), \tag{2.4}$$

where u has dimension d. We use the standard Nadaraya–Watson (NW) density estimator,

$$\hat{f}(x,y,z) \equiv \frac{1}{n} \sum_{t=1}^{n} K_h(x - X_t, y - Y_t, z - Z_t); \tag{2.5}$$

estimators for $f(x,y), f(x,z)$, and $f(x)$ are analogous. Let $\hat{F}$ be the empirical cumulative distribution function (c.d.f.) of (X,Y,Z). Our test statistic is a sample analogue of (2.3),

$$\hat{\Gamma} \equiv \Gamma(\hat{f}, \hat{F}) \equiv \int_A \left\{ 1 - \sqrt{\frac{\hat{f}(x,y)\hat{f}(x,z)}{\hat{f}(x,y,z)\hat{f}(x)}} \right\}^2 a(x,y,z)\, d\hat{F}(x,y,z)$$

$$= \frac{1}{n} \sum_{t=1}^{n} \left\{ 1 - \sqrt{\frac{\hat{f}(X_t,Y_t)\hat{f}(X_t,Z_t)}{\hat{f}(X_t,Y_t,Z_t)\hat{f}(X_t)}} \right\}^2 a(X_t,Y_t,Z_t).$$

We show that the properties of $\hat{\Gamma}$ follow from the properties of Γ. Two observations are important: (1) the first-order terms in the expansion of $\Gamma(\hat{f}, F)$ around $\Gamma(f, F)$ degenerate under the null;[3] and (2) the distance between $\Gamma(\hat{f}, \hat{F})$ and $\Gamma(\hat{f}, F)$ is asymptotically negligible. The latter is important as it is easier to study the asymptotic behavior of $\Gamma(\hat{f}, F)$. The former is important as it implies that the usual $\sqrt{n}$-asymptotics (e.g., Robinson, 1991) do not apply; different normalizations must be used (e.g., White and Hong, 1999; Hong and White, 2005).

3. THE ASYMPTOTIC DISTRIBUTION OF THE TEST STATISTIC

We now treat testing conditional independence for a continuously distributed stochastic process.

3.1. Asymptotic Null Distribution

Our assumptions are as follows. See Appendix A for definitions and other technical material.

Assumption A.1 (Stochastic process).

(a) $\{W_t \equiv (X_t', Y_t', Z_t')' \in \mathbb{R}^{d_1+d_2+d_3} \equiv \mathbb{R}^d,\ t \geq 0\}$ is a strictly stationary β-mixing process with coefficients $\beta_m = O(\rho^m)$ for some $0 < \rho < 1$.
(b) $W_t \equiv (X_t', Y_t', Z_t')'$ has joint distribution F and joint density f such that f has continuous partial derivatives of order $r \geq 4$, bounded and integrable on $\mathbb{R}^d$. The joint density f is bounded away from zero on the compact support A of $a(\cdot)$, i.e., $\inf_{w\in A} f(w) \equiv b > 0$, and 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 euclidean norm.
(c) The joint probability density function (p.d.f.) $f_{t_1,\ldots,t_l}(\cdot,\ldots,\cdot)$ of $(W_0, W_{t_1}, \ldots, W_{t_l})$ $(1 \leq l \leq 5)$ is bounded and satisfies a Lipschitz condition: $|f_{t_1,\ldots,t_l}(w_0+u_0,\ldots,w_l+u_l) - f_{t_1,\ldots,t_l}(w_0,\ldots,w_l)| \leq D_{t_1,\ldots,t_l}(w_0,\ldots,w_l)\|u\|$, where $u \equiv (u_0,\ldots,u_l)$ and $D_{t_1,\ldots,t_l}$ is integrable and satisfies $\int D_{t_1,\ldots,t_l}(w_0,\ldots,w_l)\|w\|^{2\xi}\,dw < \overline{M} < \infty$ and $\int D_{t_1,\ldots,t_l}(w_0,\ldots,w_l) f_{t_1,\ldots,t_l}(w_0,\ldots,w_l)\,dw < \overline{M} < \infty$ for some $\xi > 1$.

Assumption A.2 (Kernel). For some even integer $r \geq 4$, the kernel K is a product kernel of the bounded symmetric kernel $k:\mathbb{R} \to \mathbb{R}$ satisfying $\int_{\mathbb{R}} u^i k(u)\,du = \delta_{i0}$ $(i = 0,1,\ldots,r-1)$, $C_0 \equiv \int_{\mathbb{R}} u^r k(u)\,du < \infty$, $\int_{\mathbb{R}} u^2 k(u)^2\,du < \infty$, and $k(u) = O((1+|u|^{r+1+\delta})^{-1})$ for some $\delta > 0$, where δ_{ij} is Kronecker's delta.

Assumption A.3 (Bandwidth). As $n \to 0$, the bandwidth sequence $h \to 0$, such that

(a) $nh^{2d}/(\ln n)^{\gamma} \to \infty$ for some $\gamma > 0$;
(b) $nh^{d/2+2r} \to 0$.

Remark 1. Assumption A.1(a) is standard for application of a central limit theorem (CLT) for U-statistics for weakly dependent data (e.g., Fan and Li, 1999a). It is satisfied by many well-known processes such as linear stationary autoregressive moving average (ARMA) processes and a large class of processes implied by numerous nonlinear models, including bilinear, nonlinear autoregressive (NLAR), and autoregressive conditional heteroskedastic (ARCH) type models (see Fan and Li, 1999b). Assumptions A.1(b) and (c) are primarily smoothness conditions like those imposed by Li (1999). Assumption A.2 requires a higher order kernel, which is common in the literature (see Robinson, 1988; Fan and Li, 1996; and Li, 1999). Assumption A.3 restricts the bandwidth sequence. Although we allow different bandwidths for different kernel density estimators, we in fact use the same bandwidth h. This makes certain bias terms cancel each other under the null. For more on bandwidth choice, see Chen, Linton, and Robinson (2001). Assumption A.3(a) is explicitly used in the proof of Lemma B.7 in Appendix B. It is stronger than the common assumption $nh^d/(\ln n)^\gamma \to 0$ for some $\gamma > 0$, which suffices for Lemmas B.2–B.6. We conjecture that one can use the weaker assumption at the expense of highly technical argument to show asymptotic negligibility of the remainder in Lemma B.7. If so, as a referee comments, one can use a second-order positive kernel ($r = 2$) for the important case $d = 3$.[4]

To state the result and give the derivation, let[5] $w = (x, y, z)$ and define the following notation:

$$B_1 \equiv (C_1)^d \int_A a(w)\, dw,$$

$$B_2 \equiv (C_1)^{d-1} C_2 \sum_{i=1}^{d} \int_A \frac{1}{2} (\partial^2 f(w)/\partial w_i^2) a(w)/f(w)\, dw,$$

$$B_3 \equiv (C_1)^{d_1+d_2} \int_A a(w) f(w)/f(x,y)\, dw,$$

$$B_4 \equiv (C_1)^{d_1+d_3} \int_A a(w) f(w)/f(x,z)\, dw,$$

$$B_5 \equiv (C_1)^{d_1} \int_A a(w) f(w)/f(x)\, dw,$$

$$\sigma^2 \equiv (C_3)^d \int_A a(w)^2\, dw,$$

where $C_1 \equiv \int_{\mathbb{R}} k(u)^2\, du$, $C_2 \equiv \int_{\mathbb{R}} u^2 k(u)^2\, du$, and $C_3 \equiv \int_{\mathbb{R}} (\int_{\mathbb{R}} k(u+v)k(u)\, du)^2\, dv$. For a kernel satisfying Assumption A.2, the C_i's can be calculated explicitly; e.g., when $k(u) = (3 - u^2)\varphi(u)/2$ with $\varphi(u)$ the standard normal p.d.f., we have $C_1 = 27/(32\sqrt{\pi})$, $C_2 = 15/(64\sqrt{\pi})$, and $C_3 = 7{,}881/(8{,}192\sqrt{2\pi})$. We can now state our first result.

THEOREM 3.1. *Under Assumptions A.1–A.3 and under H_0, if $d \le 7$ and $d_1 - 4 < d_3 - d_2 < 4 - d_1$, then*

$$nh^{d/2}\{4\hat{\Gamma} - n^{-1}h^{-d}B_1 - n^{-1}h^{-d+2}B_2 + n^{-1}h^{-(d_1+d_2)}B_3 + n^{-1}h^{-(d_1+d_3)}B_4 - n^{-1}h^{-d_1}B_5\} \xrightarrow{d} N(0, 2\sigma^2).$$

The proof relies on a functional expansion of $\Gamma(\cdot, F)$, as in Aït-Sahalia, Bickel, and Stoker (2001), and some preliminary U-statistic results in Tenreiro (1997). In studying goodness-of-fit tests for kernel regression, Aït-Sahalia et al. derive the functional expansion for the sum of squared departures between restricted and unrestricted regressions. Similarly, we take a second-order expansion, as the first-order term vanishes under the null.

Not all the bias correction terms, B_i, $i = 1, \ldots, 5$, may be necessary. For example, if $d = 3$ (implying $d_1 = d_2 = d_3 = 1$), both B_2 and B_5 are asymptotically negligible. If $d_2 + d_3 > d_1$, B_5 is not needed. If $d_3 > d_1 + d_2$ (resp. $d_2 > d_1 + d_3$) then B_3 (resp. B_4) is unnecessary. If $d \le 5$, as the "curse of dimensionality" requires for realistic applications, the restriction $d_1 - 4 < d_3 - d_2 < 4 - d_1$ is redundant.

To implement, we consistently estimate the last four bias terms at certain rates as

$$\hat{B}_2 \equiv \frac{(C_1)^{d-1} C_2}{n} \sum_{t=1}^{n} \sum_{i=1}^{d} \frac{1}{2} \{ \hat{f}_i^{(2)}(W_t) a(W_t) / \hat{f}^{(0)}(W_t)^2 \},$$

$$\hat{B}_3 \equiv \frac{(C_1)^{d_1+d_2}}{n} \sum_{t=1}^{n} \{ a(W_t) / \hat{f}(X_t, Y_t) \},$$

$$\hat{B}_4 \equiv \frac{(C_1)^{d_1+d_3}}{n} \sum_{t=1}^{n} \{ a(W_t) / \hat{f}(X_t, Z_t) \},$$

$$\hat{B}_5 \equiv \frac{(C_1)^{d_1}}{n} \sum_{t=1}^{n} \{ a(W_t) / \hat{f}(X_t) \},$$

where, e.g., $\hat{f}_i^{(2)}(w) \equiv n^{-1} h_1^{-(d+2)} \sum_{t=1}^{n} k_{(2)}((w_i - W_{t,i})/h_1) \Pi_{j \ne i}^{d} k_{(0)}((w_j - W_{t,j})/h_1)$, $\hat{f}^{(0)}(w) \equiv n^{-1} h_1^{-d} \sum_{t=1}^{n} \Pi_{j=1}^{d} k_{(0)}((w_j - W_{t,j})/h_1)$, $k_{(v)}$ is the kernel of order (v, p) for estimating the vth partial derivative of a univariate density, h_1 is a bandwidth sequence, and $W_{t,i}$ is the ith element of W_t, $i = 1, 2, \ldots, d$. Following Gasser, Müller, and Mammitzsch (1985), we assume $0 \le v \le p - 2$, where $v = 0$ or 2 and p is even. The choice of $k_{(v)}$ $(v = 0, 2)$ is crucial to estimate the second-order partial derivatives effectively. For brevity, we refer the reader to Gasser et al. (1985) and Singh (1987).[6] It is not hard to show that $h^{(d_3-d_1-d_2)/2}(\hat{B}_3 - B_3)$, $h^{(d_2-d_1-d_3)/2}(\hat{B}_4 - B_4)$, and $h^{(d_2+d_3-d_1)/2}(\hat{B}_5 - B_5)$ are $o_p(1)$ by Assumptions A.1–A.3. We show in Appendix D that, for $i = 1, \ldots, d$,

$$h^{2-d/2} \left\{ \frac{1}{n} \sum_{t=1}^{n} \frac{\hat{f}_i^{(2)}(W_t) a(W_t)}{\hat{f}^{(0)}(W_t)} - \int_A \frac{\partial^2 f(w)}{\partial w_i^2} \frac{a(w)}{f(w)} dw \right\} = o_p(1), \tag{3.1}$$

given $h^{2-d/2} h_1^{-2} \upsilon_n = o(1)$, with $\upsilon_n \equiv n^{-1/2} h_1^{-d/2} (\ln n)^{\gamma} + h_1^{p}$ for $\gamma > 0$, and so $h^{2-d/2}(\hat{B}_2 - B_2) = o_p(1)$.

Then the estimation errors for the bias terms are asymptotically negligible, and we can compare

$$T_n \equiv nh^{d/2}\{4\hat{\Gamma} - n^{-1}h^{-d}B_1 - n^{-1}h^{-d+2}\hat{B}_2 + n^{-1}h^{-(d_1+d_2)}\hat{B}_3 + n^{-1}h^{-(d_1+d_3)}\hat{B}_4 - n^{-1}h^{-d_1}\hat{B}_5\}/\sqrt{2\sigma^2} \tag{3.2}$$

to the critical value z_α from the $N(0,1)$ distribution, i.e., $z_{0.05} = 1.645$ and $z_{0.10} = 1.282$, as the test is one-sided, and we reject the null when $T_n > z_\alpha$.

3.2. Consistency and Local Power Properties

We now study the consistency and local power properties of our test. Our consistency result is as follows.

PROPOSITION 3.2. *Suppose that $d \leq 7$, $d_1 - 4 < d_3 - d_2 < 4 - d_1$, and $h^{2-d/2}h_1^{-2}\upsilon_n = o(1)$. Under Assumptions A.1–A.3, the test based on the statistic (3.2) is consistent for F such that $\Gamma(f,F) \geq \epsilon > 0$.*

Note that the preceding proposition is equivalent to saying that the test is consistent when $a(x,y,z)\{1 - \sqrt{f(x,y)f(x,z)/[f(x,y,z)f(x)]}\} \neq 0$ in a region of positive density mass. In theory, we should require the support A of $a(\cdot)$ to be as large as possible. In practice, we often have that $A = A_1 \times A_2 \times A_3 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$, $A_1 = \{x \in \mathbb{R}^{d_1} : x \in [\bar{X} - 2\,\hat{S}_X, \bar{X} + 2\,\hat{S}_X]\}$, with $\bar{X}$ and $\hat{S}_X$ the sample average and standard deviation of X, respectively; and A_2 and A_3 are defined analogously.[7] Note that the support A chosen in this way is dependent on n, but this has no asymptotic impact on the distribution of our statistic.

To define local alternatives we follow the notation of Gouriéroux and Tenreiro (2001) and consider a sequence of d-dimensional strictly stationary processes $(W_{nt}, t \geq 0)$.

Assumption A.1*. (a) $\{W_{nt} \equiv (X_{nt}', Y_{nt}', Z_{nt}')' \in \mathbb{R}^{d_1+d_2+d_3} \equiv \mathbb{R}^d,\ t = 1,\ldots,n;\ n = 1,2,\ldots\}$ is a strictly stationary β-mixing process with coefficients β_m^n satisfying

$$\beta_m \equiv \sup_{n\in\mathbb{N}} \beta_m^n = O(\rho^m) \quad \text{for some } 0 < \rho < 1.$$

Let $f^{[n]}(x,y,z)$ be the joint density of X_{nt}, Y_{nt}, and Z_{nt}. Let $\alpha_n \to 0$ as $n \to \infty$. We first examine the power of our test against the sequence of local alternatives

$$H_1(\alpha_n) : f^{[n]}(y|x,z) = f^{[n]}(y|x)[1 + \alpha_n \Delta(w) + o(\alpha_n)\Delta_n(w)], \tag{3.3}$$

where $f^{[n]}(y|x,z)$ and $f^{[n]}(y|x)$ are conditional densities derived from $f^{[n]}(x,y,z)$ and $\Delta(w)$ and $\Delta_n(w)$ are specified in Assumption A.4.

Assumption A.4 (Local alternatives).

(a) $1 + \alpha_n \Delta(w) + o(\alpha_n)\Delta_n(w) \geq 0$ for all $w \in \mathbb{R}^d$ and all $n \in \mathbb{N}$.
(b) $\int_{\mathbb{R}^d} \Delta(w) f^{[n]}(x,y) f^{[n]}(z|x)\, dw = 0$ and $\int_{\mathbb{R}^d} \Delta_n(w) f^{[n]}(x,y) f^{[n]}(z|x)\, dw = 0$ for all $n \in \mathbb{N}$.
(c) $\int_A |\Delta(w)|^2 f^{[n]}(w) a(w)\, dw < \overline{M}$ and $\int_A |\Delta_n(w)|^2 f^{[n]}(w) a(w)\, dw < \overline{M}$ for some $\overline{M} < \infty$ for all $n \in \mathbb{N}$.
(d) $\lim_{n\to\infty} f^{[n]}(\cdot)$ exists and $f(w) = \lim_{n\to\infty} f^{[n]}(w)$.

Assumptions A.4(a) and (b) ensure that $f^{[n]}(x,y,z)$ is a valid p.d.f. for all $n \in \mathbb{N}$. Assumption A.4(c) ensures that the remainder term $o(\alpha_n)\Delta_n(w)$ has no impact on the asymptotic distribution of the statistic T_n and $\alpha_n \Delta(w)$ is at distance $O(\alpha_n)$ from the null. Also, we modify Assumptions A.1(b) and (c) as follows.

Assumption A.1*. (b) and (c) Assumptions A.1(b) and (c) hold with $f^{[n]}$ and $F^{[n]}$ replacing f and F, respectively.

PROPOSITION 3.3. *Suppose that $d \leq 7$, $d_1 - 4 < d_3 - d_2 < 4 - d_1$, and $h^{2-d/2} h_1^{-2} \upsilon_n = o(1)$ and that $\alpha_n = n^{-1/2} h^{-d/4}$ in $H_1(\alpha_n)$. Then under Assumptions A.1* and A.2–A.4,* $\Pr(T_n \geq z_\alpha | H_1(\alpha_n)) \to 1 - \Phi(z_\alpha - \delta/(\sqrt{2}\sigma))$, *where $\delta \equiv \int_A a(w)\Delta(w)^2 f(w)\, dw$.*

Remark 2. Proposition 3.3 indicates that our test statistic T_n has nontrivial power against $H_1(\alpha_n)$ with $\alpha_n = n^{-1/2}h^{-d/4}$ whenever $\delta \neq 0$. The rate $n^{-1/2}h^{-d/4}$ is slower than $n^{-1/2}$, as $h \to 0$. In contrast, the Linton and Gozalo (1997) and Delgado and González-Manteiga (2001) tests have nontrivial power in the direction of alternatives converging to the null at rate $n^{-1/2}$. Thus, the latter tests would be more powerful than ours against local alternatives such as (3.3).

Next, consider the following high-frequency alternatives of the type considered by Rosenblatt (1975) and, more recently, by Horowitz and Spokoiny (2001):

$$H_{1,h}(\lambda_n, \gamma_n): f^{[n]}(y|x,z)$$
$$= f^{[n]}(y|x)[1 + \lambda_n \Lambda((w - w_0)/\gamma_n) + o(\lambda_n)\Lambda_n((w - w_0)/\gamma_n)], \tag{3.4}$$

where $w_0 \in A \subset \mathbb{R}^{d_1+d_2+d_3}$ with $a(w_0) > 0$, λ_n and $\gamma_n \to 0$ as $n \to \infty$.

Assumption A.4* (Local alternatives).

(a) $1 + \lambda_n \Lambda((w - w_0)/\gamma_n) + o(\lambda_n)\Lambda_n((w - w_0)/\gamma_n) \geq 0$ for all $w \in \mathbb{R}^d$ and all $n \in \mathbb{N}$.
(b) $\int_{\mathbb{R}^d} \Lambda((w - w_0)/\gamma_n) f^{[n]}(x,y) f^{[n]}(z|x)\, dw = 0$ and $\int_{\mathbb{R}^d} \Lambda_n((w - w_0)/\gamma_n) f^{[n]}(x,y) f^{[n]}(z|x)\, dw = 0$ for all $n \in \mathbb{N}$.
(c) $\int_A |\Lambda(w)|^2\, dw < \overline{M}$ and $\int_A |\Lambda_n(w)|^2\, dw < \overline{M}$ for some $\overline{M} < \infty$ for all $n \in \mathbb{N}$.
(d) $f^{[n]}(\cdot)$ is bounded $\mathbb{R}^d$, $\lim_{n\to\infty} f^{[n]}(\cdot)$ exists, and $f(w) = \lim_{n\to\infty} f^{[n]}(w)$.

PROPOSITION 3.4. *Suppose that* $d \leq 7$, $d_1 - 4 < d_3 - d_2 < 4 - d_1$, *and* $h^{2-d/2} h_1^{-2} \upsilon_n = o(1)$. *Suppose* $H_{1,h}(\lambda_n, \gamma_n)$ *hold with* $nh_n^{d/2} \lambda_n^2 \gamma_n \to C \in (0,\infty]$. *Then under Assumptions A.1*, A.2, A.3, and A.4*,* $\Pr(T_n \geq z_\alpha | H_{1,h}(\lambda_n, \gamma_n)) \to 1 - \Phi(z_\alpha - \bar{\delta}/(\sqrt{2}\sigma))$, *where* $\bar{\delta} \equiv Ca(w_0) f(w_0) \int \Lambda(w)^2 \, dw$.

Remark 3. Proposition 3.4 indicates that our test statistic T_n has nontrivial power against $H_{1,h}(\lambda_n, \gamma_n)$ for certain sequences of λ_n and γ_n. For example, if we choose $\lambda_n = (nh^{d/2})^{-1/3}$ and $\gamma_n = (nh^{d/2})^{-1/3} (\ln \ln n)^\gamma$ for some $\gamma \geq 0$, one can easily see that the condition on λ_n and γ_n in the preceding proposition is met. Noticing that $\lambda_n \gamma_n = o(n^{-1/2})$, it is known that in this case the powers of the Linton and Gozalo (1997) and Delgado and González-Manteiga (2001) tests converge to zero as $n \to \infty$. Therefore, our test is more powerful than the latter tests for certain high-frequency alternatives of the form (3.4).

4. EXTENSIONS AND DISCUSSION

In the preceding discussion we treat a stochastic process that has continuously valued realizations. Although this case suffices for many empirical applications (e.g., nonparametric testing of Granger noncausality), our testing procedure is applicable to a much wider range of situations. We now discuss two cases that generalize the preceding basic results. Also, we propose a bootstrap approximation to the distribution of our statistic.

4.1. Discrete Random Variable

Our test can be modified to incorporate the case in which one of the random variables in (X, Y, Z) is discretely valued. For notational convenience, we explicitly assume that Z is a binary variable.[8]

Let $f_1(x,y) \equiv f(x,y) P(Z = 1 | x, y)$ be the joint density of (X, Y, Z) with respect to the product of Lebesgue measure on $\mathbb{R}^{d_1 + d_2}$ and counting measure. Similarly, one defines $f_1(x) \equiv f(x) P(Z = 1 | x)$, $f_0(x) \equiv f(x) P(Z = 0 | x)$, and $f_0(x,y) \equiv f(x,y) P(Z = 0 | x, y)$. The test is based on the functional

$$\Gamma_1(f, F) \equiv \int \left\{ 1 - \sqrt{\frac{f(x,y) f_1(x)}{f_1(x,y) f(x)}} \right\}^2 a(x,y) \, dF_1(x,y)$$
$$+ \int \left\{ 1 - \sqrt{\frac{f(x,y) f_0(x)}{f_0(x,y) f(x)}} \right\}^2 a(x,y) \, dF_0(x,y), \qquad \textbf{(4.1)}$$

where $a(x,y)$ is a nonnegative weighting function that can be understood as our previous $a(x,y,z)$ restricted to $\mathbb{R}^{d_1+d_2}$, $dF_1(x,y) \equiv f(x,y) P(Z = 1 | x, y) \, dx \, dy$, and $dF_0(x,y) \equiv f(x,y) P(Z = 0 | x, y) \, dx \, dy$. Clearly, under the null that $Y \perp Z | X$, $\Gamma(f, F) = 0$. It is easy to show that under suitable conditions, a normalized version of the sample analogue of $\Gamma_1(f, F)$ is asymptotically normally distributed, and the dimension d_3 does not affect the convergence rate. For brevity, we do not report the theoretical result here; it is available in the working paper version of this paper at http://www.econ.ucsd.edu/~lsu/.

4.2. Conditional Independence Testing with Estimated Variables

Now consider the case in which $W = (X', Y', Z')'$ is not observed directly but can be estimated. Asymptotic results for this case are useful when a conditional independence test is conducted using residuals or other estimated random variables. Let $\{M_t \in \mathbb{R}^k, t \geq 0\}$ be the observed process. Of interest are certain functions calculated from M, i.e., $W(M,\theta) \equiv (X(M,\theta)', Y(M,\theta)', Z(M,\theta)')' \in \mathbb{R}^{d_1+d_2+d_3} \equiv \mathbb{R}^d$, where the parameter $\theta \in \Theta \subset \mathbb{R}^p$. The null is, for some unknown $\theta_0 \in \Theta$,

$$H_0 : Y(M,\theta_0) \perp Z(M,\theta_0) \mid X(M,\theta_0). \tag{4.2}$$

Denote the p.d.f.s of $W(M,\theta)$ and its subvectors 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 $W(M,\theta)$. Under the null, we have

$$\Gamma_2(f,F;\theta_0) \equiv \int \left\{1 - \sqrt{\frac{f(x,y;\theta_0)f(x,z;\theta_0)}{f(w;\theta_0)f(x;\theta_0)}}\right\}^2 a(w;\theta_0)\, dF(w;\theta_0) = 0, \tag{4.3}$$

where $a(w;\theta) \equiv a(w(\theta))$ is a nonnegative weighting function that depends on θ only through w and is otherwise the same as $a(w)$ used in Section 3. We suppose that there exist estimates $\hat{\theta}$ of θ_0 that are $\sqrt{n}$-consistent under the null. To implement the test, we replace $\Gamma_2(f,F;\theta_0)$ by its sample analogue

$$\Gamma_2(\hat{f},\hat{F};\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\left\{1 - \sqrt{\frac{\hat{f}(X_t(\hat{\theta}),Y_t(\hat{\theta}))\hat{f}(X_t(\hat{\theta}),Z_t(\hat{\theta}))}{\hat{f}(W_t(\hat{\theta}))\hat{f}(X_t(\hat{\theta}))}}\right\}^2 a(W_t(\hat{\theta})),$$

where, e.g., $W_t(\hat{\theta}) \equiv W(M_t,\hat{\theta})$ and $\hat{f}(w;\theta)$ is the standard NW density estimator of $f(w;\theta)$ that uses "observations" $\{W_t(\theta), 1 \leq t \leq n\}$. Under mild regularity conditions, we can show by applying results of Andrews (1995) that estimation of $\hat{\theta}$ does not affect the asymptotics, as

$$\Gamma_2(\hat{f},\hat{F};\hat{\theta}) = \Gamma_2(\hat{f},\hat{F};\theta_0) + o_p(n^{-1}h^{-d/2}). \tag{4.4}$$

4.3. Smoothed Local Bootstrap

Generally speaking, the basic problems for the bootstrap are how to impose the null in the resampling scheme and accommodate the dependence structure in the data. We stress the fact that the theorems obtained in this paper are based on asymptotic considerations. As Neumann and Paparoditis (2000) noted, to get an asymptotically correct estimator of the null distribution of T_n, it is not necessary to reproduce the whole dependence structure of the stochastic processes generating the original observations. Simple resampling from the empirical distribution of $W_t = (X_t', Y_t', Z_t')'$ will not impose the null restriction. Paparoditis and Politis (2000) propose a local bootstrap procedure for nonparametric kernel estimators under general dependence conditions. We essentially do the same thing here, except that our conditioning variables are not necessarily lagged dependent variables. Let $\mathcal{W} \equiv \{W_t\}_{t=1}^n$. We draw bootstrap resamples $\{X_t^*, Y_t^*, Z_t^*\}_{t=1}^n$ based on the following smoothed local bootstrap procedure: (1) Draw a bootstrap sample $\mathcal{X}^* \equiv \{X_t^*\}_{t=1}^n$ from the smoothed kernel density $\tilde{f}(x) = n^{-1}\sum_{t=1}^n L_b(X_t - x)$, where $L_b(x) = b^{-d_1}L(x/b)$ with $L(\cdot)$ a product kernel of a univariate density l, and $b > 0$ the resampling bandwidth. (2) For $t = 1,\ldots,n$, given X_t^*, draw Y_t^* and Z_t^* independently from the smoothed conditional density $\tilde{f}(y|X_t^*) = \sum_{s=1}^n L_b(Y_s - y)L_b(X_s - X_t^*)/\sum_{r=1}^n L_b(X_r - X_t^*)$ and $\tilde{f}(z|X_t^*) = \sum_{s=1}^n L_b(Z_s - z)L_b(X_s - X_t^*)/\sum_{r=1}^n L_b(X_r - X_t^*)$, respectively, and denote $W_t^* \equiv (X_t^{*\prime}, Y_t^{*\prime}, Z_t^{*\prime})'$ and $\mathcal{W}^{q*} \equiv \{W_t^*\}_{t=1}^n$. (3) Compute a bootstrap statistic T_n^* in the same way as T_n, with $\mathcal{W}^*$ replacing $\mathcal{W}$. (4) Repeat steps (1) and (2) B times to obtain B bootstrap test statistics $\{T_{nj}^*\}_{j=1}^B$. Paparoditis and Politis (Rmk. 2.1) explain how to generate the bootstrap replicates computationally.

Let $\Pr^*$ denote probability conditional on the sample $\mathcal{W}$. The level α critical values $\tilde{c}_\alpha$ are computed as an approximate solution to $\Pr^*[T_n^* > \tilde{c}_\alpha] = \alpha$. The bootstrap p-value is then given by $p^* \equiv B^{-1}\sum_{j=1}^B 1(T_{nj}^* > T_n)$. Several facts are worth mentioning: (1) Conditionally on $\mathcal{W}$, the bootstrap replicates W_t^* and W_s^* are independent for $t \neq s$, and they have the same distributions; (2) conditionally on $\mathcal{W}$, Y_t^* and Z_t^* are independent given X_t^*. We shall use these facts repeatedly in the proof of Theorem 4.1 in Appendix C.

To show that the smoothed local bootstrap procedure works, we impose the following conditions on $L(\cdot)$ and b.

Assumption A.5 (Bootstrap kernel and bandwidth).

(a) The kernel L is a product kernel of a bounded symmetric kernel density $l: \mathbb{R} \to \mathbb{R}^+$ such that $\int_{\mathbb{R}} u^i l(u)\,du = \delta_{i0}$ $(i = 0,1)$.
(b) l is r times continuously differentiable such that $\int_{\mathbb{R}} u^j l^{(r)}(u)\,du = 0$ for $j = 0,1,\ldots, r-1$ and $\int_{\mathbb{R}} u^r l^{(r)}(u)\,du < \infty$.
(c) As $n \to \infty$, $b \to 0$, and $nb^{d+2r}/(\ln n)^\gamma \to C \in (0,\infty]$ for some $\gamma > 0$.

Assumption A.5(a) is standard. We impose Assumptions A.5(b) and (c) to ensure that the smoothed kernel densities $\tilde{f}$'s are well behaved, e.g., the rth derivatives of $\tilde{f}(x)$ are bounded uniformly on a compact set with probability approaching 1 as $n \to \infty$. When $r = 4$, $l = \varphi$, the standard normal density, satisfies A.5(b).

THEOREM 4.1. *Suppose Assumptions A.1–A.3 and A.5 hold; if* $d \leq 7$ *and* $d_1 - 4 < d_3 - d_2 < 4 - d_1$, *then*

(i) $T_n^* \xrightarrow{d} N(0,1)$ *conditionally on* $\mathcal{W}$;
(ii) $P(T_n > T_n^*) \to 1$ *provided that* $\Gamma(f,F) \geq \varepsilon > 0$.

Theorem 4.1(i) shows that the smoothed local bootstrap provides an asymptotic valid approximation to the null limit distribution of T_n (i.e., $N(0,1)$). This holds as long as we generate the bootstrap data by imposing the null hypothesis. Theorem 4.1(ii) implies that the test T_n based upon the bootstrap critical value is consistent against every global alternative for which $f(y|x,z) = f(y|x)$ does not hold almost everywhere. That is, $T_n \to \infty$ with probability approaching 1 under H_1. We will compare the finite-sample performance of the smoothed local bootstrap with that of the asymptotic normal approximation in our simulation.

5. NUMERICAL RESULTS

5.1. Monte Carlo Simulations

We now present Monte Carlo experiment results that illustrate the finite-sample performance of our test. First, we consider the following data generating processes (DGPs):

DGP1s: $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})'$, where $\{\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}\}$ are i.i.d. $N(0, I_3)$.

For DGPs 2s–4s and DGPs 1p–5p, which follow, $W_t = (Y_{t-1}, Y_t, Z_{t-1})'$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, $\{\varepsilon_{1,t},\ \varepsilon_{2,t}\}$ are i.i.d. $N(0, I_2)$, and

DGP2s: $Y_t = 0.5Y_{t-1} + \varepsilon_{1,t}$;
DGP3s: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2$;
DGP4s: $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}$, $Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}$, $h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2$, $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$;
DGP1p: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1,t}$;
DGP2p: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}^2 + \varepsilon_{1,t}$;
DGP3p: $Y_t = 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1,t}$;
DGP4p: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}\varepsilon_{1,t}$;
DGP5p: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$.
DGP6p: $W_t = (Y_{t-1}, Y_t, Z_{t-1})'$, where $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}$, $Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}$, $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2$, $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$, and $\{\varepsilon_{1,t},\ \varepsilon_{2,t}\}$ are i.i.d. $N(0, I_2)$.

DGPs 1s–4s allow us to examine the level of the test, whereas DGPs 1p–6p are used to study power properties. These DGPs cover a variety of linear and nonlinear stochastic processes commonly studied in time series analysis. In particular, we have Granger causality in the mean (resp. variance) in DGPs 1p–3p (resp. DGPs 4p–6p). DGPs 3s and 4s and 5p and 6p specify processes of (G)ARCH type.

We use a fourth-order kernel in estimating all required densities: $k(u) = (3 - u^2)\varphi(u)/2$. The weighting function $a(w)$ is given in note 7. Thus, $\int_{\mathbb{R}^3} a(w)\,dw = 1$ and $\int_{\mathbb{R}^3} a(w)^2\,dw = \frac{1}{27}$. As it is difficult to specify the optimal bandwidth sequence, we take $h = cn^{-1/8.5}$ for a variety of c's.

To implement our test, we rescale the data so that each variable has sample mean zero and variance 1. For each of DGPs 1s–4s, we choose $c = 1$ in calculating T_n and make a comparison between the asymptotic normal and bootstrap approximations to the distribution of T_n, with $n = 100$. For the bootstrap approximation, we choose $B = 1{,}000$, $b = n^{-1/5}$, and l the standard normal p.d.f. In Figure 1, the solid line (Hel) denotes the sample distribution of T_n obtained over 2,000 simulations. The dashed line (Normal) denotes the normal approximation and the dotted line (Hel_b) the bootstrap approximation. For each of DGPs 1s–4s, the bootstrap approximation is better than the normal approximation in the right tail. As Härdle and Mammen (1993) remark, the inaccuracy of the normal approximation increases with the dimension of (X, Y, Z), and so we recommend the use of the bootstrap in applications.

Linton and Gozalo (1997) base their tests of conditional independence on the functional $A_n(w) = \{n^{-1}\sum_{t=1}^n 1(W_t \le w)\} \times \{n^{-1}\sum_{t=1}^n 1(X_t \le x)\} - \{n^{-1}\sum_{t=1}^n 1(X_t \le x)\ 1(Y_t \le y)\}\{n^{-1}\sum_{t=1}^n 1(X_t \le x)1(Z_t \le z)\}$, where $w = (x, y, z)$. Specifically, their test statistics are of the Cramér–von Mises and Kolmogorov–Smirnov types: $CM_n = \sum_{t=1}^n A_n^2(W_t)$, $KS_n = \sqrt{n}\max_{1\le t\le n} |A_n(W_t)|$. Delgado and González-Manteiga (2001) base their tests of conditional independence on the functional $L_n(w) = n^{-1}\sum_{t=1}^n \{1(Y_t \le y) - \tilde{F}_n(y|X_t)\}\tilde{f}(X_t)1(X_t \le x)1(Z_t \le z)$, where for bandwidth h_2 and kernel K_2, $\tilde{F}_n(y|X_t) \equiv n^{-1}h_2^{-d_1}\sum_{s=1}^n 1(Y_s \le y)K_2((X_t - X_s)/h_2)/\tilde{f}(X_t)$ and $\tilde{f}(X_t) \equiv n^{-1}h_2^{-d_1}\sum_{s=1}^n K_2((X_t - X_s)/h_2)$. We denote their two test statistics as $\text{SCM}_n = \sum_{t=1}^n L_n^2(W_t)$ and $\text{SKS}_n = \sqrt{n}\max_{1\le t\le n}|L_n(W_t)|$. We choose K_2 to be the standard normal p.d.f. and let $h_2 = n^{-1/3}$ in our simulation. Note that both the Linton and Gozalo and Delgado and González-Manteiga tests were developed for i.i.d. data. To implement their tests here, we replace their bootstrap procedures by the preceding local bootstrap to account for data dependence. To compare the performance of these tests with ours, we implement our test with $c = 1$, 1.5, and 2. To save computation time, we use $B = 200$ and 250 repetitions unless otherwise stated.

Tables 1 and 2 report the estimated levels and powers for the 5% and 10% tests. Also reported in the tables are the standard linear Granger causality results (LIN_n) with 1,000 repetitions, where we examine whether Z_{t-1} should enter the regression of Y_t on Y_{t-1} linearly. From Table 1, we see that the levels of all tests behave reasonably well despite the fact that the both the Delgado and González-Manteiga (2001) test and our test (for small values of c) tend to be oversized for small sample sizes. From Table 2, we see that except for DGP1p, where the linear Granger causal relation is true, the standard linear Granger causality test performs worse than all other tests in all cases and SKS_n in some cases. It is not surprising that the CM_n and SCM_n tests beat the KS_n and SKS_n

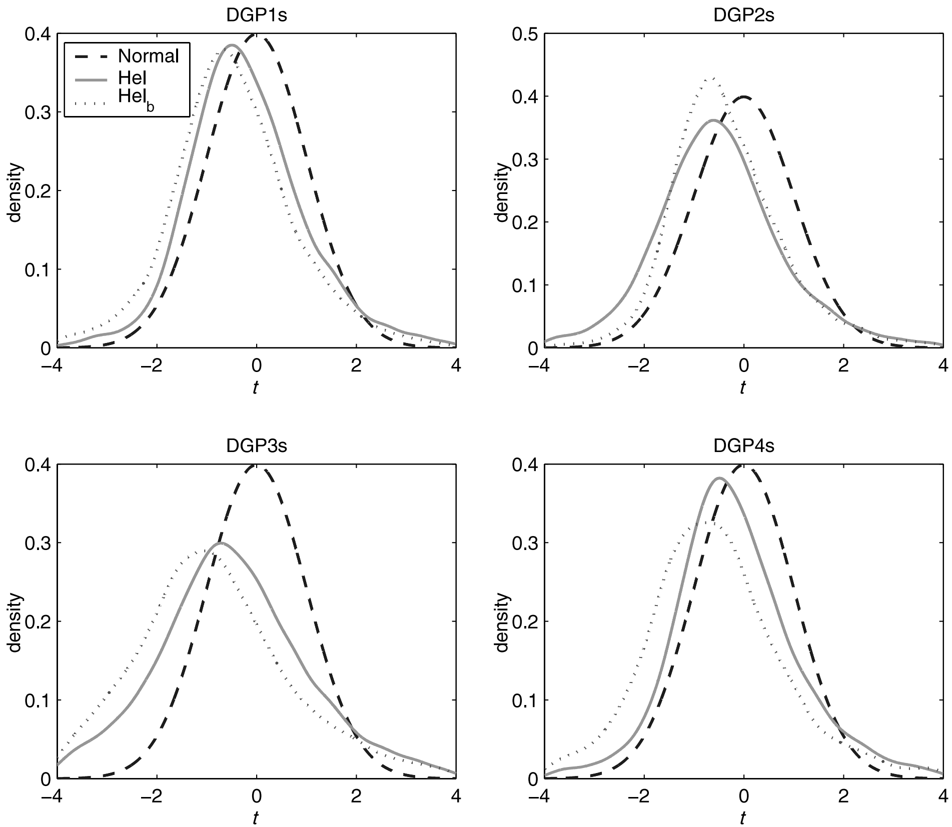

Figure 1. Comparison of a symptotic and bootstrap approximations to the distribution of T_n.

Table 1. Level comparison of the tests

	DGP1s	DGP2s	DGP3s	DGP4s	DGP1s	DGP2s	DGP3s	DGP4s
	$n = 100$, 5%				$n = 100$, 10%			
LIN_n	0.044	0.061	0.050	0.060	0.095	0.121	0.110	0.106
CM_n	0.054	0.058	0.060	0.048	0.094	0.100	0.132	0.120
KS_n	0.042	0.056	0.056	0.040	0.100	0.112	0.140	0.108
SCM_n	0.076	0.060	0.084	0.064	0.108	0.080	0.156	0.116
SKS_n	0.064	0.056	0.088	0.068	0.128	0.108	0.124	0.148
T_n, $c = 1$	0.096	0.060	0.048	0.072	0.176	0.152	0.120	0.148
T_n, $c = 1.5$	0.068	0.056	0.052	0.056	0.124	0.120	0.120	0.124
T_n, $c = 2$	0.072	0.036	0.072	0.048	0.136	0.072	0.120	0.084
	$n = 200$, 5%				$n = 200$, 10%			
LIN_n	0.043	0.053	0.042	0.050	0.840	0.109	0.101	0.090
CM_n	0.044	0.056	0.060	0.048	0.095	0.100	0.108	0.112
KS_n	0.068	0.053	0.048	0.084	0.096	0.088	0.104	0.124
SCM_n	0.048	0.060	0.064	0.068	0.092	0.100	0.124	0.140
SKS_n	0.056	0.028	0.064	0.072	0.092	0.084	0.112	0.132
T_n, $c = 1$	0.064	0.052	0.080	0.080	0.100	0.140	0.136	0.140
T_n, $c = 1.5$	0.064	0.056	0.048	0.036	0.120	0.128	0.120	0.092
T_n, $c = 2$	0.044	0.060	0.056	0.048	0.092	0.144	0.084	0.096

tests, respectively, as this has been seen in several other studies. Also, the CM_n test tends to complement the SCM_n test whereas the KS_n test tends to dominate the SKS_n test in power. As far as our test is concerned, the CM_n and SCM_n tests are more powerful than our test in detecting linear Granger causality in the mean for small values of c whereas for other cases, our test outperforms.

Next, we consider high-frequency alternatives of the form

$$Y_t = 0.5X_t + 4\tau\varphi(Z_t/\tau) + 0.5\varepsilon_t,$$

where $\{X_t, Z_t, \varepsilon_t\}$ are i.i.d. $N(0, I_3)$ and as before φ is the standard normal p.d.f. We consider $\tau \in \{0, 0.5, 1, 2\}$, where $Y_t = 0.5X_t + 0.5\varepsilon_t$ for $\tau = 0$, and we denote the corresponding DGPs as DGP1h–DGP4h. In this case, $W_t = (X_t, Y_t, Z_t)'$. We also check whether Z_t should enter the regression of Y_t on X_t linearly and denote the resulting t-test statistic as LIN_n.

Table 3 reports the rejection frequency for various tests. For $\tau = 0$ (DGP1h), the null hypothesis is true, and all tests tend to be undersized for small n. When $\tau \neq 0$, the powers of the Linton and Gozalo (1997) and Delgado and González-Manteiga (2001) tests are significantly lower than the power of our test, as expected. Also, for some values of τ, the Linton and Gozalo and Delgado and González-Manteiga tests are beaten even by the simple test LIN_n.

Table 2. Power comparison of the tests

	DGP1p	DGP2p	DGP3p	DGP4p	DGP5p	DGP6p
$n = 100$, 5%						
LIN_n	0.999	0.337	0.213	0.126	0.163	0.153
CM_n	0.920	0.548	0.504	0.412	0.384	0.188
KS_n	0.780	0.404	0.380	0.288	0.292	0.156
SCM_n	0.924	0.464	0.352	0.500	0.224	0.196
SKS_n	0.728	0.236	0.288	0.340	0.120	0.112
T_n, $c = 1$	0.668	0.756	0.388	0.860	0.828	0.680
T_n, $c = 1.5$	0.888	0.940	0.512	0.924	0.952	0.812
T_n, $c = 2$	0.952	0.944	0.576	0.940	0.988	0.912
$n = 200$, 5%						
LIN_n	1.000	0.354	0.250	0.113	0.172	0.143
CM_n	0.992	0.748	0.788	0.680	0.476	0.360
KS_n	0.952	0.552	0.660	0.532	0.336	0.284
SCM_n	0.980	0.648	0.620	0.720	0.352	0.280
SKS_n	0.964	0.324	0.512	0.552	0.148	0.136
T_n, $c = 1$	0.900	0.960	0.596	0.992	0.968	0.880
T_n, $c = 1.5$	0.980	1.000	0.808	0.992	0.972	0.972
T_n, $c = 2$	1.000	1.000	0.864	1.000	1.000	0.996
$n = 100$, 10%						
LIN_n	1.000	0.436	0.284	0.175	0.239	0.233
CM_n	0.964	0.652	0.644	0.480	0.472	0.304
KS_n	0.868	0.492	0.496	0.428	0.408	0.232
SCM_n	0.960	0.564	0.488	0.612	0.324	0.300
SKS_n	0.876	0.372	0.400	0.436	0.176	0.212
T_n, $c = 1$	0.772	0.840	0.532	0.932	0.912	0.776
T_n, $c = 1.5$	0.948	0.972	0.692	0.964	0.972	0.896
T_n, $c = 2$	0.976	0.988	0.712	0.964	0.992	0.928
$n = 200$, 10%						
LIN_n	1.000	0.442	0.327	0.176	0.253	0.209
CM_n	1.000	0.856	0.904	0.752	0.592	0.508
KS_n	0.988	0.676	0.756	0.676	0.484	0.404
SCM_n	0.988	0.732	0.728	0.812	0.480	0.424
SKS_n	0.984	0.468	0.604	0.664	0.276	0.232
T_n, $c = 1$	0.944	0.984	0.712	0.996	0.976	0.936
T_n, $c = 1.5$	0.984	1.000	0.896	0.992	0.984	0.996
T_n, $c = 2$	1.000	1.000	0.936	1.000	1.000	0.996

5.2. Application to Exchange Rate Data

Over the last two decades many studies have reported that foreign exchange rates exhibit nonlinear dependence, but researchers often neglect this when test-

TABLE 3. Comparison of tests for high-frequency alternatives

	DGP1h	DGP2h	DGP3h	DGP4h	DGP1h	DGP2h	DGP3h	DGP4h
	$n = 100$, 5%				$n = 100$, 10%			
LIN_n	0.045	0.055	0.133	0.190	0.100	0.115	0.187	0.267
CM_n	0.020	0.160	0.280	0.128	0.064	0.276	0.428	0.248
KS_n	0.024	0.128	0.176	0.112	0.072	0.256	0.316	0.172
SCM_n	0.012	0.088	0.180	0.080	0.036	0.168	0.288	0.120
SKS_n	0.036	0.156	0.196	0.116	0.072	0.220	0.292	0.152
T_n, $c = 1$	0.028	0.696	0.948	0.764	0.072	0.808	0.968	0.876
T_n, $c = 1.5$	0.044	0.828	0.980	0.892	0.068	0.916	0.984	0.948
$\mathrm{T}_n, c = 2$	0.020	0.596	0.976	0.936	0.044	0.708	0.992	0.968
	$n = 200$, 5%				$n = 200$, 10%			
LIN_n	0.047	0.059	0.124	0.202	0.094	0.115	0.207	0.286
CM_n	0.068	0.444	0.708	0.332	0.100	0.580	0.816	0.536
KS_n	0.056	0.284	0.524	0.220	0.104	0.448	0.680	0.336
SCM_n	0.024	0.196	0.356	0.104	0.036	0.272	0.488	0.196
SKS_n	0.044	0.204	0.356	0.140	0.096	0.336	0.476	0.236
$\mathrm{T}_n, c = 1$	0.040	0.980	1.000	0.964	0.072	0.988	1.000	0.988
$\mathrm{T}_n, c = 1.5$	0.028	0.988	0.992	0.996	0.064	0.996	0.992	0.996
$\mathrm{T}_n, c = 2$	0.020	0.972	1.000	0.996	0.080	0.988	1.000	0.996

ing Granger causality. One exception is Hong (2001), who proposes a test for volatility spillover and applies it to study the volatility spillover between two weekly nominal U.S. dollar exchange rates, Deutschemark (DM) and Japanese yen (YEN).

In this application, we apply our nonparametric test to examine the causal relationship between DM and YEN and that between DM and the British pound (PD), and we compare this to some previous tests. The data are obtained from Datastream for the sample period from 19 January 1994 to 19 January 2004, with 2,609 observations total. The exchange rates are the local currency against the U.S. dollar. As is standard, we let DM, YEN, and PD stand for the natural logarithm of the preceding three exchange rates multiplied by 100. The augmented Dickey–Fuller test indicates that there is a unit root in all three level series but not in the first-differenced series, ΔDM, ΔYEN, and ΔPD. Johansen's likelihood test indicates that DM is not cointegrated with YEN or PD. Therefore, both the linear and nonlinear Granger causality tests will be conducted on the first differenced data.

For conciseness, we only consider the dynamic interaction between exchange rates at the one-day lag. For example, for testing whether YEN Granger-causes DM linearly, we check whether $\beta = 0$ in $\Delta\mathrm{DM}_t = \alpha_0 + \alpha\Delta\mathrm{DM}_{t-1} + \beta\Delta\mathrm{YEN}_{t-1} + \varepsilon_t$; for testing whether YEN Granger-causes DM nonlinearly, we check $H_{0,NL}$: $\Delta\mathrm{DM}_t \perp \Delta\mathrm{YEN}_{t-1} | \Delta\mathrm{DM}_{t-1}$.

The results are summarized in Table 4. The linear Granger causality test (LIN) does not reveal a Granger causal relationship between DM and YEN or PD at a

Table 4. Applications to Deutschemark (DM), Japanese yen (YEN), and British pound (PD)

	DM and YEN		DM and PD	
Tests/H_0	ΔYEN $\not\Rightarrow$ ΔDM	ΔDM $\not\Rightarrow$ ΔYEN	ΔPD $\not\Rightarrow$ ΔDM	ΔDM $\not\Rightarrow$ ΔPD
LIN_n	0.219	0.431	0.997	0.234
CM_n	0.915	0.790	0.900	0.320
KS_n	0.915	0.625	0.785	0.365
SCM_n	0.710	0.865	0.925	0.340
SKS_n	0.620	0.905	0.770	0.635
T_n, $c = 1$	0.185	0.020	0.005	0.105
T_n, $c = 1.5$	0.385	0.025	0.020	0.110
T_n, $c = 2$	0.450	0.020	0.065	0.200

Note: The notation $\not\Rightarrow$ means "does not Granger cause." The central entries are the p-values for each test. Bandwidth sequences and kernels are chosen as in the simulations.

one-day lag, similar to the Linton and Gozalo (1997) and Delgado and González-Manteiga (2001) tests. In contrast, our nonparametric test reveals unidirectional Granger causality from DM to YEN and from PD to DM. This suggests that at a one-day lag the exchange rates across countries interact strongly with each other. One obvious reason for the failure of the linear Granger causality test and the Linton and Gozalo and Delgado and González-Manteiga tests in detecting such causal linkages is that exchange rates exhibit unambiguously nonlinear dependence across markets. The volatility spillover between exchange rates is a special case of such nonlinear dependence.

6. CONCLUSION

This paper develops asymptotic distribution theory for a nonparametric test of conditional independence under weak dependence conditions. The test is directly applicable to testing Granger non-causality. It also applies to cases where not all variables are continuous or observable. Monte Carlo experiments indicate that the our test outperforms the Linton and Gozalo (1997) and Delgado and González-Manteiga (2001) tests significantly in a variety of DGPs. An application to exchange rate data demonstrates the power of our test in detecting nonlinear Granger causal relationships.

To improve the asymptotic approximation to the finite-sample distribution of our test statistic, one can consider higher order refinements. If the distributions of our test statistic and its bootstrap analogue admit Edgeworth expansion, 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 of $O(h^{d_2} + h^{d_3})$. Recently, Nishiyama and Robinson (2000) and Linton (2002) have established the validity of Edgeworth expansion for a degenerate U-statistic with variable kernel. This suggests that a rigorous proof 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; we leave this for future research.

Other interesting directions for future research are to accommodate nonstationary processes and to extend our test to the case in which some of the random variables are nonparametrically estimated.

NOTES

1. For categorical data there are numerous tests of independence and conditional independence; see Rosenbaum (1984) and Yao and Tritchler (1993), among others.

2. We adopt the same notational convention for the kernel K as for the density f, namely, to indicate the kernel by the list of its arguments or by specifying the dimension of its arguments.

3. Fernandes and Flores (2000) employ a generalized entropy measure that includes Hellinger distance as a special case to test conditional independence. The first-order terms of their functional expansion are also degenerate, and so they use a weight function to avoid the degeneracy, which unfortunately results in poor small-sample performance.

4. In simulations we find that for some DGPs, there is about a 0.1% chance that $\hat{f}(x,y)\hat{f}(x,z)/[\hat{f}(x,y,z)\hat{f}(x)]$ takes negative values when (x,y,z) lies two standard deviations from the sample mean of (X,Y,Z). To avoid negative density estimates, we recommend replacing $\hat{f}(\cdot)$ by $\max(\hat{f}(\cdot), 0.1/n)\ 1\{\hat{f}(\cdot) \le 0\} + \hat{f}(\cdot)1\{\hat{f}(\cdot) > 0\}$; this change does not affect the asymptotic theory.

5. For the vector argument in a function, we find it convenient to assume that every vector is a row vector to avoid proliferation of transposes.

6. We thank a referee who kindly brought to our attention these two references.

7. Alternatively, one can use the Bartlett kernel function (or other density-form function) as the weighting function a. For example, if the ith element of W, W_i, has mean zero and standard deviation one (perhaps after being recentered and rescaled), for $i = 1,\dots,d$, one can use $a(w) = \Pi_{i=1}^{d}[(\frac{1}{2} + \frac{1}{4}w_i)1\{-2 \le w_i \le 0\} + (\frac{1}{2} - \frac{1}{4}w_i)1\{0 < w_i \le 2\}]$.

8. If Y is a binary variable, one can exchange the role of Y and Z because $Y \perp Z \mid X$ if and only if $Z \perp Y \mid X$. The case for which both Y and Z are discretely valued is treated in Rosenbaum (1984).

REFERENCES

Aït-Sahalia, Y., P.J. Bickel, & T.M. Stoker (2001) Goodness-of-fit for kernel regression with an application to option implied volatilities. *Journal of Econometrics* 105, 363–412.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Angrist, J.D. (1997) Conditional independence in sample selection models. *Economics Letters* 54, 103–112.

Baek, E. & W. Brock (1992) A General Test for Nonlinear Granger Causality: Bivariate Model. Working paper, Iowa State University and University of Wisconsin, Madison.

Bell, D., J. Kay, & J. Malley (1996) A non-parametric approach to non-linear causality testing. *Economics Letters* 51, 7–18.

Bickel, P.J. & M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071–1095.

Blum, J.R., J. Kiefer, & M. Rosenblatt (1961) Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics* 32, 485–498.

Brock, W., W. Dechert, J. Scheinkman, & B. LeBaron (1996) A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235.

Chen, X., O. Linton, & P.M. Robinson (2001) The Estimation of Conditional Densities. STICERD Econometrics Discussion paper, London School of Economics and Political Science.

Dawid, A.D. (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* 41, 1–31.

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

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

Fan, Y. & 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.

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

Fernandes, M. (2000) Nonparametric Entropy-Based Tests of Independence between Stochastic Processes. Discussion paper, European University Institute.

Fernandes, M. & R.G. Flores (2000) Tests for Conditional Independence, Markovian Dynamics, and Noncausality. Discussion paper, European University Institute.

Florens, J.P. & D. Fougere (1996) Noncausality in continuous time. *Econometrica* 64, 1195–1212.

Florens, J.P. & M. Mouchart (1982) A note on causality. *Econometrica* 50, 583–591.

Gasser, T., H.-G. Müller, & V. Mammitzsch (1985) Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society Series B* 47, 238–252.

Gouriéroux, C. & C. Tenreiro (2001) Local power properties of kernel based goodness of fit tests. *Journal of Multivariate Analysis* 78, 161–190.

Granger, C.W.J. (1980) Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* 2, 329–352.

Gronau, R. (1974) Wage comparisons: A selectivity bias. *Journal of Political Economy* 82, 1119–1144.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Heckman, J.J. (1974) Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.

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

Hoeffding, W. (1948) A nonparametric test of independence. *Annals of Mathematical Statistics* 58, 546–557.

Hong, Y. (2001) A test for volatility spillover with application to exchange rates. *Journal of Econometrics* 103, 183–224.

Hong, Y. & H. White (2005) Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73, 837–901.

Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.

Klein, R.W. & R.H. Spady (1993) An efficient semiparametric estimator for discrete choice models. *Econometrica* 61, 387–421.

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

Liebscher, E. (1996) Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Process and Their Applications* 65, 69–80.

Linton, O. (2002) Edgeworth approximations for semiparametric instrumental variable estimators and test statistics. *Journal of Econometrics* 106, 325–368.

Linton, O. & P. Gozalo (1997) Conditional Independence Restrictions: Testing and Estimation. Discussion paper, Cowles Foundation for Research in Economics, Yale University.

Manski, C.F. (1975) The maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228.

Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571–599.

Neumann, M.H. & E. Paparoditis (2000) On bootstrapping L_2-type statistics in density testing. *Statistics and Probability Letters* 50, 137–147.

Newey, W.K. (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10, 233–253.

Nishiyama, Y. & P.M. Robinson (2000) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68, 931–980.

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

Phillips, P.C.B. (1988) Conditional and unconditional statistical independence. *Journal of Econometrics* 38, 341–348.

Pitman, E.J.G. (1979) *Some Basic Theory for Statistical Inference*. Wiley.

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

Robinson, P.M. (1991) Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437–453.

Rosenbaum, P.R. (1984) Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika* 49, 425–435.

Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3, 1–14.

Singh, R.S. (1987) MISE of kernel estimates of a density and its derivatives. *Statistics and Probability Letters* 5, 153–159.

Skaug, H.J. & D. Tjøstheim (1993) A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* 80, 591–602.

Tenreiro, C. (1997) Loi asymptotique des erreurs quadratiques intégrées des estimateurs a noyau de la densité et de la régression sous des conditions de dependance. *Portugaliae Mathematica* 54, 197–213.

White, H. & Y. Hong (1999) M-testing using finite and infinite dimensional parameter estimators. In R. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, pp. 326–345. Oxford University Press.

Yao, Q. & D. Tritchler (1993) An exact analysis of conditional independence in several 2×2 contingency tables. *Biometrics* 49, 233–236.

Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichskeitstheorie und Verwandte Gebiete* 35, 237–252.

Yoshihara, K. (1989) Limiting behavior of generalized quadratic forms generated by absolutely regular processes. In P. Mandl & M. Hušková (eds.), *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, pp. 539–547. Charles University Press.

APPENDIX A: Some Useful Definitions, Lemmas, and Theorems

Here we provide a definition, two lemmas, and one theorem that are used in the proofs of the main theorems and propositions in the text.

DEFINITION A.1. *Let $\{U_t,\, t \in \mathbb{Z}\}$ be a d-dimensional strictly stationary stochastic process and let $\mathcal{F}_s^t$ denote the σ-algebra generated by $(U_s, \ldots, U_t)$ for $s \le t$. The process is called β-mixing or absolutely regular, if as $m \to \infty$,*

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

For a sequence of d-dimensional strictly stationary processes $(U_{nt},\, t \in \mathbb{Z})$, denote by β_m^n the β-mixing coefficient of process $(W_{nt},\, t \in \mathbb{Z})$:

$$\beta_m^n = \mathrm{E}\left[\sup_{A \epsilon \mathcal{F}_{n,m}^{\infty}} \{|P(A|\mathcal{F}_{n,-\infty}^{0}) - P(A)|\} \right],$$

where $\mathcal{F}_{n,m}^{\infty}$ (resp. $\mathcal{F}_{n,-\infty}^{0}$) is the σ-algebra generated by U_{nt}, $t \ge m$ (resp. U_{nt}, $t \le 0$).

LEMMA A.2 (Yoshihara, 1976). *Let $\{U_t,\, t \ge 0\}$ be a d-dimensional stochastic process satisfying Assumption A.1(a) in the text. Let $h(v_1, \ldots, v_k)$ be a Borel measurable function on $\mathbb{R}^{kd}$ such that for some $\delta > 0$ and given j, $M \equiv \max\{\int_{\mathbb{R}^{kd}} |h(v_1, \ldots, v_k)|^{1+\delta} dF(v_1, \ldots, v_k), \int\int_{\mathbb{R}^{kd}} |h(v_1, \ldots, v_k)|^{1+\delta} dF^{(1)}(v_1, \ldots, v_j)\, dF^{(2)}(v_{j+1}, \ldots, v_k)\}$ exists. Then $|\int_{\mathbb{R}^{kd}} h(v_1, \ldots, v_k)\, dF(v_1, \ldots, v_k) - \int\int_{\mathbb{R}^{kd}} h(v_1, \ldots, v_k)\, dF^{(1)}(v_1, \ldots, v_j)\, dF^{(2)}(v_{j+1}, \ldots, v_k)| \le 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)}$, where $m \equiv i_{j+1} - i_j$, F, $F^{(1)}$, and $F^{(2)}$ are distributions of random vectors $(U_{i_1}, \ldots, U_{i_k})$, $V_1 \equiv (U_{i_1}, \ldots, U_{i_j})$, and $V_2 \equiv (U_{i_{j+1}}, \ldots, U_{i_k})$, respectively; and $i_1 < i_2 < \cdots < i_k$.*

LEMMA A.3 (Yoshihara, 1989). *Let h be defined as before; then* $\mathrm{E}|\mathrm{E}[h(V_1, V_2)|V_1] - \mathrm{E}_{V_1} h(V_1, V_2)| \le 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)}$, *where* $\mathrm{E}_{V_1} h(V_1, V_2) \equiv H(V_1)$ *with* $H(v_1) \equiv \mathrm{E}[h(v_1, V_2)]$.

Now let $g_n(\cdot)$ and $h_n(\cdot,\cdot)$ be Borel measurable functions on $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$, respectively. Suppose $\mathrm{E}[g_n(U_0)] = 0$, $\mathrm{E}[h_n(U_{0,}v)] = 0$, and $h_n(u,v) = h_n(v,u)$ for all $(u,v) \in \mathbb{R}^d \times \mathbb{R}^d$. Define $\mathcal{G}_n \equiv n^{-1/2} \sum_{i=1}^n g_n(U_i)$ and $\mathcal{H}_n \equiv n^{-1} \sum_{1 \le i < j \le n} [h_n(U_i, U_j) - \mathrm{E}h_n(U_i, U_j)]$. Clearly, $\mathcal{G}_n$ and $\mathcal{H}_n$ are degenerate U-statistics of respective orders 1 and 2. Let $p > 0$ and let $\{\bar{U}_t,\, t \ge 0\}$ be an i.i.d. sequence where $\bar{U}_0$ is an independent copy of U_0. Define

$$u_n(p) \equiv \max\left\{ \max_{1 \le i \le n} \|h_n(U_i, U_0)\|_p, \|h_n(U_0, \bar{U}_0)\|_p \right\},$$

$$v_n(p) \equiv \max\left\{ \max_{1 \le i \le n} \|G_{n0}(U_i, U_0)\|_p, \|G_{n0}(U_0, \bar{U}_0)\|_p \right\},$$

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

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

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

THEOREM A.4 (Tenreiro, 1997). *Using the preceding notation, suppose there exist $\delta_0 > 0$, $\gamma_0 < \frac{1}{2}$, and $\gamma_1 > 0$ such that*

(i) $\|g_n(U_0)\|_4 = O(1)$;
(ii) $\mathrm{E}[g_n(U_i)g_n(U_0)] = c_i + o(1)$, $i = 0, 1, 2, \ldots$;
(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) $\mathrm{E}[h_n(U_0, \bar{U}_0)]^2 = 2\tilde{\sigma}_2^2 + o(1)$.

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

APPENDIX B: Proof of Theorem 3.1

We begin by expanding the functional $\Gamma(\hat{f}, F)$ using the functional delta method. The only difference between $\Gamma(\hat{f}, F)$ and $\hat{\Gamma} \equiv \Gamma(\hat{f}, \hat{F})$ is that the latter is an average over the empirical distribution function $\hat{F}$ instead of F. We will show in Lemma B.6 that this difference is asymptotically inconsequential. To bound the remainder term in the functional expansion of $\Gamma(\hat{f}, F)$, we define the sup norm, $\|g\| \equiv \sup_{u \in A \cap \mathbb{R}^p} |g(u)|$. In what follows, the dimension p of u will be d, $d_1 + d_2$, $d_1 + d_3$, or d_1, depending on which subset of $w \equiv (x, y, z)$ we are referring to (in this Appendix all vectors are row vectors). Define $\Omega_i \equiv \{g : \mathbb{R}^{p_i} \to \mathbb{R}$, g is bounded, $\int g = 0$, and $\|g\| < b/2\}$, with $p_i = d$, $d_1 + d_2$, $d_1 + d_3$, and d_1, for $i = 1, \ldots, 4$, respectively. Throughout this Appendix, C denotes a generic constant that may vary from one place to another. The bar notation denotes an i.i.d. copy of the corresponding processes, independent of that 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.4 and B.6 for details.

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

LEMMA B.1. *Let F be a c.d.f. on $\mathbb{R}^d$. Let g_{xyz}, g_{xy}, g_{xz}, and g_x belong to Ω_i, $i =$ 1, 2, 3, and 4, respectively. Then under Assumption A.1(b) and H_0, $\Gamma(\cdot, F)$ has the following expansion:*

$$\Gamma(f + g; F) = \frac{1}{4} \int \left\{ \frac{g_{xyz}}{f(x, y, z)} - \frac{g_{xy}}{f(x, y)} - \frac{g_{xz}}{f(x, z)} + \frac{g_x}{f(x)} \right\}^2 a(w)\, dF(w) + R(g, F),$$

where $\sup\{|R(g, F)|/(\|g_{xyz}\|^3 + \|g_{xy}\|^3 + \|g_{xz}\|^3 + \|g_x\|^3) : (g_{xyz}, g_{xy}, g_{xz}, g_x) \in \Omega_1 \times \Omega_2 \times \Omega_3 \times \Omega_4\} < \infty$.

Proof. Define

$$\Psi(\tau) = \int \left\{ 1 - \sqrt{\frac{(f(x, y) + \tau g_{xy})(f(x, z) + \tau g_{xz})}{(f(x, y, z) + \tau g_{xyz})(f(x) + \tau g_x)}} \right\}^2 a(w)\, dF(w),$$

where $(g_{xyz}, g_{xy}, g_{xz}, g_x)$ are such that $(\tau g_{xyz}, \tau g_{xy}, \tau g_{xz}, \tau g_x) \in \Omega_1 \times \Omega_2 \times \Omega_3 \times \Omega_4$ for all $0 \leq \tau \leq 1$. From the explicit expression for $\Psi(\tau)$ and the properties of the f's and

g's, it follows that Ψ is three times continuously differentiable in τ on $[0,1]$. Applying Taylor's formula with Lagrange remainder to Ψ, we get

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

where $0 \leq \tau^* \leq \tau$. Note that $\Psi(0) = 0$ under H_0. Define $\varphi_1(\tau,w) \equiv [f(x,y) + \tau g_{xy}][f(x,z) + \tau g_{xz}]$, $\varphi_2(\tau,w) \equiv [f(x,y,z) + \tau g_{xyz}][f(x) + \tau g_x]$. It is immediate that

$$\Psi'(\tau) = \int \left\{1 - \sqrt{\frac{\varphi_2(\tau,w)}{\varphi_1(\tau,w)}}\right\} \times \left\{\frac{\partial\varphi_1(\tau,w)/\partial\tau}{\varphi_2(\tau,w)} - \frac{\varphi_1(\tau,w)\partial\varphi_2(\tau,w)/\partial\tau}{\varphi_2(\tau,w)^2}\right\} a(w)\, dF(w). \tag{B.1}$$

Under the null, $\Psi'(0) = 0$. That is, the first-order term vanishes in the expansion of $\Psi(\tau)$ around $\tau = 0$.

Next, we have

$$\begin{aligned}\Psi''(\tau) = {} & \frac{1}{2}\int \sqrt{\frac{\varphi_1(\tau,w)}{\varphi_2(\tau,w)}} \left\{\frac{\varphi_2(\tau,w)\partial\varphi_1(\tau,w)/\partial\tau}{\varphi_1(\tau,w)^2} - \frac{\partial\varphi_2(\tau,w)/\partial\tau}{\varphi_1(\tau,w)}\right\} \\ & \times \left\{\frac{\partial\varphi_1(\tau,w)/\partial\tau}{\varphi_2(\tau,w)} - \frac{\varphi_1(\tau,w)\partial\varphi_2(\tau,w)/\partial\tau}{\varphi_2(\tau,w)^2}\right\} a(w)\, dF(w) \\ & + \int \left\{1 - \sqrt{\frac{\varphi_2(\tau,w)}{\varphi_1(\tau,w)}}\right\} \\ & \times \left\{\frac{\partial^2\varphi_1(\tau,w)/\partial\tau^2}{\varphi_2(\tau,w)} - \frac{2\partial\varphi_1(\tau,w)/\partial\tau\partial\varphi_2(\tau,w)/\partial\tau}{\varphi_2(\tau,w)^2}\right. . \\ & \left. - \frac{\varphi_1(\tau,w)\partial^2\varphi_2(\tau,w)/\partial\tau^2}{\varphi_2(\tau,w)^2} + \frac{2\varphi_1(\tau,w)(\partial\varphi_2(\tau,w)/\partial\tau)^2}{\varphi_2(\tau,w)^3}\right\} a(w)\, dF(w).\end{aligned}$$

Note that under H_0, at $\tau = 0$, the second term in the last expression vanishes and that $\partial\varphi_1(0,w)/\partial\tau = g_{xy}f(x,z) + g_{xz}f(x,y)$, $\partial\varphi_2(0,w)/\partial\tau = g_{xyz}f(x) + g_x f(x,y,z)$, and so we can easily obtain that under H_0,

$$\Psi''(0) = \frac{1}{2}\int \left\{\frac{g_{xyz}}{f(x,y,z)} - \frac{g_{xy}}{f(x,y)} - \frac{g_{xz}}{f(x,z)} + \frac{g_x}{f(x)}\right\}^2 a(w)\, dF(w).$$

Further, notice that $\partial^2\varphi_1(\tau,w)/\partial\tau^2 = 2g_{xy}g_{xz}$ and $\partial^2\varphi_2(\tau,w)/\partial\tau^2 = 2g_{xyz}g_x$, both of which are free of τ. One can characterize the remainder term by first computing $\Psi'''(\tau)$. The explicit formula for $\Psi'''(\tau)$ is lengthy. By the Cauchy–Schwartz inequality and Assumption A.1(b), we can bound this remainder by a factor of $(\|g_{xyz}\|^3 + \|g_{xy}\|^3 + \|g_{xz}\|^3 + \|g_x\|^3)$. Consequently, for $\tau = 1$, we obtain that under H_0

$$\Psi(1) = \frac{1}{4}\int\left\{\frac{g_{xyz}}{f(x,y,z)} - \frac{g_{xy}}{f(x,y)} - \frac{g_{xz}}{f(x,z)} + \frac{g_x}{f(x)}\right\}^2 a(w)\,dF(w)$$
$$+ O(\|g_{xyz}\|^3 + \|g_{xy}\|^3 + \|g_{xz}\|^3 + \|g_x\|^3),$$

and the lemma follows. ∎

LEMMA B.2. *Under Assumptions A.1, A.2, and A.3(a), and H_0, we have for any c.d.f. F,*

$$\Gamma(\hat f, F) = \frac{1}{4}\int\left\{\frac{\hat f(x,y,z)}{f(x,y,z)} - \frac{\hat f(x,y)}{f(x,y)} - \frac{\hat f(x,z)}{f(x,z)} + \frac{\hat f(x)}{f(x)}\right\}^2 a(w)\,dF(w)$$
$$+ O_p(\|\hat f(x,y,z) - f(x,y,z)\|_\infty^3),$$

where $\|\hat f(x,y,z) - f(x,y,z)\|_\infty \equiv \sup_{(x,y,z)\in A}|\hat f(x,y,z) - f(x,y,z)|$.

Proof. We apply Lemma B.1 with $g_{xyz} = \hat f(x,y,z) - f(x,y,z)$, $g_{xy} = \hat f(x,y) - f(x,y)$, $g_{xz} = \hat f(x,z) - f(x,z)$, and $g_x = \hat f(x) - f(x)$. First note that the β-mixing condition in Assumption A.1 implies α-mixing. One can modify the proof of Theorem 4.3 in Liebscher (1996) with Assumption A.2 in place of his condition on the kernel function K and get

$$\|\hat f(w) - f(w)\|_\infty = O_p(n^{-1/2}h^{-d/2}(\ln n)^{\gamma/6} + h^r) = o_p(1) \tag{B.2}$$

for some $\gamma > 0$. Similar expressions hold for $\|\hat f(u) - f(u)\|_\infty$, with $u = (x,y)$, (x,z), or x. Let $S \equiv \{\|\hat f(w) - f(w)\|_\infty \ge b/2,\ \|\hat f(x,y) - f(x,y)\|_\infty \ge b/2,\ \|\hat f(x,z) - f(x,z)\|_\infty \ge b/2$, and $\|\hat f(x) - f(x)\|_\infty \ge b/2\}$. Then $\Pr[S] \to 0$ so that $\Pr[(g_{xyz}, g_{xy}, g_{xz}, g_x) \in \Omega_1 \times \Omega_2 \times \Omega_3 \times \Omega_4] \to 1$. Last, notice that $\|\hat f(w) - f(w)\|_\infty$ dominates $\|\hat f(u) - f(u)\|_\infty$ for $u = (x,y), (x,z)$, or x. The result follows. ∎

To facilitate the presentation, we introduce some new notation. Let

$$I_n \equiv \int\left\{\frac{\hat f(x,y,z)}{f(x,y,z)} - \frac{\hat f(x,y)}{f(x,y)} - \frac{\hat f(x,z)}{f(x,z)} + \frac{\hat f(x)}{f(x)}\right\}^2 a(w)\,dF(w) \equiv \int r_n(w)^2 a(w)\,dF(w).$$

Then $I_n = \int[r_n(w) - \mathrm{E}r_n(w)]^2 a(w)\,dF(w) + 2\int[r_n(w) - \mathrm{E}r_n(w)]\mathrm{E}r_n(w)a(w)\,dF(w) + \int[\mathrm{E}r_n(w)]^2 a(w)dF(w)$ and $I_n - \mathrm{E}[I_n] = 2\int[r_n(w) - \mathrm{E}r_n(w)]\mathrm{E}r_n(w)a(w)\,dF(w) + \int\{[r_n(w) - \mathrm{E}r_n(w)]^2 - \mathrm{E}[r_n(w) - \mathrm{E}r_n(w)]^2\}a(w)\,dF(w)$. Throughout the rest of this Appendix, we let $w \equiv (x,y,z) \in \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}^{d_3}$, $u \equiv (x',y',z') \in \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}^{d_3}$, and $v \equiv (\tilde x, \tilde y, \tilde z) \in \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\mathbb{R}^{d_3}$. Define

$$R(w,u) \equiv \frac{K_h(w-u)}{f(w)} - \frac{K_h(x-x')K_h(y-y')}{f(x,y)}$$
$$- \frac{K_h(x-x')K_h(z-z')}{f(x,z)} + \frac{K_h(x-x')}{f(x)} \equiv \sum_{i=1}^{4} R_i(w,u),$$

$\widetilde R(w,u) \equiv \sum_{i=1}^4[R_i(w,u) - \mathrm{E}R_i(w,W)] \equiv \sum_{i=1}^4 \widetilde R_i(w,u)$, $G_n(u) \equiv \int\widetilde R(w,u)h^{-r}\mathrm{E}r_n(w)a(w)\,dF(w)$, and $H_n(u,v) \equiv h^{d/2}\int\widetilde R(w,u)\widetilde R(w,v)a(w)\,dF(w)$. Note that we have suppressed the dependence of $R(\cdot,\cdot)$, $R_i(\cdot,\cdot)$, $\tilde{R}(\cdot,\cdot)$, and $\tilde{R}_i(\cdot,\cdot)$ on n. Then we can write

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

It is easy to verify that $U_{n,3} = O_p(n^{-1/2}h^{-d/2}) = o_p(1)$ under Assumptions A.1–A.3. We shall 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 restrictions on bandwidth and kernel (Lemma B.3). To get the asymptotic distribution of our test statistic, we need to calculate both asymptotic variance (Lemma B.4) and bias correction terms (Lemma B.5).

LEMMA B.3. *Let $h \to 0$. Under Assumptions A.1 and A.2 and H_0, $U_{n,1} \xrightarrow{d} N(0,\tilde{\sigma}^2)$, where $\tilde{\sigma}^2 \equiv \mathrm{Var}(\gamma(W_0)) + 2\sum_{t=1}^{\infty}\mathrm{Cov}(\gamma(W_t),\gamma(W_0))$ and $\gamma(\cdot)$ is defined in equation (B.4), which follows.*

Proof. First, $h^{-r}\mathrm{E}r_n(w) = h^{-r}\mathrm{E}[R(w,W)] = \triangle_n^r f(w)/f(w) - \triangle_n^r f(x,y)/f(x,y) - \triangle_n^r f(x,z)/f(x,z) + \triangle_n^r f(x)/f(x) \equiv \tilde{\gamma}_n(w)$, where

$$\triangle_n^r f(w) \equiv \frac{(-1)^r}{(r-1)!}\sum_{i_1,\ldots,i_r=1}^{d}\int_{\mathbb{R}^d} u_{i_1}\ldots u_{i_r}K(u)\int_0^1 \frac{\partial^r f(w-hut)}{\partial w_{i_1}\ldots\partial w_{i_r}}(1-t)^{r-1}\,dt\,du,$$

and $\triangle_n^r f(x,y)$,$\triangle_n^r f(x,z)$, and $\triangle_n^r f(x)$ are defined analogously. Because $h \to 0$, by the dominated convergence theorem and Assumption A.2, $\lim_{n\to\infty}\tilde{\gamma}_n(w) = \triangle^r f(w)/f(w) - \triangle^r f(x,y)/f(x,y) - \triangle^r f(x,z)/f(x,z) + \triangle^r f(x)/f(x) \equiv \tilde{\gamma}(w)$, where $\triangle^r f(w) \equiv ((-1)^r/r!)C_0\sum_{i=1}^{d}\partial^r f(w)/\partial w_i^r$, C_0 is defined in Assumption A.2, and $\triangle^r f(x,y)$,$\triangle^r f(x,z)$, and $\triangle^r f(x)$ are defined analogously.

Notice that $\mathrm{E}G_n(W) = 0$ by construction and $\sup_{n\in\mathbb{N}}\sup_{w\in A}|G_n(w)| < \infty$ under Assumptions A.1(b) and A.2. Now $\lim_{n\to\infty}\mathrm{E}[G_n(W_i)G_n(W_0)] = \int\tilde{\gamma}(w_i)a(w_i)\tilde{\gamma}(w_0)a(w_0)\{1 + f(y_i,z_i|x_i) - f(z_i|x_i,y_i) - f(y_i|x_i,z_i)\}\{1 + f(y_0,z_0|x_0) - f(z_0|x_0,y_0) - f(y_0|x_0,z_0)\}f_i(w_0,w_i)\,dw_i\,dw_0 - \{\int\tilde{\gamma}(w)a(w)[1 + f(y,z|x) - f(z|x,y) - f(y|x,z)]f(w)\,dw\}^2 = \mathrm{Cov}(\gamma(W_i),\gamma(W_0))$, where

$$\gamma(w) \equiv a(w)\tilde{\gamma}(w)[1 + f(y,z|x) - f(y|x,z) - f(z|x,y)]. \quad \textbf{(B.4)}$$

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

LEMMA B.4. *Under Assumptions A.1, A.2, and A.3(a) and H_0, $U_{n,2} \xrightarrow{d} N(0,\sigma^2/2)$, where σ^2 is defined before Theorem 3.1 in Section 3.1.*

Proof. Note that $U_{n,2} = n^{-1}\sum_{1\le i<j\le n}[H_n(W_i,W_j) - \mathrm{E}H_n(W_i,W_j)]$. By construction, $H_n(u,v) = H_n(v,u)$, and $\mathrm{E}H_n(W_0,v) = 0$. We verify conditions (iii)–(vii) in Theorem A.4. First, $H_n(W_i,W_0) = h^{d/2}\int \tilde{R}(w,W_i) \times \tilde{R}(w,W_0)a(w)\,dF(w) = \sum_{j=1}^4\sum_{k=1}^4 h^{d/2}\int \tilde{R}_j(w,W_i)\tilde{R}_k(w,W_0)a(w)\,dF(w)$, and so for $p \ge 1$, $\|H_n(W_i,W_0)\|_p \le \sum_{j=1}^4\sum_{k=1}^4 \|h^{d/2}\int \tilde{R}_j(w,W_i)\tilde{R}_k(w,W_0)a(w)\,dF(w)\|_p \le C\|h^{d/2}\int \tilde{R}_1(w,W_i)\tilde{R}_1(w,W_0)a(w)\,dF(w)\|_p \equiv C\|H_{n1}(W_i,W_0)\|_p$, where the first inequality is due to the triangle inequality for the L_p norm and the second follows from the fact that $\|H_{n1}(W_i,W_0)\|_p$ is the dominant term in the double summation. Notice that $H_{n1}(u,v) = h^{d/2}\int_A K_h(w-u)K_h(w-v)a(w)/f(w)\,dw + O(h^{d/2})$, and by Assumptions A.1(b) and (c)

$$
\begin{aligned}
&\mathrm{E}\left|\int_A K_h(w-W_i)K_h(w-W_0)\frac{a(w)}{f(w)}\,dw\right|^p \\
&\quad = h^{-dp}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|\int_A K(w)K\left(w+\frac{u-v}{h}\right)\frac{a(u+hw)}{f(u+hw)}\,dw\right|^p f_i(u,v)\,du\,dv \\
&\quad \le h^{-d(p-1)}\sup_{w\in A}\left(\frac{a(w)}{b}\right)^p \sup_{i\in\mathcal{N}}\sup_{u,v\in A} f_i(u,v)\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|K(u)K(u+v)|^p\,du\,dv,
\end{aligned}
$$

and so we have $\|H_n(W_i,W_0)\|_p \le Ch^{d/2}h^{-d(p-1)/p} = C(h^d)^{(1/p-1/2)}$.

Letting $\overline{W}_0$ be an independent copy of W_0, one can show by similar argument that $\|H_n(W_0,\overline{W}_0)\|_p \le C(h^d)^{(1/p-1/2)}$. Consequently, $u_n(p) \le C(h^d)^{(1/p-1/2)}$ for some $C>0$. Now we show $v_n(p) \le C(h^d)^{1/p}$. Note that $G_{n0}(u,v) \equiv \mathrm{E}[H_n(W_0,u)H_n(W_0,v)] = G_{n0,1}(u,v)(1+o(1))$, where $G_{n0,1}(u,v) = h^d\mathrm{E}\{\int\int \tilde{R}_1(w,W_0)\tilde{R}_1(w,u)\tilde{R}_1(w',W_0)\tilde{R}_1(w',v)a(w)a(w')\,dF(w)\,dF(w')\} \le C\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(w)K(w+w')K(\tilde{w})K(\tilde{w}+w'+(u-v)/h)\,dw\,dw'\,d\tilde{w} + O(h^d)$, and so $\|G_{n0,1}(W_i,W_0)\|_p \le C(h^{d/p}+h^d)$, and $\|G_{n0}(W_i,W_0)\|_p \le Ch^{d/p}$. Similarly, one can show $\|G_{n0}(W_0,\overline{W}_0)\|_p \le Ch^{d/p}$, and thus $v_n(p) \le C(h^d)^{1/p}$.

By the same argument, we have $w_n(p) \equiv \|G_{n0}(W_0,W_0)\|_p \le C$ and $z_n(p) \le Ch^d$. For some fixed $\delta_0>0$, conditions (iv) and (v) in Theorem A.4 are satisfied by Assumption A.3(a). By Assumption A.3(a), $n^{\gamma_1}h^d < \infty$ for some $\gamma_1 \in (0,1)$, and so condition (vi) in Theorem A.4 is satisfied. Now take $\gamma_0 = (2+\delta_0)/(8+2\delta_0) \in (0,\frac{1}{2})$; then condition (iii) in Theorem A.4 is satisfied again by Assumption A.3(a). Finally, $\mathrm{E}[H_n(W_0,\overline{W}_0)^2] = \int_A a(w)^2\,dw\int_{\mathbb{R}^d}[\int_{\mathbb{R}^d} K(u+v)K(u)\,du]^2\,dv + o(1) = \sigma^2 + o(1)$. It follows that $U_{n,2} \xrightarrow{d} N(0,\sigma^2/2)$. ■

LEMMA B.5. *Under Assumptions A.1–A.3 and H_0, if $d \le 7$ and $d_1 - 4 < d_3 - d_2 < 4 - d_1$, then*

$$
\begin{aligned}
nh^{d/2}\mathrm{E}I_n &= h^{-d/2}B_1 + h^{-d/2+2}B_2 - h^{(d_2-d_1-d_3)/2}B_3 - h^{(d_3-d_1-d_2)/2}B_4 \\
&\quad + h^{(d_2+d_3-d_1)/2}B_5 + o(1).
\end{aligned}
$$

Proof. $\mathrm{E}I_n = \int_A[\mathrm{E}r_n(w)]^2 a(w)\,dF(w) + \mathrm{E}\int_A[r_n(w) - \mathrm{E}r_n(w)]^2 a(w)\,dF(w) \equiv A_{n,1} + A_{n,2}$. From the proof of Lemma B.3, we obtain

$$nh^{d/2}A_{n,1} = nh^{d/2+2r}\int_A \tilde{\gamma}(w)^2 a(w)\,dF(w) + o(nh^{d/2+2r}) = o(1), \tag{B.5}$$

where the last equality follows from Assumptions A.1(b) and A.3(b). Now write

$$\begin{aligned} A_{n,2} &= n^{-2}\sum_{t=1}^{n} \mathrm{E}\left\{\int_A \tilde{R}(w,W_t)^2 a(w)\,dF(w)\right\} \\ &\quad + 2n^{-2}\sum_{1\le i<j\le n} \mathrm{E}\left\{\int_A \tilde{R}(w,W_i)\tilde{R}(w,W_j)a(w)\,dF(w)\right\} \\ &= n^{-1}h^{-d/2}\left\{\mathrm{E}H_n(W_0,W_0) + 2n^{-1}\sum_{1\le i<j\le n}\mathrm{E}H_n(W_i,W_j)\right\}. \end{aligned}$$

We want to show

$$\begin{aligned} \mathrm{E}H_n(W_0,W_0) &= h^{-d/2}B_1 + h^{-d/2+2}B_2 - h^{(d_2-d_1-d_3)/2}B_3 - h^{(d_3-d_1-d_2)/2}B_4 \\ &\quad + h^{(d_2+d_3-d_1)/2}B_5 \\ &\quad + O(h^{d/2} + h^{-d/2+4} + h^{(d_3-d_1-d_2+4)/2} + h^{(d_2-d_1-d_3+4)/2} \\ &\qquad + h^{(d_2+d_3-d_1+4)/2}) \end{aligned} \tag{B.6}$$

and

$$D_n \equiv 2n^{-1}\sum_{1\le i<j\le n}\mathrm{E}H_n(W_i,W_j) = o_p(1). \tag{B.7}$$

Now $\mathrm{E}H_n(W_0,W_0) = \mathrm{E}[h^{d/2}\int\tilde{R}(w,W_0)\tilde{R}(w,W_0)a(w)\,dF(w)] = \sum_{i,j=1}^{4}h^{d/2}\mathrm{E}\{\int R_i(w,W_0)R_j(w,W_0)a(w)\,dF(w)\} + O(h^{d/2}) \equiv \sum_{i=1}^{10}B_{n,i} + O(h^{d/2})$, where $B_{n,i} = h^{d/2}\mathrm{E}\int R_i(w,W_0)^2 a(w)\,dF(w)$ for $1\le i\le 4$, $B_{n,i} = h^{d/2}\mathrm{E}\int 2R_1(w,W_0)R_{i-3}(w,W_0)a(w)\,dF(w)$ for $5\le i\le 7$, $B_{n,i} = h^{d/2}\mathrm{E}\int 2R_2(w,W_0)R_{i-5}(w,W_0)a(w)\,dF(w)$ for $8\le i\le 9$, and $B_{n,10} = h^{d/2}\mathrm{E}\int 2R_3(w,W_0)R_4(w,W_0)a(w)\,dF(w)$. We can expand each term to the order of negligible asymptotic effects to obtain (B.6). For example,

$$\begin{aligned} B_{n,1} &= h^{-d/2}\int_{\mathbb{R}^d}K(u)^2\,du\int_A a(w)\,dw + h^{-d/2+2}\int_{\mathbb{R}^d}K(u)^2u_1^2\,du \\ &\quad \times\sum_{i=1}^{d}\int_A \frac{1}{2}\frac{\partial^2 f(w)}{\partial w_i^2}\frac{a(w)}{f(w)}\,dw + O(h^{-d/2+4}). \end{aligned}$$

To show $D_n = o_p(1)$, let $m = [L\log n]$ (the integer part of $L\log n$), where L is a large positive constant so that $n^4\beta_m^{\delta/(1+\delta)} = o(1)$ for some $\delta > 0$ by Assumption A.1(a). (For example, for fixed $\delta > 0$, if $\rho < 1/2.71828$ in Assumption A.1(a), $B = 4(1+\delta)/\delta$ would suffice.) 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 $u_n(p) \le C(h^d)^{1/p-1/2}$ with $p = 1+\delta$ (see the proof of Lemma B.4) to obtain $D_{n,a} = n^{-1}\sum_{j-i>m}\mathrm{E}H_n(W_i,W_j) \le Cn^{-1}n^2(h^d)^{1/(1+\delta)-1/2}\beta_m^{\delta/(1+\delta)} =$

$o(nh^{-d/2}\beta_m^{\delta/(1+\delta)}) = o(1)$. For case (b), use the bound $u_n(1) \leq Ch^{d/2}$ to obtain $D_{n,b} = n^{-1}\sum_{j-i\leq m} \mathrm{E}H_n(W_i, W_j) \leq Cn^{-1}nmh^{d/2} = O(mh^{d/2}) = o(1)$. Consequently, (B.7) holds.

Last, given $h = o(1)$ and the restrictions on d_i, $i = 1, 2, 3$, it is easy to verify

$$O(h^{d/2} + h^{-d/2+4} + h^{(d_3-d_1-d_2+4)/2} + h^{(d_2-d_1-d_3+4)/2} + h^{(d_2+d_3-d_1+4)/2}) = o(1), \quad \textbf{(B.8)}$$

where, e.g., $h^{(d_2+d_3-d_1+4)/2} = o(1)$ because $d \leq 7$ implies $d_1 \leq 5$. Combining (B.5)–(B.8), the conclusion follows. ∎

LEMMA B.6. *Let* $\tilde{\Delta}_n = \Gamma(\hat{f}, \hat{F}) - \Gamma(\hat{f}, F)$. *Then under Assumptions A.1–A.3 and* H_0, $nh^{d/2}\tilde{\Delta}_n = o_p(1)$.

Proof. By the same argument used to obtain the expansion of $\Gamma(\hat{f}, F)$, we obtain that under H_0,

$$\Gamma(\hat{f}, \hat{F}) = \frac{1}{4}\int \left\{ \frac{\hat{f}(x,y,z)}{f(x,y,z)} - \frac{\hat{f}(x,y)}{f(x,y)} - \frac{\hat{f}(x,z)}{f(x,z)} + \frac{\hat{f}(x)}{f(x)} \right\}^2 a(w)\, d\hat{F}(w)$$

$$+ O_p(\|\hat{f}(x,y,z) - f(x,y,z)\|_\infty^3).$$

It thus suffices to show that

$$\Delta_n \equiv \int \left\{ \frac{\hat{f}(x,y,z)}{f(x,y,z)} - \frac{\hat{f}(x,y)}{f(x,y)} - \frac{\hat{f}(x,z)}{f(x,z)} + \frac{\hat{f}(x)}{f(x)} \right\}^2 a(w)\, d[\hat{F}(w) - F(w)]$$

$$= o_p(n^{-1}h^{-d/2}).$$

Write $\Delta_n = \int_A r_n(w)^2 a(w)\, d[\hat{F}(w) - F(w)] = n^{-3}\sum_{j,k,l=1}^n \{R(W_l, W_j)R(W_l, W_k)a(W_l) - \int R(w, W_j)R(w, W_k)a(w)\, dF(w)\} = \sum_{i=1}^4 \Delta_{n,i}$, where $\Delta_{n,1} \equiv n^{-3}\sum_{l\neq j,k}^n \{R(W_l, W_j) R(W_l, W_k)a(W_l) - \int R(w, W_j)R(w, W_k)\, a(w)\, dF(w)\}$ is the summation of the centered terms with $l \neq j$, $l \neq k$, and $j \neq k$, $\Delta_{n,2} \equiv 2n^{-3}\sum_{j\neq k}^n R(W_j, W_j)R(W_j, W_k)a(W_j)$ is the summation of the terms with $l = j \neq k$, $\Delta_{n,3} \equiv n^{-3}\sum_{j=1}^n R(W_j, W_j)^2 a(W_j)$ is the summation of the terms with $l = j = k$, and $\Delta_{n,4} \equiv -n^{-3}\sum_{j,k=1}^n \int R(w, W_j)R(w, W_k)a(w)\, dF(w)$ is the summation of the centering terms for $\Delta_{n,2}$ and $\Delta_{n,3}$.

Dispensing with the simpler terms first, we have by Assumption A.3(a), (B.3), the remarks following (B.3), and Lemmas B.3 and B.4,

$$-\Delta_{n,4} = n^{-1}I_n = n^{-1}(I_n - \mathrm{E}I_n) + n^{-1}\mathrm{E}I_n = n^{-2}h^{-d/2}\{nh^{d/2}(I_n - \mathrm{E}I_n)\}$$

$$+ n^{-1}(h^{2r} + n^{-1}h^{-d})$$

$$= O_p(n^{-2}h^{-d/2}) + O(n^{-1}h^{-d/2}h^{d/2+2r}) + O(n^{-2}h^{-d}) = o_p(n^{-1}h^{-d/2}) \quad \textbf{(B.9)}$$

and $\mathrm{E}|\Delta_{n,3}| = n^{-3}\sum_{j=1}^n \mathrm{E}[R(W_j, W_j)^2 a(W_j)] = O(n^{-2}h^{-2d}) = o(n^{-1}h^{-d/2})$. Consequently, by the Markov inequality,

$$\Delta_{n,3} = o_p(n^{-1}h^{-d/2}). \quad \textbf{(B.10)}$$

It is difficult to show that the other two terms are small. Our strategy is to use Lemmas A.2 and A.3 repeatedly and show these terms are asymptotically negligible in that

$\Delta_{n,i} = o_p(n^{-1}h^{-d/2})$, $i = 1$ and 2. For $j \neq k$, we can show that (recall the bar notation previously defined)

$$\mathrm{E}[R(W_j, W_j)R(W_j, W_k)a(W_j)] = O(h^{-d}) \tag{B.11}$$

and

$$\mathrm{E}[R(\overline{W}_j, \overline{W}_j)R(\overline{W}_j, \overline{W}_k)a(\overline{W}_j)] = O(h^{r-d}). \tag{B.12}$$

To bound $D_{n,1} \equiv \mathrm{E}(\Delta_{n,2}) = 2n^{-3}\sum_{j\neq k}^{n}\mathrm{E}[R(W_j, W_j)R(W_j, W_k)a(W_j)]$, we consider two different cases for $D_{n,1}$: (a) $|j-k| > m$ and (b) $|j-k| \leq m$. We use $D_{n,1a}$ and $D_{n,1b}$ to denote these two cases. By Lemma A.2 and (B.12), $D_{n,1a} = 2n^{-3}\sum_{|j-k|>m}\mathrm{E}[R(W_j, W_j)R(W_j, W_k)a(W_j)] \leq C\{n^{-1}h^{r-d} + n^{-3}n^2(h^{-d})^{(1+2\delta)/(1+\delta)}\beta_m^{\delta/(1+\delta)}\} = O(n^{-1}h^{-d/2}h^{r-d/2}) + o(n^{-1}h^{-d}\beta_m^{\delta/(1+\delta)}) = o(n^{-1}h^{-d/2})$. By (B.11), $D_{n,1b} = 2n^{-3}\sum_{|j-k|\leq m}\mathrm{E}[R(W_j, W_j)R(W_j, W_k)a(W_j)] \leq Cn^{-3}nmh^{-d} = O(n^{-2}mh^{-d}) = o(n^{-1}h^{-d/2})$. So $D_{n,1} = o(n^{-1}h^{-d/2})$.

Let $D_{n,2} \equiv \mathrm{E}(\Delta_{n,2})^2 = 4n^{-6}\sum_{t_1\neq t_2}\sum_{t_3\neq t_4}\mathrm{E}\{R(W_{t_1}, W_{t_1})R(W_{t_1}, W_{t_2})a(W_{t_1})R(W_{t_3}, W_{t_3})R(W_{t_3}, W_{t_4})\, a(W_{t_3})\}$. 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 $D_{n,2s}$ to denote these cases $(s = a, b)$. Observe that by Lemma A.2 and (B.11), $D_{n,2a} \leq (D_{n,1a})^2 + C(n^{-2}(h^{-d})^{(2+4\delta)/(1+\delta)}\beta_m^{\delta/(1+\delta)}) = o(n^{-2}h^{-d})$. 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 $O(n^3 m)$. For $t_1 \neq t_2$ and $t_3 \neq t_4$, one can bound $\mathrm{E}|R(W_{t_1}, W_{t_1})R(W_{t_1}, W_{t_2})a(W_{t_1})R(W_{t_3}, W_{t_3})R(W_{t_3}, W_{t_4})a(W_{t_3})|$ by Ch^{-2d} if $\{t_1, t_2\} \cap \{t_3, t_4\} \neq \{t_1, t_2\}$ and by Ch^{-3d} otherwise. Consequently, $D_{n,2b} \leq C(n^{-6}n^3mh^{-2d} + n^{-6}n^2h^{-3d}) = o(n^{-2}h^{-d})$. So $\mathrm{E}(\Delta_{n,2})^2 = o(n^{-2}h^{-d})$, and by the Chebyshev inequality, we have

$$\Delta_{n,2} = o_p(n^{-1}h^{-d/2}). \tag{B.13}$$

Now, we want to show

$$\Delta_{n,1} = o_p(n^{-1}h^{-d/2}). \tag{B.14}$$

Write $\Delta_{n,1} = n^{-3}\sum_{l\neq j,k}^{n}\{R(W_l, W_j)R(W_l, W_k)a(W_l) - \mathrm{E}[R(W_l, W_j)R(W_l, W_k)a(W_l)| W_j, W_k]\} + n^{-3}\sum_{l\neq j,k}^{n}\{\mathrm{E}[R(W_l, W_j)R(W_l, W_k)a(W_l)|W_j, W_k] - \int R(w, W_j)R(w, W_k)\, a(w)\, dF(w)\} \equiv \Delta_{n,1,1} + \Delta_{n,1,2}$. By Lemma A.3,

$$\begin{aligned}\mathrm{E}|\Delta_{n,1,2}| &\leq n^{-3}\sum_{l\neq j,k}^{n}\mathrm{E}\left|\mathrm{E}[R(W_l, W_j)R(W_l, W_k)a(W_l)|W_j, W_k]\right.\\ &\qquad\left. - \int R(w, W_j)R(w, W_k)a(w)\, dF(w)\right|\\ &\leq C\{[(h^{-d})^{2\delta/(1+\delta)} + n^{-1}(h^{-d})^{(1+2\delta)/(1+\delta)}]\beta_m^{\delta/(1+\delta)} + (n^{-1}m + n^{-2}mh^{-d})\}\\ &= o(n^{-1}h^{-d/2}),\end{aligned}$$

and by the Markov inequality

$$\Delta_{n,1,2} = o_p(n^{-1}h^{-d/2}). \qquad \textbf{(B.15)}$$

Now let $S_{j,k,l} \equiv R(W_l, W_j)R(W_l, W_k)a(W_l) - \mathrm{E}[R(W_l, W_j)R(W_l, W_k)a(W_l)|W_j, W_k]$; then $\Delta_{n,1,1} = n^{-3}\sum_{l\neq j,k}^{n} S_{j,k,l}$ with $\mathrm{E}(\Delta_{n,1,1}) = 0$ because $\mathrm{E}(S_{j,k,l}) = 0$ for all $l \neq j$ and $l \neq k$. Denote

$$D_{n,3} \equiv \mathrm{E}(\Delta_{n,1,1})^2 = n^{-6} \sum_{t_1\neq t_3,\, t_2\neq t_3,\, t_3} \sum_{t_4\neq t_6,\, t_5\neq t_6,\, t_6} \mathrm{E}\{S_{t_1,t_2,t_3}S_{t_4,t_5,t_6}\}.$$

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. Using $D_{n,3s}$ to denote these cases $(s = a, b, c, d)$, by Lemma A.2, one can show that $|D_{n,3s}| = o(n^{-2}h^{-d})$ for $s = a, b, c, d$. In sum, $D_{n,3} = o(n^{-2}h^{-d})$, and thus by the Chebyshev inequality

$$\Delta_{n,1,1} = o_p(n^{-1}h^{-d/2}). \qquad \textbf{(B.16)}$$

Combining (B.15) and (B.16), we have (B.14). The conclusion follows. ■

LEMMA B.7. *Under Assumptions A.1–A.3,* $nh^{d/2}\|\hat{f}(x,y,z) - f(x,y,z)\|_\infty^3 = o_p(1)$.

Proof. By (B.2) and Assumption A.3, $nh^{d/2}\|\hat{f}(x,y,z) - f(x,y,z)\|_\infty^3 = nh^{d/2}O_p(n^{-3/2}h^{-3d/2}(\ln n)^{\gamma/2} + h^{3r}) = O_p(n^{-1/2}h^{-d}(\ln n)^{\gamma/2} + nh^{d/2+3r}) = o_p(1)$. ■

Putting Lemmas B.2–B.7 together, we have proved Theorem 3.1.

APPENDIX C: Proof of Theorem 4.1

Let $\hat{f}^*(x), \hat{f}^*(x,y), \hat{f}^*(x,z)$, and $\hat{f}^*(x,y,z)$ be defined as $\hat{f}(x), \hat{f}(x,y), \hat{f}(x,z)$, and $\hat{f}(x,y,z)$ with $\mathcal{W}^*$ replacing $\mathcal{W}$. Let $\tilde{f}(w) \equiv \tilde{f}(x,y,z)$ denote the p.d.f. of $W_t^* = (X_t^{*\prime}, Y_t^{*\prime}, Z_t^{*\prime})'$, i.e., $\tilde{f}(x,y,z) \equiv \tilde{f}(y|x)\tilde{f}(z|x)\tilde{f}(x)$, and denote the corresponding c.d.f. as $\tilde{F}(w)$. Let $\hat{F}^*(w)$ denote the empirical distribution of $\{W_t^*\}$. We first state a lemma that is proved in Appendix D.

LEMMA C.1. *Suppose that Assumptions A.1–A.3 and A.5 hold. Then*

(i) $\sup_{x\in A\cap\mathbb{R}^{d_1}}|\hat{f}^*(x) - \tilde{f}(x)| = O_p(n^{-1/2}\, h^{-d_1/2}(\ln n)^{\gamma/2} + h^r)$;
(ii) $\sup_{x\in A\cap\mathbb{R}^{d_1+d_2}}|\hat{f}^*(x,y) - \tilde{f}(x,y)| = O_p(n^{-1/2}h^{-(d_1+d_2)/2}(\ln n)^{\gamma/2} + h^r)$;
(iii) $\sup_{x\in A\cap\mathbb{R}^{d_1+d_3}}|\hat{f}^*(x,z) - \tilde{f}(x,z)| = O_p(n^{-1/2}h^{-(d_1+d_3)/2}(\ln n)^{\gamma/2} + h^r)$;
(iv) $\sup_{x\in A}|\hat{f}^*(x,y,z) - \tilde{f}(x,y,z)| = O_p(n^{-1/2}\, h^{-d/2}(\ln n)^{\gamma/2} + h^r)$.

Let $A_1 = A \cap \mathbb{R}^{d_1}$, $A_2 = A \cap \mathbb{R}^{d_1+d_2}$, and $A_3 = A \cap \mathbb{R}^{d_1+d_3}$. Define $\mathcal{S}_1(C) = \{\sup_{x\in A_1}|\tilde{f}(x) - f(x)| \le C(n^{-1/2}b^{-d_1/2}(\ln n)^{\gamma/2} + b^2),\ \max_{1\le i\le d_1}\sup_{x\in A_1}|\partial^r\tilde{f}(x)/\partial^r x_i| \le C\}$. Similarly, we define the random sets $\mathcal{S}_2(C), \mathcal{S}_3(C)$, and $\mathcal{S}_4(C)$ for $\tilde{f}(x,y) \equiv \tilde{f}(y|x)\tilde{f}(x)$, $\tilde{f}(x,z) \equiv \tilde{f}(z|x)\tilde{f}(x)$, and $\tilde{f}(x,y,z)$, respectively. A standard result gives $\sup_{x\in A_1}|\tilde{f}(x) - f(x)| = O_p(n^{-1/2}b^{-d_1/2}(\ln n)^{\gamma/2} + b^2)$ (e.g., Fan and Yao, 2003). By the proof of Lemma C.1 in Appendix D, $\sup_{x\in A_1}|\partial^r\tilde{f}(x)/\partial^r x_i| = O_p(1)$ for $i = 1, \ldots, d_1$.

So for any $\epsilon > 0$, there exists a sufficiently large constant C_1 such that $P(\mathcal{S}_1^c(C_1)) \le \epsilon/4$ for sufficiently large n, where $\mathcal{S}_1^c$ is the complement of $\mathcal{S}_1$. Similarly, for any $\epsilon > 0$, there exists a sufficiently large constant C_j such that $P(\mathcal{S}_j^c(C_j)) \le \epsilon/4$ for sufficiently large n and for $j = 2,3,4$. Let $C = \max_{1\le i\le 4} C_i$ and $\mathcal{S}(C) = \cap_{i=1}^4 \mathcal{S}_i(C)$. Then the Bonferroni inequality implies that for any $\epsilon > 0$, there exists a sufficiently large constant C such that $P(\mathcal{S}^c(C)) \le \epsilon$.

To show (i), by the law of iterated expectations, we have

$$P(T_n^* \le u|\mathcal{W}) = P(T_n^* \le u|\mathcal{W} \cap \mathcal{S}(C))P(\mathcal{S}(C)) \\ + P(T_n^* \le u|\mathcal{W} \cap \mathcal{S}^c(C))P(\mathcal{S}^c(C)).$$

Because the second term in the last expression can be made arbitrarily small for sufficiently large n, it suffices to show $P(T_n^* \le u|\mathcal{W} \cap \mathcal{S}(C)) \to \Phi(u)$ for all $u \in \mathbb{R}$, where $\Phi(\cdot)$ is the standard normal c.d.f.

Define

$$\hat{\Gamma}^* \equiv \Gamma(\hat{f}^*, \hat{F}^*) = n^{-1}\sum_{t=1}^n \left\{1 - \sqrt{\frac{\hat{f}^*(X_t^*,Y_t^*)\hat{f}^*(X_t^*,Z_t^*)}{\hat{f}^*(X_t^*,Y_t^*,Z_t^*)\hat{f}(X_t^*)}}\right\}^2 a(X_t^*,Y_t^*,Z_t^*).$$

One can modify the proofs of Lemmas B.1 and B.2 to obtain

$$\Gamma(\hat{f}^*, \tilde{F}) = \frac{1}{4}\int\left\{\frac{\hat{f}^*(x,y,z)}{\tilde{f}(x,y,z)} - \frac{\hat{f}^*(x,y)}{\tilde{f}(x,y)} - \frac{\hat{f}^*(x,z)}{\tilde{f}(x,z)} + \frac{\hat{f}^*(x)}{\tilde{f}(x)}\right\}^2 a(w)\, d\tilde{F}(w) \\ + O_p(\|\hat{f}^*(x,y,z) - \tilde{f}(x,y,z)\|_\infty^3),$$

where

$$\|\hat{f}^*(x,y,z) - \tilde{f}(x,y,z)\|_\infty \equiv \sup_{(x,y,z)\in A} |\hat{f}^*(x,y,z) - \tilde{f}(x,y,z)|.$$

Let $r_n^*(w)$, $R^*(w,u)$, $\tilde{R}^*(w,u)$, I_n^*, $G_n^*(u)$, $H_n^*(u,v)$ be defined as $r_n(w)$, $R(w,u)$, $\tilde{R}(w,u)$, I_n, $G_n(u)$, $H_n(u,v)$ with $\hat{f}^*$, $\tilde{f}$, and $\tilde{F}$ replacing $\hat{f}$, f, and F. Throughout, let E^* denote the expectation with respect to the smoothed kernel density $\tilde{f}(x,y,z)$ conditional on $\mathcal{W} \cap \mathcal{S}(C)$. Noticing that $\mathrm{E}^* G_n^*(W_i^*) = 0$ and $\mathrm{E}^* H_n^*(W_i^*, W_j^*) = 0$ for $i \ne j$, we have

$$I_n^* - \mathrm{E}^*[I_n^*] = 2n^{-1/2}h^r\left\{n^{-1/2}\sum_{i=1}^n G_n^*(W_i^*)\right\} + 2n^{-1}h^{-d/2}\left\{n^{-1}\sum_{1\le i<j\le n} H_n^*(W_i^*,W_j^*)\right\} \\ + n^{-1}h^{-d/2}\left\{n^{-1}\sum_{i=1}^n [H_n^*(W_i^*,W_i^*) - \mathrm{E}^* H_n^*(W_i^*,W_i^*)]\right\} \\ \equiv 2n^{-1/2}h^r U_{n,1}^* + 2n^{-1}h^{-d/2}U_{n,2}^* + n^{-1}h^{-d/2}U_{n,3}^*.$$

Conditional on $\mathcal{W}$, $\{W_i^*\}$ forms a triangular array of independent random variables, and so do $\{G_n^*(W_i^*)\}$ and $\{H_n^*(W_i^*, W_i^*)\}$. It is easy to verify that $U_{n,1}^* = O_p(1)$ and $U_{n,3}^* = O_p(n^{-1/2}h^{-d/2}) = o_p(1)$ conditional on $\mathcal{W} \cap \mathcal{S}(C)$ following the proof of Lemma 5.2 in Paparoditis and Politis (2000). By construction, $H_n^*(u,v) = H_n^*(v,u)$, and $\mathrm{E}^*H_n^*(W_i^*,v) = 0$. Let $G_n^*(u,v) = \mathrm{E}^*\{H_n^*(W_1^*,u)\ H_n^*(W_1^*,v)\}$. We verify that $\mathrm{E}^*[H_n^{*2}(W_1^*,W_2^*)] = \sigma^2 + o(1)$, $\mathrm{E}^*[H_n^{*4}(W_1^*,W_2^*)] \le Ch^{-d}$ and $\mathrm{E}^*[G_n^{*2}(W_1^*,W_2^*)] \le Ch^d$, and thus $\{\mathrm{E}^*[G_n^{*2}(W_1^*,W_2^*)] + n^{-1}\mathrm{E}^*[H_n^{*4}(W_1^*,W_2^*)]\}/\{\mathrm{E}^*[H_n^{*2}(W_1^*,W_2^*)]\}^2 \to 0$. Consequently, $U_{n,2}^* \xrightarrow{d} N(0,\sigma^2/2)$ conditional on $\mathcal{W} \cap \mathcal{S}(C)$ by Theorem 1 of Hall (1984).

Next, we can show that $nh^{d/2}\mathrm{E}^*[I_n^*] = \mathrm{E}^*H_n^*(W_1^*,W_1^*) = h^{-d/2}B_1 + h^{-d/2+2}B_2^* - h^{(d_2-d_1-d_3)/2}B_3^* - h^{(d_3-d_1-d_2)/2}B_4^* + h^{(d_2+d_3-d_1)/2}B_5^* + o(1)$, where for $i = 2,3,4,5$, B_i^* is defined as B_i with $\tilde{f}$'s replacing f's, e.g., $B_5^* = C_1^{d_1}\int_A a(w)\tilde{f}(w)/\tilde{f}(x)\,dw$. Let $\tilde{\Delta}_n^* = \hat{\Gamma}^* - \Gamma(\hat{f}^*,\tilde{F})$. We can show conditional on $\mathcal{W} \cap \mathcal{S}(C)$ that $\tilde{\Delta}_n^* = o_p(n^{-1}h^{-d/2})$ with arguments similar to but simpler than those used in the proof of Lemma B.6 because $\{W_t^*\}$ is an i.i.d. sequence given $\mathcal{W}$. Let $\hat{B}_i^*$, $i = 2,3,4,5$, be defined as $\hat{B}_i$ with $\{\mathcal{W}^*,\hat{f}^*\}$ replacing $\{\mathcal{W},\hat{f}\}$, e.g., $\hat{B}_5^* \equiv (C_1)^{d_1}n^{-1}\sum_{t=1}^n\{a(W_t^*)/\hat{f}^*(X_t^*)\}$. Applying Lemma C.1, we can show that $h^{2-d/2}(\hat{B}_2^* - B_2^*)$, $h^{(d_3-d_1-d_2)/2}(\hat{B}_3^* - B_3^*)$, $h^{(d_2-d_1-d_3)/2}(\hat{B}_4^* - B_4^*)$, $h^{(d_2+d_3-d_1)/2}(\hat{B}_5^* - B_5^*)$, and $nh^{d/2}\|\hat{f}^*(x,y,z) - \tilde{f}(x,y,z)\|_\infty^3$ are $o_p(1)$ conditional on $\mathcal{W} \cap \mathcal{S}(C)$. The completes the proof of part (i) of Theorem 4.1.

To show (ii), by the law of iterated expectations, it suffices to show that $P(T_n > T_n^*|\mathcal{W}) \to 1$ with probability approaching 1 (wpa.1) when $\Gamma(f,F) \ge \varepsilon > 0$. Using the notation defined previously, we only need to show $P(T_n > T_n^*|\mathcal{W} \cap \mathcal{S}(C)) \to 1$ wpa.1 provided that $\Gamma(f,F) \ge \varepsilon > 0$. Conditional on $\mathcal{W} \cap \mathcal{S}(C)$, we still have by part (i) that $T_n^* \xrightarrow{d} N(0,1)$, which holds as long as we generate the bootstrap data by imposing the null hypothesis. Hence, $T_n^* = O_p(1)$ conditional on $\mathcal{W} \cap \mathcal{S}(C)$, and so $P(T_n > T_n^*|\mathcal{W} \cap \mathcal{S}(C)) \to 1$ when $\Gamma(f,F) \ge \varepsilon > 0$ by Proposition 3.2. ■

APPENDIX D: Other Proofs

Proof of Proposition 3.2. The analysis is similar to that of Lemmas B.1 and B.6, now keeping the additional terms that do not vanish under the alternative. First, $\Psi(\tau) = \Psi(0) + \tau\Psi'(0) + o(\Psi'(0))$, where $\Psi(0) = \Gamma(f,F)$ and $\Psi'(0)$ is obtained from (B.1). So $\Gamma(\hat{f},F) = \Gamma(f,F) + \Psi'(0) + o(\Psi'(0))$. Noticing that Lemma B.6 also holds under the alternative (i.e., $\Gamma(\hat{f},\hat{F}) = \Gamma(\hat{f},F) + o_p(n^{-1}h^{-d/2})$), we have $\Gamma(\hat{f},\hat{F}) = \Gamma(f,F) + \Psi'(0) + o(\Psi'(0)) + o_p(1)$. It is easy to show that $n^{1/2}\Psi'(0) = O_p(1)$ when $\Gamma(f,F) > 0$. Thus $T_n = 4nh^{d/2}\Gamma(f,F)/\sqrt{2\sigma^2} + n^{1/2}h^{d/2}O_p(1) \xrightarrow{p} \infty$ if $\Gamma(f,F) > 0$. ■

Proof of Proposition 3.3. First, for the double array stochastic process $\{W_{nt}, 0 \le t \le n\}$, 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 $H_1(\alpha_n)$, $T_n - 4nh^{d/2}\Gamma(f^{[n]},F^{[n]})/\sqrt{2\sigma^2} \xrightarrow{d} N(0,1)$. Moreover, under $H_1(\alpha_n)$, $\Gamma(f^{[n]},F^{[n]}) = (\alpha_n^2/4)\int\triangle(w)^2a(w)\,dF^{[n]}(w) + o(\alpha_n^2)$. For $\alpha_n = n^{-1/2}h^{-d/4}$, $4nh^{d/2}\Gamma(f^{[n]},F^{[n]}) = \int\triangle(w)^2a(w)\,dF^{[n]}(w) \to \int\triangle(w)^2a(w)\,dF(w) \equiv \delta$ as $n \to \infty$. Consequently, $\Pr(T_n \ge z_\alpha|H_1(\alpha_n)) \to 1 - \Phi(z_\alpha - \delta/(\sqrt{2}\sigma))$. ■

Proof of Proposition 3.4. Under $H_{1,h}(\lambda_n,\gamma_n)$, $T_n - 4nh^{d/2}\Gamma(f^{[n]},F^{[n]})/\sqrt{2\sigma^2} \xrightarrow{d} N(0,1)$, and $4nh^{d/2}\Gamma(f^{[n]},F^{[n]}) = nh^{d/2}\lambda_n^2\int\Lambda((w - w_0)/\gamma_n)^2a(w)\,dF^{[n]}(w)\{1 + o(1)\} = nh^{d/2}\lambda_n^2\gamma_n a(w_0)f(w_0)\int\Lambda(w)^2\,dw\{1 + o(1)\} \to Ca(w_0)f(w_0)\int\Lambda(w)^2\,dw \equiv \bar{\delta}$ as $n \to \infty$. Consequently, $\Pr(T_n \ge z_\alpha|H_1(\alpha_n)) \to 1 - \Phi(z_\alpha - \bar{\delta}/(\sqrt{2}\sigma))$. ■

Proof of (3.1). To sketch the proof, first note that $\|\hat{f}^{(0)}(w) - f(w)\|_\infty = O_p(v_n)$ and $\|\hat{f}_i^{(2)}(w) - \partial^2 f(w)/\partial w_i^2\|_\infty = O_p(h_1^{-2}v_n)$, where $v_n \equiv n^{-1/2}h_1^{-d/2}(\ln n)^\gamma + h_1^p$ for some $\gamma > 0$. So

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\hat{f}_i^{(2)}(W_t)a(W_t)}{\hat{f}_{h_1}(W_t)^2} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 f(W_t)}{\partial w_i^2}\frac{a(W_t)}{f(W_t)^2}\{1 + h_1^{-2}v_n\}.$$

By Assumption A.1(b), $\xi_t \equiv (\partial^2 f(W_t)/\partial w_i^2)a(W_t)/f(W_t)^2$ is a bounded random variable with compact support A, and $\{\xi_t\}$ is a mixing process with the same mixing coefficients as $\{W_t\}$. One can thus apply a CLT for mixing processes to obtain

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 f(W_t)}{\partial w_i^2}\frac{a(W_t)}{f(W_t)^2} - \int_A\left(\frac{\partial^2 f(w)}{\partial w_i^2}\right)\frac{a(w)}{f(w)}\,dw = O_p(n^{-1/2}).$$

It then suffices to ensure that $h^{2-d/2}h_1^{-2}v_n = o(1)$, which holds by assumption. ■

Proof of Lemma C.1. We only prove part (i), because the proof of parts (ii) and (iii) is analogous and part (iv) follows from (i)–(iii). Let $A_1 = A \cap \mathbb{R}^{d_1}$. Write

$$\hat{f}^*(x) - \tilde{f}(x) = \{\hat{f}^*(x) - \mathrm{E}^*[\hat{f}^*(x)]\} + \{\mathrm{E}^*[\hat{f}^*(x)] - \tilde{f}(x)\} \equiv V_n(x) + B_n(x),$$

where $V_n(x)$ and $B_n(x)$ contribute to the variance and bias of the estimate of $\tilde{f}(x)$, respectively. Let $\epsilon > 0$ and $A_1^\epsilon = \{u : |u - v| \le \epsilon \text{ for } v \in A_1\}$. By Assumptions A.1, A.2, and A.5 and the uniform consistency of the kernel estimate of density derivatives on a compact set (cf. Hong and White, 2005, p. 899),

$$\sup_{x\in A_1}\left|\frac{\partial^r\tilde{f}(x)}{\partial x_i^r}\right| \le \sup_{x\in A_1}\left|\frac{\partial^r\tilde{f}(x)}{\partial x_i^r} - \mathrm{E}\left[\frac{\partial^r\tilde{f}(x)}{\partial x_i^r}\right]\right| + \sup_{x\in A_1}\left|\mathrm{E}\left[\frac{\partial^r\tilde{f}(x)}{\partial x_i^r}\right]\right|$$
$$= O_p(n^{-1/2}b^{-(d_1+2r)/2}(\ln n)^{\gamma/2}) + O(1) = O_p(1),$$

where x_i is the ith component of x. Noting that $\mathrm{E}^*[\hat{f}^*(x)] = \mathrm{E}^*[K_h(X_t^* - x)] = \int_{\mathbb{R}^{d_1}} K(u)\tilde{f}(x + hu)\,du$, by the rth-order Taylor expansion, the bias is

$$\sup_{x\in A_1}|B_n(x)| = \sup_{x\in A_1}\left|\int_{\mathbb{R}^{d_1}} K(u)\{\tilde{f}(x+hu) - \tilde{f}(x)\}\,du\right|$$
$$\le \frac{Ch^r}{r!}\int_{\mathbb{R}^{d_1}}|K(u)|\|u\|^r\,du \sup_{x\in A_1^\epsilon}\sum_{i=1}^{d_1}\left|\frac{\partial^r}{\partial x_i^r}\tilde{f}(x)\right| = O_p(h^r).$$

Noting that conditional on $\mathcal{W}$, $\{X_t^*\}$ is an i.i.d. sequence, standard arguments show that $\sup_{x\in A_1}|V_n(x)| = O_p(n^{-1/2}h^{-d_1/2}(\ln n)^{\gamma/2})$ (e.g., Newey, 1994, Lem. B.1; Masry, 1996, Thm. 2). Hence

$$\sup_{x\in A_1}|\hat{f}^*(x) - \tilde{f}(x)| = O_p(n^{-1/2}h^{-d_1/2}(\ln n)^{\gamma/2} + h^r).$$

■