Econometric Theory, **37**, 2021, 959–1003.
doi:10.1017/S0266466620000407

A NONPARAMETRIC TEST OF SIGNIFICANT VARIABLES IN GRADIENTS*

FENG YAO
Shenzhen University
West Virginia University
Guangdong University of Foreign Studies

TAINING WANG
Capital University of Economics and Business

We propose a nonparametric test of significant variables in the partial derivative of a regression mean function. The derivative is estimated by local polynomial estimation and the test statistic is constructed through a variation-based measure of the derivative in the direction of variables of interest. We establish the asymptotic null distribution of the test statistic and demonstrate that it is consistent. Motivated by the null distribution, we propose a wild bootstrap test, and show that it exhibits the same null distribution, whether the null is valid or not. We perform a Monte Carlo study to demonstrate its encouraging finite sample performance. An empirical application is conducted showing how the test can be applied to infer certain aspects of regression structures in a hedonic price model.

1. INTRODUCTION

Nonparametric estimation and hypothesis tests have gained popularity among practitioners because they are robust to functional misspecification under less restrictive assumptions (Li and Racine, 2007). The test of significant variables has been of interest to many researchers in regression analysis because it is often used to support or reject an economic theory or considered for model selection (see Hart, 1997 for a review of nonparametric tests). The test can be performed by using the sample analog of a moment condition (Fan et al., 1996; Zheng, 1996; Li and Wang, 1998; Li, 1999; Chen and Fan, 1999; Lavergne and Vuong, 2000; Hsiao et al., 2007; Gu et al., 2007; Li et al., 2016), by comparing the

*We would like to express our sincere thanks to the Editor, Peter C. B. Phillips, the Co-Editor, Liangjun Su, and three anonymous referees for their constructive suggestions and comments that improved the paper substantially. We also thank Carlos Martins-Filho, the participants in the 2017 Asian Meeting of the Econometric Society and the 4th Econometric Workshop in Dongbei University of Finance and Economics (DUFE) for valuable comments. Any remaining errors are the authors' responsibility. Address correspondence to Taining Wang, International School of Economics and Management, Capital University of Economics and Business, Beijing, China; e-mail: taining.wang@cueb.edu.cn.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

difference between sums of squared residuals (Ullah, 1985; Dette, 1999; Fan and Li, 2002), by an R^2-based statistic (Su and Ullah, 2013; Yao and Ullah, 2013), or by bootstrap procedures (Racine, 1997; Delgado and Manteiga, 2001). General tests of a parametric model can also be performed by comparing the difference between parametric and nonparametric fits (among many others, Eubank and Spiegelman, 1990; Staniswalis and Severini, 1991; Wooldridge, 1992; Yatchew, 1992; Azzalini and Bowman, 1993; Gozalo, 1993; Härdle and Mammen, 1993; Hjellvik and Tjøstheim, 1995; Hong and White, 1995; Bierens and Ploberger, 1997; Fan and Li, 1999; Li, 1999; Whang, 2000; Fan and Huang, 2001; Horowitz and Spokoiny, 2001; Stengos and Sun, 2001; Gao et al., 2009; Hong and Lee, 2013; Su and Lu, 2013; Lin et al., 2014).

Estimation of the derivative, or gradient, of the conditional expectation function frequently provides direct and valuable information to empirical researchers. Pioneered by Härdle and Stoker (1989) and Powell et al. (1989), average derivative estimation has received a lot of attention in the literature, because it does not need to assume parametric assumptions on either the derivative or the conditional expectation functions, provides a solution to estimate coefficients of index models, and achieves a root-n convergence rate through averaging. Many notable papers have been published lately. Among many others, following Härdle and Stoker (1989) and Powell et al. (1989), Coppejans and Sieg (2005) develop estimation in the context of bounded and discrete regressors, Cattaneo et al. (2014) derive a robust standard error formula for the average derivative estimator, and Kaido (2017) provides an efficient estimation of average derivatives with an interval-censored variable. With a local polynomial estimation, Li et al. (2003b), Banerjee (2007), and Altonji et al. (2012) estimate the average derivative, and Chaudhuri et al. (1997) estimate the average derivative quantile.

However, relatively little attention has been paid to testing significant variables in the partial derivative of a regression function. Racine (1997) proposes a bootstrap test for significant variables in the regression function based on partial derivative estimates. Sperlich et al. (2002) propose a test based on a cross derivative for the interaction term in an additive regression model. According to our review of the literature, no works have tested significant variables in the gradients, which would provide valuable insights. First, this test would provide independent insights into the structure of the gradient functions, regarding whether the partial derivative depends on all, some, or none of the variables in the regression. For example, in applied nonparametric estimation, a common practice is to report the partial derivative estimate at the mean value of independent variables, which can be misleading unless the gradient estimate does not change much over the support of independent variables. In estimating the stochastic frontier to evaluate the productive efficiency, for example in Yao et al. (2019), one important question arises regarding whether the marginal product of a regular input depends on computerization, an environment variable. Second, this test further allows one to infer about certain aspects of the structure of the

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

regression function.[1] For instance, in a bivariate regression with X and Z as regressors, if the gradient of X does not depend on X, one can expect that X enters the model linearly, implied by a varying coefficient that can depend on the other variable Z. If the gradient of X does not depend on X or Z, then not only X enters the model linearly, but also there is no interaction between X and Z variables, that is, implied by a partially linear model (PLM). Finally, as pointed out by Sperlich et al. (2002), "even though a certain test based on the estimation of a functional form is superior in detecting a general deviation from the hypothetical one, a single peak or bump can often be better detected by tests based on the derivatives."

In this paper, we fill the gap by providing a simple nonparametric test of significant variables in the partial derivative of a regression function. Recall that a general principle of the aforementioned nonparametric tests is to compare the difference between the null (parametric or semiparametric) and alternative (nonparametric) estimates. In our case, how to construct the gradient estimate satisfying the null restriction such that it does not depend on some variables is unclear. We overcome this obstacle by proposing a variation-based test that uses a local polynomial gradient estimate, to assess whether the gradients exhibit variations along the direction of variables that we test in the null.

Facilitated by the local polynomial gradient estimates, we obtain a simple asymptotic null distribution, whose bias and variance only depend on the error terms in the regression model. We further demonstrate that the test is consistent. Motivated by the simple structure of the null distribution, we propose a wild bootstrap test statistic, constructed by only resampling the estimated residuals. We demonstrate that the bootstrap test statistic exhibits the same asymptotic null distribution, whether the null is valid or not. This result substantially facilitates our ability to perform the test, because the bootstrap procedure will be valid across all different types of null hypotheses; thus, modifying either the procedure or its asymptotic justification is unnecessary. We further illustrate its encouraging finite sample performances through a Monte Carlo study.

Additivity is widely used in parametric and semiparametric models of economic data, because additivity alleviates the *curse of dimensionality*. The testing of additivity has been addressed in the literature (see Sperlich et al., 2002, Wang and Carriere, 2011, and references therein). We can apply our test for additivity. For example, in a bivariate regression, if the gradient of one variable X does not depend on the other variable Z throughout the support, one can infer an additive structure in the regression. Section 3.3 illustrates this applicability and compares it with alternative additivity tests.

Many researchers have considered tests for a semiparametric structure in the regression models. For example, Lewbel (1995) proposes a moment-based test of the Slutsky symmetry structure in the demand functions, which are estimated with a kernel method. Chen and Fan (1999), Aït-Sahalia et al. (2001), and Delgado

[1] We assume below that the variables in the regression are significant. Otherwise, one can think of the test of significance through whether the partial derivative is zero as a special case of our proposed test.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

and Manteiga (2001) have considered testing a nonparametric/semiparametric null model, with the null estimated with a kernel method. Li et al. (2003a) propose a series-based test for a semiparametric null model by weighting the unconditional moment. Recent work by Korolev (2018) proposes a consistent Lagrange Multiplier (LM)-type specification test for semiparametric models with an increasing number of series-based unconditional moments. Like the other omnibus specification tests, our test does not offer a model selection procedure, and remains silent on how to proceed if the null is rejected, that is, we should not use the test in a sequential multiple testing set-up. Our test has a different focus, that is, to test for the significant variables in the gradient function. Although we do not test for the regression structure directly, we can infer certain aspects of the regression model structure implied by the insignificant variables in the gradient function, as illustrated by an empirical illustration.

The remainder of the paper is as follows. We present the test in Section 2, the theoretical properties in Section 3, a bootstrap procedure and its property in Section 4, a simulation study in Section 5, an empirical illustration in Section 6, and conclusion in Section 7. We relegate the proof of two Theorems to Appendix 1, and collect additional proofs and the simulations' results in an online supplement.

2. A TEST FOR SIGNIFICANT VARIABLES IN GRADIENT

We consider a regression model

$$Y_i = m(W_i) + \epsilon_i, i = 1, 2, \ldots, n,$$

based on identically and independently distributed (i.i.d.) random variables $\{W_i, Y_i\}_{i=1}^n$, $W_i = (X_i, Z_i')' \in \Re^d$, $X_i \in \Re$, $Z_i \in \Re^{d-1}$, and we let (w_i, y_i) be a realization of (W_i, Y_i), with $w_i = (x_i, z_i')'$. With $E(\epsilon|W) = 0$, we define the gradient (first-order partial derivative) of the conditional mean with respect to the conditioning variable X as $g_X(W) \equiv (\partial E(Y|W))/(\partial X) = (\partial m(W))/(\partial X)$. Clearly, here, we can consider X to be any variable in W, and we assume that W are all significant in the conditional mean function, an assumption that can be tested with any of the aforementioned conditional mean significant variable tests.

2.1. A Variation-Based Test

We are interested in testing the significance of variables W^r in $g_X(W)$, with $W^r \subseteq W$ and $W^r \in \Re^{d_1}$, where $1 \leq d_1 \leq d$. We partition the vector W into W^r, d_1 variables to be tested, and W^c, the other $d - d_1$ variables in W excluding W^r. For notational simplicity, we write $\xi(W) = \xi(W^r; W^c)$ for any generic function $\xi(\cdot)$, highlighting the d_1 variables W^r whose significance is to be tested.

We assume that all W variables are significant in $m(\cdot)$, and express our null hypothesis that W^r are not significant in $g_X(W)$ formally as

$$H_0 : P(g_X(W^r; W^c) = g_X^0(W^c)) = 1 \text{ for some function } g_X^0(W^c). \quad \textbf{(1)}$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Because $g_X^0(W^c)$ is only a function of W^c, the null requires that $g_X(W)$'s value does not change with W^r almost surely for all W. Notably, nonparametric gradient estimates generally vary over their domain. Thus, we must formulate the test to detect whether the gradient only changes with W^c over the entire domain for all W. The tests of significant variables in the conditional mean generally use restricted conditional mean estimates. Because we test for the significance of variables in the gradient, we cannot rely on the restricted conditional mean directly. Furthermore, based on the regression model, how to estimate $g_X(W)$ subject to the null restriction is unclear. Thus, we create a different type of test formulation.

Denoting the density of random variable W by $f(W)$ and the density of W^r by $f_{W^r}(W^r)$, we propose the consideration of the following variation-based measure

$$T = \int \left[g_X(W_i^r; W_i^c) - \int g_X(W_j^r; W_i^c) f_{W^r}(W_j^r) dW_j^r \right]^2 f(W_i) dW_i,$$

for $j \neq i$. When $d_1 > 1$, the integral in T with respect to W^r is meant to be multivariate, that is, with respect to each of the d_1 elements in W^r. We introduce the density weight so that the integrals are equivalently expressed as expectations; thus, it captures the variation of $g_X(\cdot)$ in the directions of W^r. Notably, T measures the distance between $g_X(W_i^r; W_i^c)$ and $\int g_X(W_j^r; W_i^c) f_{W^r}(W_j^r) dW_j^r$ over the entire domain with a square of an L_2-norm. Equivalently, we define that $E_{W_j^r}(g_X(W_j^r; W_i^c)) \equiv \int g_X(W_j^r; W_i^c) f_{W^r}(W_j^r) dW_j^r$,

$$T = E[g_X(W_i^r; W_i^c) - E_{W_j^r}(g_X(W_j^r; W_i^c))]^2.$$

Notably, $T = 0$ under the null, and $T > 0$ when the null is not true. Thus, this variation-based T provides a valid target measure for us to perform the significance test of the gradient.

We construct an empirical estimate of T to perform the desired test by replacing the unknowns in T by their estimates, and we use the local polynomial gradient estimates detailed in Section 2.2,

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{g}_X(W_i) - \frac{1}{n-1} \sum_{\substack{j=1 \\ i \neq j}}^{n} \hat{g}_X(W_j^r; W_i^c) \right]^2. \tag{2}$$

$\hat{T}$ estimates its population counterpart T, replacing expectations with sample averages. To simplify asymptotic analysis, we restrict $i \neq j$, and rely on the "leave-one-out" local polynomial gradient estimate $\hat{g}_X(W_i)$, where the data used in estimation will not include the evaluation point W_i. $\hat{T}$ also calls for a local polynomial gradient estimate $\hat{g}_X(W_j^r; W_i^c)$, where the estimation of $\hat{g}_X$ will not use the ith and jth observations. Finally, our test statistic is presented as $nh^{2+\frac{d}{2}}\hat{T}$, a scaled version of $\hat{T}$. To manage the random denominator problem, in practice, one must perform trimming or introduce a weight function $\pi(W_i)$ to reweight the components in the test so that the derivative is estimated in areas where the density

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

is not too small, that is,

$$\hat{T} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_X(W_i)\pi(W_i) - \frac{1}{n-1}\sum_{\substack{j=1\\ i\neq j}}^{n}\hat{g}_X(W_j^r; W_i^c)\pi(W_j^r; W_i^c)\right]^2.$$

Here, we follow Li et al. (2003b) and Banerjee (2007) to consider a compact support for W, a reasonable assumption in many applications, that is, the number of hours worked is bounded from above by the legal limit of regular hours. Furthermore, we consider the local polynomial estimation below, which enjoys better boundary properties relative to the Nadaraya–Watson estimator. In this case, as indicated by the recent work of Fan and Guerre (2016), trimming can have negative consequences for testing problems, because trimming may decrease the power against some standard alternatives and against a more specific boundary alternative, which will be poorly detected. We present the test in (2) to illustrate the main properties with simplified notations, for example, $\pi(\cdot) = 1$.

2.2. Local Polynomial Gradient Estimate

We estimate the gradient with the popular local polynomial estimation (see Fan and Gijbels, 1996, Masry, 1996, Calonico et al., 2014 for discussions on the attractive properties of local polynomials and Su and Ullah, 2008 for applications in simultaneous equation models). We use the following notation to facilitate the theoretical representation. Let

$$\mathbf{j} = (j_1, \ldots, j_d),\ \mathbf{j}! = j_1! \times \cdots \times j_d!,$$
$$|\mathbf{j}| = \sum_{i=1}^{d} j_i,\ w^{\mathbf{j}} = w_1^{j_1} \times \cdots \times w_d^{j_d},\ \sum_{0\le|\mathbf{j}|\le p} = \sum_{i=0}^{p}\sum_{j_1=0}^{i}\cdots\sum_{\substack{j_d=0\\ |\mathbf{j}|=i}}^{i},$$

and

$$(D^{\mathbf{j}}m)(w) = \frac{\partial^{\mathbf{j}} m(w)}{\partial w_1^{j_1}\cdots\partial w_d^{j_d}}$$

A local polynomial estimator of order p, $\hat{a}_{\mathbf{j}}$, is obtained from minimizing the multivariate weighted least square criterion

$$\sum_{i=1}^{n}\left(Y_i - \sum_{0\le|\mathbf{j}|\le p} a_{\mathbf{j}}(W_i - w)^{\mathbf{j}}\right)^2 K\left(\frac{W_i - w}{h}\right), \tag{3}$$

where $K(\cdot)$ is a nonnegative kernel function on $\Re^d$, and $h \equiv h(n)$ is a scalar bandwidth sequence that goes to zero as n goes to infinity, assumed to be the same across different variables in W for notational simplicity.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Following Masry (1996), let $N_i = \binom{i+d-1}{d-1}$ be the number of distinct d-tuples $\mathbf{j}$ with $|\mathbf{j}| = i$. Arrange these N_i d-tuples as a sequence in lexicographical order (with the highest priority to the last position or that $(0,\ldots,0,i)$ is the first in the sequence and $(i,0,\ldots,0)$ is the last), and let G_i^{-1} to denote this one-to-one map. Then, the gradient estimate is $\hat{g}_X(w) = \hat{a}_{G_1(d)}$, because X appears as the first element in $W = (X,Z')'$.

Local quadratic estimation is frequently used to estimate the first-order partial derivative (Lu, 1996; Ruppert and Wand, 1994) and can be sufficient in some empirical situations considered for testing significant variables of gradient. The use of local polynomial estimation offers its generality and convenience in the theoretical consideration, illustrated more clearly through the requirements on the order of p to reduce the bias (see our Assumption A4(2) and its discussions below). Hence, we consider a general pth order local polynomial gradient estimation in constructing the test statistics.

3. ASYMPTOTIC PROPERTIES

We characterize the asymptotic properties of our proposed test $\hat{T}$ under the following assumptions. Below we denote a generic constant by C, the magnitude of which is inconsequential for the asymptotic analysis, and can vary from one place to another. We denote a generic function $\xi(w) \in C^j$ if $\xi(w)$, and all of its partial derivatives of order $\leq j$ are continuous and uniformly bounded on $\Re^d$.

3.1. Assumptions

A1. (1) $\{Y_i, W_i\}$ is i.i.d., and we denote the density of $W = (X,Z')'$ by $f(W)$; (2) $f(W) \in C^2$ and $0 < inf_{W\in\mathcal{W}} f(W)$, where $\mathcal{W}$, the support of W, is a compact subset of $\Re^d$. (3) The conditional density of W given ϵ is $f_{W|\epsilon}(W)$, and it is uniformly bounded on $\Re^d$.

A2. $K(\Psi) : \Re^d \to \Re$ is a product kernel $K(\Psi) = \prod_{j=1}^d \mathcal{K}(\psi_j)$ with symmetric $\mathcal{K}(\psi_j) : \Re \to \Re$ such that: (1) $\left|\mathcal{K}(\psi)\psi^j\right| \leq C$ for all $\psi \in \Re$ with $j = 0,1,\ldots,2p+1$; (2) $\int |\psi^j \mathcal{K}(\psi)| d\psi \leq C$ for $j = 0,1,\ldots,2p+1$; (3) $\int \mathcal{K}(\psi)d\psi = 1$, $\int \psi\mathcal{K}(\psi)d\psi = 0$; and (4) $\mathcal{K}(\psi)$ is continuously differentiable on $\Re$ with $|\psi^j d/(d\psi)\mathcal{K}(\psi)| \leq C$ for all $\psi \in \Re$ and $j = 0,1,\ldots,2p+1$.

A3. (1) $E(\epsilon|W) = 0$, $E(\epsilon^2|W) = \sigma^2(W)$ almost surely, $\sigma^2(w) \in C^1$, and $E(\epsilon^4|w) \in C^1$; (2) $m(w) \in C^{p+1}$.

A4. (1) $nh^{d+2} \to \infty$ as $n \to \infty$; (2) $nh^{2p+\frac{d}{2}+2} \to 0$ as $n \to \infty$, and $p+1 > \frac{d}{4}$; and (3) $nh^{\frac{3}{2}d} \to \infty$ as $n \to \infty$.

A5. Let $\mu_{k,\mathbf{j}} = \int_{\Re^d} \Psi^{\mathbf{j}} K(\Psi) d\Psi$. Define $S_{i,j}$ to be an $N_i \times N_j$ matrix with its $(l,m)th$ element being $[S_{i,j}]_{l,m} = \mu_{k,G_i(l)+G_j(m)}$. Then, S is a positive definite matrix

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

such that

$$S = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p,0} & S_{p,1} & \cdots & S_{p,p} \end{bmatrix}.$$

Assumption A1 requires that observations be i.i.d. across i, a typical assumption applicable to cross-sectional data. Notably, Härdle and Stoker (1989) require that the density vanish at the boundary of its support, a crucial assumption for their estimator to be consistent. We assume that W has a compact support, with a smooth density bounded away from zero, enabling a uniform argument on the density estimator component in the local polynomial estimation. The conditional density $f_{W|\epsilon}(W)$ is assumed to be uniformly bounded, to obtain the uniform convergence of the local polynomial estimator. A2 provides the standard moment and smoothness conditions on the kernel function to be used in the local polynomial estimation (see Masry, 1996 and Su and Ullah, 2008). A3 requires that the conditional mean function, the conditional heteroskedasticity function, and the fourth conditional moment of the error term be smooth and bounded. A4 specifies the bounds on the speed at which the bandwidth approaches zero as sample sizes increase, which allows a wide range of choices of bandwidths. Specifically, A4(2) reveals the delicate requirement between the order of the local polynomial and the dimension of W that are necessary for the leading bias term to approach zero. We discuss the choice of bandwidth that easily satisfies A4 in more detail in the simulation section. Finally, A5 enables us to state the probability limit of the local polynomial estimation (e.g., see A1 in Su and Ullah, 2008).

3.2. Asymptotic Distribution

We present the asymptotic null distribution in Theorem 1. We denote the (i,j)th element of matrix A by $[A]_{i,j}$, notably by the definition of $G(\cdot)$ function, $i = G_{|\mathbf{j}|}^{-1}(\mathbf{j})$, and we define

$$SK(\Psi) = \sum_{0\le|\mathbf{j}|\le p}\sum_{i=1}^{N_{|\mathbf{j}|}}[S^{-1}]_{1+d,\,\sum_{i'=0}^{|\mathbf{j}|-1}N_{i'}+i}K(\Psi)\Psi^{G_{|\mathbf{j}|}(i)} \tag{4}$$

$$\hat{SK}\left(\frac{W_j-W_i}{h}\right) = \sum_{0\le|\mathbf{j}|\le p}\sum_{i=1}^{N_{|\mathbf{j}|}}[S_n^{-1}(W_i)]_{1+d,\,\sum_{i'=0}^{|\mathbf{j}|-1}N_{i'}+i}K\left(\frac{W_j-W_i}{h}\right)\left(\frac{W_j-W_i}{h}\right)^{G_{|\mathbf{j}|}(i)}. \tag{5}$$

Let

$$S_{n,\mathbf{j}}(w) = \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{\mathbf{j}}.$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Define $S_{n,i,j}(w)$ to be an $N_i \times N_j$ matrix with its (l,m)th element being $[S_{n,i,j}(w)]_{l,m} = S_{n,G_i(l)+G_j(m)}(w)$. Then

$$S_n(w) = \begin{bmatrix} S_{n,0,0}(w) & S_{n,0,1}(w) & \cdots & S_{n,0,p}(w) \\ S_{n,1,0}(w) & S_{n,1,1}(w) & \cdots & S_{n,1,p}(w) \\ \vdots & \vdots & \cdots & \vdots \\ S_{n,p,0}(w) & S_{n,p,1}(w) & \cdots & S_{n,p,p}(w) \end{bmatrix}.$$

THEOREM 1. *Suppose the null H_0 such that W^r is not significant in $g_X(W)$ stated in (1) holds. Define*

$$B_{1n} = \frac{1}{n^2 h^d} \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} \hat{SK}^2 \left(\frac{W_j - W_i}{h} \right) \epsilon_j^2,$$

$$B_{3n} = \frac{1}{n^2 (n-1) h^{d+d_1}} \sum_{\substack{i=1 \\ i\neq j\neq l}}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \hat{SK} \left(\frac{W_j - W_i}{h} \right) \hat{SK} \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) \epsilon_j^2,$$

$$B_{2n} = \frac{1}{n^2 (n-1)^2 h^{d+d_1}} \sum_{\substack{i=1 \\ i\neq j\neq t\neq l}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \hat{SK} \left(\frac{W_t^r - W_j^r}{h}; \frac{W_t^c - W_i^c}{h} \right) \times \hat{SK} \left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h} \right) \epsilon_t^2,$$

and

$$\Omega = 2 \left(\int \sigma^4(W) dW \right) \left[\int \left(\int SK(\Psi_1 + \Psi) SK(\Psi) d\Psi \right)^2 d\Psi_1 \right].$$

Then under assumptions A1(1), (2), A2, A3, A4(1), (2), and A5, as $n \to \infty$ we have

$$nh^{2+\frac{d}{2}} \left[\hat{T} - \left(\frac{1}{nh^{2+d}} B_{1n} + \frac{1}{nh^{2+d-d_1}} (B_{2n} - 2B_{3n}) \right) \right] \xrightarrow{d} \mathcal{N}(0, \Omega). \quad \textbf{(6)}$$

The asymptotic null distribution contains bias terms. Using local polynomial estimation, Calonico et al. (2014) analytically develop the asymptotic distribution accounting for the effects of bias estimation. Interestingly, their Remark 7 formally justifies performing inference based on increasing the order of a local polynomial estimation, which automatically performs a bias-correction, and allows one to use Mean Squared Error (MSE)-optimal bandwidths. Due to the nature of constructing a variation-based test in (2), the bias in estimating the gradient is squared, of smaller order and controlled by Assumption A4(2). The bias in the asymptotic null arises from the "variance" term in estimating the gradient (see also the

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

nonlinearity bias term discussed in the two-step estimation procedure in Cattaneo and Jansson, 2018).

The asymptotic null distribution of $\hat{T}$ in Theorem 1 provides a basis for performing a hypothesis test under H_0. For implementation, we require consistent estimates of B_{in} for $i = 1, 2, 3,$ and Ω. We propose to estimate them with

$$\hat{B}_{1n} = \frac{1}{n^2 h^d} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \hat{SK}^2 \left(\frac{W_j - W_i}{h} \right) \hat{\epsilon}_j^2,$$

$$\hat{B}_{3n} = \frac{1}{n^2 (n-1) h^{d+d_1}} \sum_{\substack{i=1 \\ i \neq j \neq l}}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \hat{SK} \left(\frac{W_j - W_i}{h} \right) \hat{SK} \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) \hat{\epsilon}_j^2,$$

$$\hat{B}_{2n} = \frac{1}{n^2 (n-1)^2 h^{d+d_1}} \sum_{\substack{i=1 \\ i \neq j \neq t \neq l}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \hat{SK} \left(\frac{W_t^r - W_j^r}{h}; \frac{W_t^c - W_i^c}{h} \right) \times \hat{SK} \left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h} \right) \hat{\epsilon}_t^2$$

and

$$\hat{\Omega} = \frac{2}{n^2 h^d} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \frac{\hat{\epsilon}_i^2 \hat{\epsilon}_j^2}{\hat{f}(W_i) \hat{f}(W_j)} K \left(\frac{W_i - W_j}{h} \right) \times \left[\int \left(\int SK(\Psi_1 + \Psi) SK(\Psi) d\Psi \right)^2 d\Psi_1 \right],$$

where $\hat{\epsilon}_i = Y_i - \hat{m}(W_i)$ uses the local polynomial estimate $\hat{m}(W_i)$. We define the feasible centered test statistic as

$$\hat{T}_c = [nh^{2+\frac{d}{2}} \hat{T} - h^{-\frac{d}{2}} \hat{B}_{1n} - h^{-(\frac{d}{2} - d_1)} (\hat{B}_{2n} - 2\hat{B}_{3n})] / \sqrt{\hat{\Omega}}. \tag{7}$$

We provide the characterization of $\hat{B}_{in}$, $\hat{\Omega}$ and the asymptotic distribution of $\hat{T}_c$ under H_0.

COROLLARY 1. *Under H_0 and assumptions A1–A5, we have $\hat{B}_{1n} - B_{1n} = o_p(h^{\frac{d}{2}})$, $\hat{B}_{in} - B_{in} = o_p(h^{\frac{d}{2} - d_1})$ for $i = 2, 3$, $\hat{\Omega} - \Omega = o_p(1)$, and $\hat{T}_c \xrightarrow{d} \mathcal{N}(0, 1)$.*

The additional assumptions A1(3) and A4(3), compared with what we assume in Theorem 1, are necessary to demonstrate $\hat{B}_{1n} - B_{1n} = o_p(h^{\frac{d}{2}})$. The last statement in Corollary 1 demonstrates that the feasible centered test $\hat{T}_c$ is asymptotically pivotal. Thus, we can conduct a hypothesis test and reject H_0 when $\hat{T}_c > z_{1-\alpha}$ at a significance level of α for z_α being the αth quantile of the standard normal distribution.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

THEOREM 2. *(i) Suppose the alternative H_A holds, that is, W^r is significant in $g_X(W)$ or (1) does not hold. Then, under assumptions A1–A5,*

$$\frac{\hat{T}_c}{nh^{2+\frac{d}{2}}} = \frac{E[g_X(W_i) - E_{W_l^r}(g_X(W_l^r; W_i^c))]^2}{\sqrt{\Omega}} + o_p(1).$$

Thus, we have as $n \to \infty$, $P(\hat{T}_c > c_n) \to 1$ for any nonstochastic, positive sequence c_n such that $c_n = o(nh^{2+\frac{d}{2}})$. Thus, the test $\hat{T}_c$ is globally consistent.

(ii) Suppose the local alternative $H_A(\delta_n) : P(g_X(W^r; W^c) = g_X^0(W^c) + \delta_n D(W)) = 1$ holds with $\delta_n = n^{-\frac{1}{2}} h^{-1-\frac{d}{4}}$, $ED^4(W_i^r; W_j^c) < C$ for all i,j, and $D(w_i^r; w_j^c) \in C^1$. Define $E_{W_j^r}[D(W_j^r, W_i^c)] \equiv \int D(W_j^r, W_i^c) f_{W^r}(W_j^r) dW_j^r$. Then, under assumptions A1–A5, for $\Phi(\cdot)$ the cumulative distribution function of the standard normal, as $n \to \infty$ the local power of $\hat{T}_c$ satisfies

$$P\left(\hat{T}_c \geq z | H_A\left(n^{-\frac{1}{2}} h^{-1-\frac{d}{4}}\right)\right) \to 1 - \Phi\left(z - \frac{E[D(W_i) - E_{W_l^r}(D(W_l^r; W_i^c))]^2}{\sqrt{\Omega}}\right).$$

Below, we briefly sketch the idea used to analyze $\hat{T}$, providing insights into its asymptotic behavior.

$$\begin{aligned}
\hat{T} = {} & \frac{1}{n}\sum_{i=1}^{n}\left[g_X(W_i) - \frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n} g_X(W_l^r; W_i^c)\right]^2 \\
& + \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_X(W_i) - g_X(W_i) - \frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}(\hat{g}_X(W_l^r; W_i^c) - g_X(W_l^r; W_i^c))\right]^2 \\
& + \frac{2}{n}\sum_{i=1}^{n}\left[g_X(W_i) - \frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n} g_X(W_l^r; W_i^c)\right] \\
& \times \left[\hat{g}_X(W_i) - g_X(W_i) - \frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}(\hat{g}_X(W_l^r; W_i^c) - g_X(W_l^r; W_i^c))\right] \\
= {} & T_{1n} + T_{2n} + T_{3n}.
\end{aligned}$$

Under the null hypothesis, T_{1n} and T_{3n} vanish; thus, only T_{2n} determines the asymptotic null distribution in Theorem 1. From its proof, we observe that a quadratic form of the innovations $\{\epsilon_i\}$ plays the dominating role in determining the asymptotic null distribution (an observation also appears in Härdle and Mammen, 1993 and Kress et al., 2008). Furthermore, only the expressions based on ϵ

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

contribute to the asymptotic bias terms. This observation greatly simplifies the asymptotic distribution expressions, and further motivates our proposal for the bootstrap procedure in the next section.

Under the alternative, $g_X(W^r; W^c)$ varies with W^r and W^c; thus, T_{1n} and T_{3n} do not vanish. We can demonstrate that $T_{1n} = E[g_X(W_i) - E_{W_l^r}(g_X(W_l^r; W_i^c))]^2 + o_p(1)$ and $T_{3n} = o_p(1)$. It leads to Theorem 2(i), which establishes the global consistency of our test $\hat{T}_c$. For a comparison of our test with other tests, we obtain the asymptotic local power in Theorem 2(ii) under the local alternative $H_A(\delta_n)$, where $D(W)$ is a continuous function of W, representing the deviation of $g_X(\cdot)$ from $g_X^0(W^c)$. The deviation is weighted by $\delta_n \to 0$ as $n \to \infty$; thus, the alternative deviates from the null locally. Specifically, we should choose δ_n such that the asymptotic local power is bounded away from one and from the level, and we obtain $\delta_n = n^{-\frac{1}{2}} h^{-1-\frac{d}{4}}$. For example, when $d = 1$, and $h = O(n^{-1/5})$, the power of the test is asymptotically constant on regions of the form $\{g_X^0(W^c) + n^{-1/4} D(W) : E[D(W_i) - E_{W_j^r}[D(W_j^r, W_i^c)]]^2 = \text{ constant}\}$.

Theorem 2(ii) implies that the test has a nontrivial asymptotic power against alternatives that converges to the null at the rate of $n^{-\frac{1}{2}} h^{-1-\frac{d}{4}}$. By taking a large bandwidth, we can make the alternative magnitude (of order δ_n) against which the test has nontrivial power arbitrarily close to the parametric rate $n^{-\frac{1}{2}}$. Furthermore, as expected, the power increases with the magnitude of the departure from the null indicated by $(E[D(W_i) - E_{W_l^r}(D(W_l^r; W_i^c))]^2)/\sqrt{\Omega}$, and we provide more details on the power analysis in the next section.

3.3. Test of Additivity

A special case of our test is for $W^r = Z$ and $W^c = X$. Our $H_0 : P(g_X(W^r; W^c) = g_X^0(W^c)) = 1$ is equivalent to $H_0 : P(g_X(Z; X) = g_X^0(X)) = 1$. Because $g_X(\cdot)$ is only a function of X over the domain, $m(W)$ is additive in X and Z, that is, $m(W) = m_1(X) + m_2(Z)$; thus, we use this special case as a test of additivity. For $d_1 = d - 1$, $d_1 - \frac{d}{2} = \frac{d}{2} - 1$, the test statistic is $\hat{T}_c = [nh^{2+\frac{d}{2}} \hat{T} - h^{-\frac{d}{2}} \hat{B}_{1n} - h^{\frac{d}{2}-1} (\hat{B}_{2n} - 2\hat{B}_{3n})]/\sqrt{\hat{\Omega}}$, where $\hat{B}_{1n}$ and $\hat{\Omega}$ are the same as before,

$$\hat{B}_{3n} = \frac{1}{n^2(n-1)h^{2d-1}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \neq l}}^{n} \sum_{l=1}^{n} \hat{SK}\left(\frac{W_j - W_i}{h}\right) \hat{SK}\left(\frac{Z_j - Z_l}{h}; \frac{X_j - X_i}{h}\right) \hat{\epsilon}_j^2,$$

and

$$\hat{B}_{2n} = \frac{1}{n^2(n-1)^2 h^{2d-1}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \neq t \neq l}}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \hat{SK}\left(\frac{Z_t - Z_j}{h}; \frac{X_t - X_i}{h}\right) \times \hat{SK}\left(\frac{Z_t - Z_l}{h}; \frac{X_t - X_i}{h}\right) \hat{\epsilon}_t^2.$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

From Theorem 2(ii), we observe that the local power of $\hat{T}_c$ satisfies

$$P(\hat{T}_c \geq z | H_A(n^{-\frac{1}{2}} h^{-1-\frac{d}{4}})) \to 1 - \Phi\left(z - \frac{E[D(W_i) - E_{Z_l}(D(Z_l; X_i))]^2}{\sqrt{\Omega}}\right).$$

The local power is inversely related to the magnitude of Ω, which we can evaluate for a given order p of the local polynomial estimator and a kernel function. In practice, typical choices of p are 1 and 2, which correspond to the local linear and local quadratic regression, respectively. With the formula for partitioned inverse, we can demonstrate that for the Gaussian kernel, $\mathcal{K}(\psi) = (1/\sqrt{2\pi}) exp(-\psi^2/2)$,

$$\Omega = 2 \int \sigma^4(W) dW \left[\int \left(\int K(\Psi_1 + \Psi) K(\Psi) \Psi^{G_1(d)} (\Psi_1 + \Psi)^{G_1(d)} d\Psi \right)^2 d\Psi_1 \right]$$
$$= 2 \int \sigma^2(W) dW \left[\frac{3}{16} \left(\frac{1}{2\sqrt{2\pi}} \right)^d \right].$$

Thus, the local power does not vary with either $p = 1$ or $p = 2$. With $p = 3$, we can obtain the same expression for Ω when $d = 3$; thus, the test exhibits similar local power with $p = 1$ and 2. For general d and p, we do not have a simplified formula for Ω (S contains many nonzero elements and does not exhibit a simple structure for the partitioned inverse); thus, it must be derived in a case by case fashion.

We now compare the power of our test $\hat{T}_c$ according to $H_A(\delta_n)$ for a given bandwidth with two alternative tests of additivity. The first test is proposed in Sect. 6.2 of Sperlich et al. (2002), denoted by $\hat{T}_s$. In an additive regression model with an interaction term $f_{\alpha\beta}(x_\alpha, x_\beta)$, $\hat{T}_s$ is constructed as the square of the L_2-norm of the cross derivative of estimate of $f_{\alpha\beta}(x_\alpha, x_\beta)$. The test rule is to reject $H_{0,\alpha\beta} : f_{\alpha\beta}(x_\alpha, x_\beta) = 0$ in favor of the alternative $H_{1,\alpha\beta}(a_n) : f_{\alpha\beta} \in F_{\alpha\beta}(a_n)$ when $\hat{T}_s$ is sufficiently large, where $F_{\alpha\beta}(a)$ denotes a collection of $f_{\alpha\beta}(x_\alpha, x_\beta)$ such that the L_2 norm of the cross derivative of $f_{\alpha\beta}(x_\alpha, x_\beta)$ is greater than or equal to a. Notably, a_n in $H_{1,\alpha\beta}(a_n)$ denotes the magnitude of the deviation of the properly centered interaction term $f_{\alpha\beta}$ away from zero, and our local alternative $H_A(\delta_n)$ is of a very different nature, that is, $g_X(W) = g_X^0(X) + \delta_n D(W)$. Bearing this difference in mind, we further compare the two tests in terms of a_n and δ_n. By Thm. 10 in Sperlich et al. (2002)[2], $\hat{T}_s$ has a nontrivial power against the alternative $F_{\alpha\beta}(a_n)$ such that $a_n^{-1} = o(\sqrt{nh^5} + h^{-2})$. Using their Assumption A6 that $h = O(n^{-1/10})$, we obtain that $a_n = O(n^{-1/5+\delta})$ for some $\delta > 0$. So $\hat{T}_s$ has a nontrivial power against the alternative $F_{\alpha\beta}(a_n)$ with $a_n = O(n^{-1/5+\delta})$. For our test, with $h = O(n^{-1/(2p+d)})$, $\delta_n = n^{-1/2} h^{-1-d/4} = O(n^{-1/4})$ for $d = 2$ and $p = 2$. Thus, $\delta_n = o(a_n)$, and our $\hat{T}_c$ exhibits a larger local power against $\hat{T}_s$, in the sense that $\hat{T}_c$ can detect local alternatives that converge to the null at a faster rate.

[2]By inspecting their proof, we believe that the rate of a_n^{-1} should be $o(\sqrt{nh^5} + h^{-2})$, not $o(nh^5 + h^{-2})$ as stated.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

The second test is proposed in Wang and Carriere (2011), which tests the null of additivity of $m(W)$ against the local alternative $H_A(\delta_n^w) : m(W) = m^0(W) + \delta_n^w \Delta(W)$, where $m^0(W)$ denotes the additive specification of $m(W)$. Again, their local alternative $H_A(\delta_n^w)$ is on $m(W)$, and our $H_A(\delta_n)$ is on the derivative $g_X(W)$. Gozalo and Linton (2001) propose four tests of additivity in $m(\cdot)$, and one of them, $\hat{\tau}_0$, by taking a direct difference between the nonparametric estimate and the additive estimate, is shown to be most powerful. We focus on the test by Wang and Carriere (2011) because it exhibits better performance than $\hat{\tau}_0$. Based on Thm. 2 in Wang and Carriere (2011), we construct their centered test as $\hat{T}_{wc}$, and the local power of $\hat{T}_{wc}$ satisfies $P(\hat{T}_{wc} \geq z | H_A(\delta_n^w)) \to 1 - \Phi(z - (\int \Delta^2(W) dW)/\sqrt{V_2})$ as $n \to \infty$ for $\delta_n^w = n^{-1/2} h^{-d/4}$, where V_2 is the asymptotic variance of $\hat{T}_{wc}$ defined in Wang and Carriere (2011). Assuming that $\hat{T}_{wc}$ uses a bandwidth h of the same order as ours, we observe that $\hat{T}_{wc}$ exhibits larger local power, because $\hat{T}_{wc}$ has a nontrivial power against alternatives that converge to the null at rate $\delta_n^w = n^{-1/2} h^{-d/4}$, which goes to zero at a rate faster than our rate $\delta_n = n^{-1/2} h^{-1-d/4}$. If we focus on comparing just the power statements that are independent of the sample size, the power of $\hat{T}_{wc}$ increases with $(\int \Delta^2(W) dW)/\sqrt{V_2}$, and the power of $\hat{T}_c$ increases with $(E[D(W_i) - E_{Z_l}(D(Z_l; X_i))]^2)/\sqrt{\Omega}$. Assuming further that $\int \Delta^2(W) dW$ is of similar magnitude as $E[D(W_i) - E_{Z_l}(D(Z_l; X_i))]^2$, a special case in nature because $\Delta(W)$ is on the level of $m(W)$ while $D(W)$ is on the derivative, we can compare their powers through V_2 and Ω. From Wang and Carriere (2011) and Härdle and Mammen (1993), we obtain that $V_2 = 2 \int \sigma^4(W) dW \int \kappa(\Psi) \kappa(-\Psi) d\Psi = 2 \int \sigma^4(W) dW (1/(2\sqrt{2\pi}))^d > \Omega$, where $\kappa(\cdot)$ denotes the convolution of the kernel function and the second equality is valid for the Gaussian kernel. Thus, our test $\hat{T}_c$ can exhibit a larger power through a smaller asymptotic variance, but the gain is limited to small samples or when the deviation of the alternative from the null is significant, as we demonstrate in the simulation section.

4. A BOOTSTRAP TEST

The asymptotic distributions of our test $\hat{T}_c$ obtained in the last section provide guidance to perform the test. Specifically, we can compare the feasible centered test $\hat{T}_c$ with the critical value from a standard normal distribution and draw conclusions. However, many papers have revealed that the asymptotic normal approximation performs poorly in finite sample settings. Specifically, the consistent nonparametric test often suffers from substantial finite sample size distortions, because the distribution of the nonparametric test statistic approaches asymptotically the normal distribution at a slow convergence rate (e.g., Härdle and Mammen, 1993; Li and Wang, 1998; Fan et al., 2006; Hsiao et al., 2007; Gu et al., 2007), or the approximation from the first-order asymptotic theory is far too crude to be useful in practice unless the sample size is tremendously large (Hjellvik et al., 1998). Therefore, we provide a wild bootstrap test as a viable alternative for approximating the finite sample null distribution of the test statistic.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Many researchers have constructed the bootstrap test by imposing the null restriction in the bootstrap sample (i.e., Li and Wang, 1998; Fan et al., 2006; Hsiao et al., 2007; Gu et al., 2007), such that even when the null is not true, one still obtains the null distribution of the test statistics. This is a strategy that we could have followed. For example, when $W^r = Z$ in $g_X(W^r; W^c)$ is not significant, one can infer additivity in $m(\cdot)$, and we can use the additive estimate to construct the bootstrap sample. We note first that our null hypothesis is with regard to significant variables in $g_X(\cdot)$. Hence, placing inferred structural restrictions on $m(\cdot)$, although feasible, is not direct. Second, this strategy must be modified when W^r changes, requiring one to change the argument significantly depending on what appears in W^r.

Guided by the analysis of the null distribution of the test $\hat{T}_c$, we focus on the terms involving only the innovations $\{\epsilon_i\}$ mentioned in the last section. They have an unvarying form for all types of null hypotheses considered in this paper; thus, there is no need to modify either the bootstrap procedure or its argument for its asymptotic validity. As aforementioned, they play the dominating role in determining the asymptotic null distribution. We will show that they are asymptotically equivalent to the test statistics under the null hypothesis, that is, we practically bootstrap from a population that always reflects the null hypothesis.

We follow Härdle and Mammen (1993) to adopt a wild bootstrap procedure. Let $\hat{\epsilon}_i = Y_i - \hat{m}(W_i)$, where we focus on the aforementioned local polynomial estimate $\hat{m}(W_i)$. The bootstrap test contains three steps:

Step 1: generate ϵ_i^* as the wild bootstrap error. For example, ϵ_i^* is generated independently from the two-point distribution $\hat{F}_i$ such that $\epsilon_i^* = a\hat{\epsilon}_i$ for $a = (1-\sqrt{5})/2$ with probability $p = (\sqrt{5}+1)/(2\sqrt{5})$, and $\epsilon_i^* = b\hat{\epsilon}_i$ for $b = (1+\sqrt{5})/2$ with probability $1-p$. It is called the wild bootstrap error because we use only a single residual $\hat{\epsilon}_i$ to estimate the conditional distribution of ϵ_i given W_i by $\hat{F}_i$. It does not mimic the i.i.d. structure of $\{W_i, Y_i\}_{i=1}^n$. Let $E^*(\cdot) = E(\cdot|\{W_i, Y_i\}_{i=1}^n)$ be the expectation under the bootstrap distribution, that is, the conditional distribution given $\{W_i, Y_i\}_{i=1}^n$. We easily verify that $E^*(\epsilon_i^*) = 0$, $E^*(\epsilon_i^*)^2 = \hat{\epsilon}_i^2$, and $E^*(\epsilon_i^*)^3 = \hat{\epsilon}_i^3$.

Step 2: construct the bootstrap test statistic $\hat{T}_c^* = [nh^{2+\frac{d}{2}}\hat{T}^* - h^{-\frac{d}{2}}\hat{B}_{1n}^* - h^{-(\frac{d}{2}-d_1)}(\hat{B}_{2n}^* - 2\hat{B}_{3n}^*)]/\sqrt{\hat{\Omega}^*}$. Here,

$$\hat{T}^* = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_X^*(W_i) - \frac{1}{n-1}\sum_{\substack{j=1\\ i\neq j}}^{n}\hat{g}_X^*(W_j^r; W_i^c)\right]^2,$$

where $\hat{g}_X^*(W_i)$ and $\hat{g}_X^*(W_j^r; W_i^c)$ are calculated with the bootstrap sample $\{W_i, \epsilon_i^*\}_{i=1}^n$. That is, we use ϵ_i^* as the bootstrap dependent variable. $\hat{B}_{in}^*$ for $i = 1, 2, 3$ and $\hat{\Omega}^*$ are defined analogously to $\hat{B}_{in}$ and $\hat{\Omega}$ in equation (7), but with $\hat{\epsilon}_i$ replaced by ϵ_i^*.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Step 3: repeat Steps 1 and 2 B times, with B a large number. Then, the B bootstrap test statistics indexed by $\{\hat{T}^*_{c,b}\}_{b=1}^B$ yield the empirical distribution of the bootstrap statistics, which is then used to approximate the finite sample null distribution of $\hat{T}_c$. The empirical p-value is obtained as the percentage of the number of times that $\hat{T}^*_c$ exceeds $\hat{T}_c$ in the B repetitions.

Theorem 3 provides asymptotic justification for the aforementioned bootstrap procedure.

THEOREM 3. *With assumptions A1–A5, let z^*_α be the $(1-\alpha)$ quantile of the empirical distribution of $\{\hat{T}^*_{c,b}\}_{b=1}^B$. (i) We have as $n \to \infty$, $\hat{T}^*_c \xrightarrow{d} \mathcal{N}(0,1)$, conditionally on $\{W_i, Y_i\}_{i=1}^n$. (ii) $\lim_{n\to\infty} P(\hat{T}_c \geq z^*_\alpha) = \alpha$ under H_0. (iii) $\lim_{n\to\infty} P(\hat{T}_c \geq z^*_\alpha) = 1$ under H_A.*

Notably, our uncentered test statistics $\hat{T}$ and its associated bootstrap test statistic $\hat{T}^*$ are easier to implement. However, $\hat{T}_c$ is asymptotically pivotal; thus, we construct the bootstrap test statistic $\hat{T}^*_c$ based on $\hat{T}_c$, which should have a better size (see Horowitz, 1997).

We choose the bootstrap sample as $\{W_i, \epsilon^*_i\}_{i=1}^n$, with ϵ^*_i as the dependent variable. The asymptotic equivalence of $\hat{B}^*_{in}$ with the "true" ones for $i = 1, 2$, and 3 (as demonstrated in Corollary 1) is automatically satisfied. As mentioned by Cattaneo and Jansson (2018), this provides us the bootstrap consistency condition. It indicates that the bootstrap provides an asymptotic valid approximation to the null limiting distribution of $\hat{T}_c$. Theorem 3(i) holds regardless of whether H_0 is true. When H_0 is true, the bootstrap procedure leads asymptotically to the correct size of the test as in Theorem 3(ii), because $\hat{T}^*_c$ converges in distribution to the same limiting distribution under H_0 as in Corollary 1. When H_0 is false, $\hat{T}_c$ converges to infinity in Theorem 3, but asymptotically the bootstrap critical value z^*_α remains finite for any significance level α different from 0. Thus, $P(\hat{T}_c > z^*_\alpha) \to 1$ as in Theorem 3(iii), and the bootstrap method is consistent.

Our bootstrap method is nonstandard, because we essentially bootstrap the leading variance term of the original statistic. Our method enables convenient construction of the test and theoretical analysis. The construction of $\hat{T}^*_c$ focuses on the main components in $\hat{T}_c$, whose distribution does not depend on whether the null hypothesis holds, although $\hat{T}_c$'s distribution does. Therefore, the derived bootstrap test automatically follows the first guideline set by Hall and Wilson (1991). Thus, our bootstrap approximation to the null hypothesis distribution of $\hat{T}_c$ is always valid even when the data $\{W_i, Y_i\}_{i=1}^n$ were drawn from a population for which the null does not hold (see similar bootstrap in Kress et al., 2008). This ensures the adequate power of $\hat{T}^*_c$ against the departure from the null. On the other hand, it implies that the restricted conditional mean is zero. Because ϵ^*_i does not mimic the real data world regression under the null closely, we expect a cost of a less accurate approximation (size) than when using Y^*_i from a nonparametric regression model with H_0 directly imposed. In the test of additivity, where the null can be directly imposed in estimation to construct Y^*_i, we recommend using Y^*_i in $\hat{T}^*_c$ instead of

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

ϵ_i^*. In the simulations, we implement the bootstrap using both ϵ_i^* and Y_i^*. With ϵ_i^*, we do observe a slight relative size distortion, which disappears in larger sample sizes.

Remark 1. Cattaneo et al. (2019a) recently proposed an interesting approach to perform inference in a two-step estimation procedure where a first-order bias emerges with many included covariates in the first step. They propose a jacknife estimate of the bias and standard error and a new automatic bias-corrected two-step estimator. They further develop a wild bootstrap post-bias-correction inference. The newly proposed estimator and inference improve significantly over conventional two-step point estimators and inference procedures. We conjecture that a jacknife approach could be appropriately adopted, offering an interesting inference procedure for our test; however, this method is beyond the scope of this study, and we recommend it as a topic for further research.

Remark 2. The *curse of dimensionality* prevents the applicability of fully nonparametric models in large dimensions, and clearly our test is no exception. An empirically feasible model should be the model with some structures. For example, it would be interesting to use the PLM in Robinson (1988), modeling a subset of W nonparametrically, that is, $W_1 \equiv (X, Z_1)$, where Z_1 is a subset of Z, while leaving the other variables Z_2 in Z to enter the model linearly. We conjecture that our test can be applied to W_1, after we estimate Z_2's coefficient consistently at a $\sqrt{n}$ rate. In simulations, our test exhibits encouraging finite sample performance in the PLM (in the online supplement), and we apply it in the empirical application.

5. MONTE CARLO STUDY

5.1. Choice of Bandwidths

We discuss the important issue for the choice of bandwidths, because the performance of the test depends heavily on the bandwidths. First, in the kernel-based test for the semiparametric null model, it has been common to utilize different bandwidths for the testing and for the estimation step, such that the impact of the estimation step can be properly controlled, that is, see Gozalo and Linton (2001), Sperlich et al. (2002), and Wang and Carriere (2011). Sperlich et al. (2002) and Wang and Carriere (2011) also suggest a double-bandwidth strategy in estimating different components in the additive part, and it is a key tactic for achieving an increased finite sample precision of the estimate. Our test for significant variables in the gradient, different from the others aforementioned, requires choosing only one set of bandwidths.

Second, our assumptions A4(1)–(3) allow a relatively wide range for the bandwidths. To be specific, because we use a pth order local polynomial estimator for the first-order derivative and assume that $p-1$ is odd, the optimal bandwidth for estimating the first-order derivative is of order $O(n^{-1/(2(p+1)+d)})$, which violates our Assumption A4(2). However, when we estimate the regression function, one

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

normally selects a p'th order local polynomial estimator where p' is odd, that is, $p' = p - 1$. Then, the order of the optimal bandwidth for estimating regression function, that is, $h = O(n^{-(1/(2p+d))})$, satisfies the assumptions for $d < 4$. Thus, the regression optimal bandwidth selected from a cross-validation can serve as the tuning strategy. For $d > 4$, a slightly undersmoothed bandwidth should be used. This practice is in line, from a different perspective, with the recommendation in Calonico et al. (2014). Interestingly, they construct a robust confidence interval for a bias-corrected regression discontinuity estimator, together with a novel standard error estimator. Their Remark 7 provides a justification for using MSE-optimal bandwidth for the regression estimator (i.e., local linear), but constructs a confidence interval based on the higher order estimator suitable for the derivative (i.e., local quadratic). Below, we follow this choice, that is, $h = O(n^{-(1/(2p+d))})$, and investigate the robustness of our test using bandwidths of very different magnitudes, including a cross-validation bandwidth. We present a Monte Carlo study for our bootstrap test statistic $\hat{T}_c^*$, illustrating its finite sample performance and comparing it with alternative tests for additivity, and more detailed simulations of $\hat{T}_c^*$ are in the online supplement.[3]

Using higher order expansion, Calonico et al. (2018) show that robust bias-correction (RBC) in constructing confidence intervals with local polynomial estimates, through a proper studentization with a new standard error that incorporates the randomness in estimating the bias, performs better than undersmoothing. Such intervals are more robust to bandwidth choice, justifying the popular empirical practice of using MSE-optimal bandwidths. Using a novel Edgeworth expansion for t-statistics based on local polynomial methods uniformly over relevant classes of data generating processes, Calonico et al. (2019) demonstrate that a confidence interval constructed with RBC minimizes the worse-case coverage error and length in large samples. They develop optimal bandwidth and kernel selections in terms of optimal coverage error and interval length. RBC inference allows the use of the MSE-optimal bandwidth and this choice can be coverage error optimal for interior points. Cattaneo et al. (2019b) present more generally partition-based series estimators and obtain general pointwise and uniform distributional approximations for the RBC-based inference procedures. It is encouraging to observe that RBC-based inferences allow for the tuning parameter choices that minimize the mean squared error, offering a data-driven method that combines optimal point estimation and valid inference.

5.2. A Comparison Study

$\hat{T}_c^*$ for the null $W^r = Z$ provides a test for $m(W)$ to be additive in X and Z, as in Section 3.3. We illustrate and compare its finite performance with two

[3]We have conducted extensive simulation studies by testing various null hypotheses with $d = 2$ and $d = 3$, for both $\hat{T}_c^*$ and $\hat{T}^*$. The regression structures under the null include the additive, varying coefficient, and PLMs. We further probe the sensitiveness of our tests to the bandwidth by varying its magnitude or selecting it via cross-validation. The results remain encouraging and are reported in Appendix 3 of the online supplement.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

kernel-based tests for additivity in a simple bivariate regression. We first consider the additivity test by Sperlich et al. (2002) (STY hereafter). Their uncentered test $\hat{T}_s$ investigates the cross derivative of the interaction between X and Z in an additive model. Specifically,

$$\hat{T}_s = \int_{\mathbb{R}^2} \left(\hat{m}^{(1,1)}(X,Z)\right)^2 f(W)dW, \tag{8}$$

where $\hat{m}^{(1,1)}(X,Z)$ estimates the cross derivative, that is, $m^{(1,1)}(X,Z) \equiv (\partial^2 m(W)/\partial X \partial Z)$. A vanishing $\hat{T}_s$ indicates that the cross derivative is close to zero everywhere, supporting the additivity. We follow STY to calculate $\hat{m}^{(1,1)}(X,Z)$ with a local quadratic estimate and implement its bootstrap test as $\hat{T}_s^*$.

The second test considered is Wang and Carriere (2011) (WC hereafter). They recognize that the test in Gozalo and Linton (2001), which directly compares the L_2-norm of the difference between the fully nonparametric and additive estimates (i.e., $\hat{m}(W) - \hat{m}^0(W)$), contains a bias from estimating a nonlinear additive function. Furthermore, the existing choice of optimal bandwidth for regression may not be appropriate for the additivity test. WC propose a different uncentered test $\hat{T}_w$, which first performs a conditional smoothing on the residual from an additive estimation, and then constructs a L_2-norm of the smoothed residual, that is,

$$\hat{T}_w = \int_{\mathbb{R}^d} \hat{E}(Y - \hat{m}^0(W)|W)^2 f(W)dW, \tag{9}$$

where $\hat{E}(\cdot|W)$ denotes the Nadaraya–Watson estimator, and $\hat{m}^0(W) = \hat{\mu} + \hat{m}_1(X) + \hat{m}_2(Z)$, with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\hat{m}_j(\cdot)$ being the marginal integration estimate in Linton and Nielsen (1995). The test was shown to ameliorate the bias from estimating the nonlinear additive components, and requires less on the choice of the bandwidth (an optimal bandwidth for regression is feasible when the dimension is higher than two). We follow their suggestion to implement its bootstrap test as $\hat{T}_w^*$.

We use the following three bivariate DGPs adapted from Sperlich et al. (2002),

$$DGP_1 : Y_i = 0.5 + 2X_i + 1.5\sin(-1.5Z_i) + \delta_1 m_1(X_i, Z_i) + \epsilon_i$$
$$DGP_2 : Y_i = 0.5 + 1.5\sin(-1.5X_i) - Z_i^2 + \delta_1 m_2(X_i, Z_i) + \epsilon_i$$
$$DGP_3 : Y_i = 0.5 + 2X_i - Z_i^2 + \delta_1 m_3(X_i, Z_i) + \epsilon_i,$$

where X_i and Z_i are each i.i.d. and drawn independently from a uniform distribution $U(-2,2)$, and $\epsilon_i \sim \mathcal{N}(0,1)$ is the error term. The interaction functions $m_1(X_i,Z_i) = X_iZ_i$, $m_2(X_i,Z_i) = X_ie^{Z_i}$, and $m_3(X_i,Z_i) = X_i\sin(Z_i)$ reflect the alternative nonadditive structure. When $\delta_1 = 0$, $DGP_1 - DGP_3$ become additive.

Our test $\hat{T}_c^*$ uses the bootstrap dependent variable ϵ_i^*, offering convenient construction of the test and theoretical analysis. Its validity arises from the fact that our bootstrap approximation to the null distribution of $\hat{T}_c$ is always valid even when the data $\{W_i, Y_i\}_{i=1}^n$ were drawn from a population for which the null does not hold. As mentioned after Theorem 3, ϵ_i^* does not mimic the real data world under the null of additivity closely. We expect a more accurate approximation (size) by

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

using Y_i^* with the null directly imposed, which is totally feasible in an additivity test. Specifically, we use $Y_i^* = \hat{m}^0(W_i) + \epsilon_i^*$, where we use the aforementioned additive estimates as $\hat{m}^0(W_i)$. We perform Steps 2–3 in our bootstrap, construct the uncentered test $\hat{T}_a^*$ as in $\hat{T}^*$, but use the bootstrap sample $\{W_i, Y_i^*\}_{i=1}^n$. To construct the centered test $\hat{T}_{ac}^*$, we define $\hat{B}_{in}^*$ and $\hat{\Omega}^*$ analogously to $\hat{B}_{in}$ and $\hat{\Omega}$ in equation (7), but with $\hat{\epsilon}_i$ replaced by $\hat{\epsilon}_i^* = Y_i^* - \hat{m}^*(W_i)$, where $\hat{m}^*(W_i)$ are the local polynomial estimates using Y_i^* as the dependent variable. We implement both $\hat{T}_c^*$ and $\hat{T}_{ac}^*$, and expect a potential size improvement by using $\hat{T}_{ac}^*$ relative to $\hat{T}_c^*$, but not in power.

Implementing the bootstrap tests $\hat{T}_w^*$, $\hat{T}_s^*$, and $\hat{T}_{ac}^*$ requires estimating the additive functions, $\hat{m}_j(\cdot)$, with $j = 1, 2$. Through a marginal integration, we average the bivariate estimate of $m(W)$ over the other direction. Following STY and WC, we use two different bandwidths in $\hat{m}_j(\cdot)$, one for the direction of interest (h_1) and the other for the other direction to be averaged over (h_2). Sperlich et al. (2002) and Wang and Carriere (2011) both suggest using a double-bandwidth strategy that utilizes another set of bandwidths h for the testing step, potentially different from the bandwidths used in the estimation step. In contrast, our test $\hat{T}_c^*$ only requires a single set of bandwidths h.

For a fair comparison, we implement the four test statistics with the same Gaussian kernel function and carefully choose the bandwidth sequences. Because the optimal bandwidths for regression are compatible with our assumptions, we start with the *rule-of-thumb* bandwidth $h_\xi = C_h \hat{\sigma}_\xi n^{-\frac{1}{6}}$, where $\xi = X$ or Z. We have used different values of C_h to evaluate our test performance (see Appendix 3 in the online supplement). However, the test $\hat{T}_s^*$ requires a relatively large bandwidth (see Assump. A.6 in Sperlich et al., 2002); thus, we choose $C_h = 1.5$ and call the set of bandwidths selected as h.[4] Although both $\hat{T}_s^*$ and $\hat{T}_w^*$ recommend different bandwidths for estimation, that is, h_1 for the direction of interest in the additive function estimation and h_2 for the other direction to be averaged over, the recommended h_1, h_2, and h in the simulation of Sperlich et al. (2002) do not change much in magnitude; thus, we use h as discussed to estimate $m^0(W)$ and construct $\hat{T}_s^*$. In contrast, WC emphasize the double-bandwidth strategy and use h, h_1 and h_2 of very different magnitudes in their simulation. To construct their test, we mimic their choice of bandwidths by choosing h as aforementioned. To estimate $m^0(W)$ for $\hat{T}_w^*$, we use bandwidth $h_\xi = C_h \hat{\sigma}_\xi n^{-\frac{1}{6}}$, where $C_h = 0.5$ in h_1 for the direction of interest (undersmoothed), and $C_h = 1.0$ in h_2 for the other direction to be averaged over (oversmoothed).[5] We adopt the same estimation strategy of $m^0(W)$ for $\hat{T}_{ac}^*$.

[4] We observe from our simulation that an oversmoothing bandwidth h is critical for the power of $\hat{T}_s^*$. For example, when C_h is less than 1.3, the power of $\hat{T}_s^*$ is small and does not change much with the sample sizes, making the comparisons difficult. Thus, we adopt a larger constant $C_h = 1.5$ for all three tests.

[5] For $\hat{T}_w^*$ implementation, we have also set $C_h = (0.4, 1.6)$, $C_h = (0.3, 1.7)$ for the constant in the bandwidth of (h_1, h_2). The conclusion of the tests' comparison in terms of size and power does not change qualitatively. Thus, to save space, we do not report these results.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

TABLE 1. Additivity test comparison from a bivariate regression

Panel A	DGP_1											
	$n = 50$				$n = 100$				$n = 200$			
δ_1	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$
0.0	0.068	0.064	0.092	0.058*	0.060	0.056*	0.094	0.072	0.054	0.048*	0.078	0.056
0.2	0.252	0.214	0.218	0.138	0.424	0.416	0.586	0.224	0.672	0.668	0.978	0.264
0.4	0.726	0.688	0.448	0.366	0.938	0.908	0.904	0.494	1	1	1.000	0.674
0.6	0.932	0.914	0.538	0.540	1	0.994	0.968	0.664	1	1	1.000	0.858
0.8	0.992	0.990	0.592	0.614	1	1	0.994	0.744	1	1	1.000	0.906
1.0	1.000	1	0.656	0.672	1	1	1.000	0.802	1	1	1.000	0.948
Panel B	DGP_2											
	$n = 50$				$n = 100$				$n = 200$			
δ_1	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$
0.0	0.036	0.040*	0.136	0.066	0.044	0.052*	0.158	0.058	0.056	0.048*	0.184	0.048*
0.2	0.180	0.154	0.304	0.268	0.406	0.398	0.622	0.482	0.870	0.842	0.998	0.772
0.4	0.818	0.744	0.448	0.486	0.994	0.972	0.880	0.698	1	1	1.000	0.878
0.6	0.986	0.976	0.496	0.608	1	1	0.960	0.702	1	1	1.000	0.910
0.8	1	0.994	0.574	0.650	1	1	0.976	0.748	1	1	1.000	0.924
1.0	1	1	0.640	0.662	1	1	0.996	0.786	1	1	1.000	0.962
Panel C	DGP_3											
	$n = 50$				$n = 100$				$n = 200$			
δ_1	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$	$\hat{T}_c^*$	$\hat{T}_{ac}^*$	$\hat{T}_w^*$	$\hat{T}_s^*$
0.0	0.040	0.044*	0.144	0.044*	0.054	0.050*	0.146	0.086	0.048*	0.052*	0.174	0.056
0.2	0.426	0.468	0.264	0.242	0.846	0.788	0.662	0.328	1	0.958	0.988	0.458
0.4	0.992	0.986	0.452	0.572	1	1	0.942	0.768	1	1	1.000	0.882
0.6	1	1	0.566	0.670	1	1	0.982	0.864	1	1	1.000	0.952
0.8	1	1	0.652	0.786	1	1	0.990	0.872	1	1	1.000	0.980
1.0	1	1	0.698	0.784	1	1	0.994	0.916	1	1	1.000	0.990

Note: We report the empirical rejection frequency at 5% significance level from 500 simulations with 299 bootstrap repetitions. $\hat{T}_c^*$ and $\hat{T}_{ac}^*$ are our centered test statistics with $W^r = Z$ using the residual bootstrap method outlined in Section 4 and using additivity under H_0, respectively. $\hat{T}_w^*$ and $\hat{T}_s^*$ denote, respectively, the additivity tests by Wang and Carriere (2011) and Sperlich et al. (2002). All tests adopt Gaussian kernel and *rule-of-thumb* bandwidths.

We choose sample sizes $n = (50, 100, 200)$ and perform 500 simulations with 299 bootstrap repetitions to evaluate the empirical size by setting $\delta_1 = 0$ and power with $\delta_1 = (0.2, 0.4, \ldots, 1.0)$. Finally, we set the significance level α at 5%.

Table 1 presents the size and power of the four bootstrap tests under DGP_1, DGP_2, and DGP_3 in *Panel A*, *B*, and *C*, respectively. To facilitate the comparison, we put a $\star$ in the upper right corner of the size closest to the nominal level. In terms of size, $\hat{T}_{ac}^*$ clearly outperforms the others, being closest to the target significance level most of the time. The second best is $\hat{T}_c^*$, followed by $\hat{T}_s^*$. Their size approaches the target level as the sample size increases. $\hat{T}_w^*$ is generally over-sized in the

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

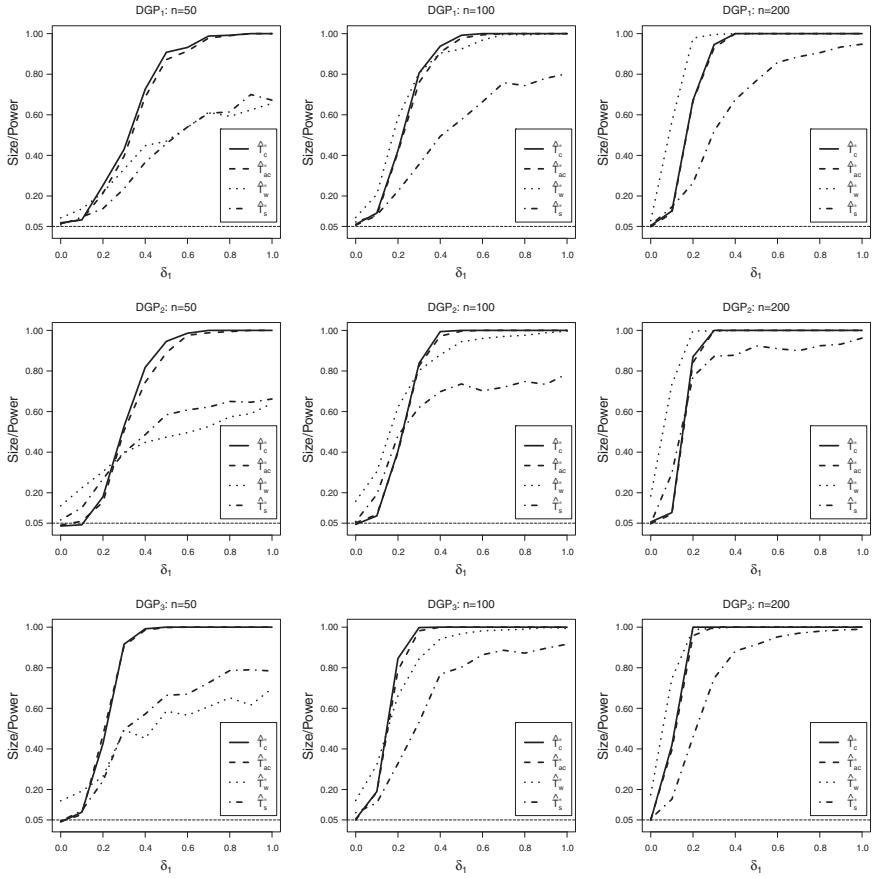


FIGURE 1. Power curve comparison of additivity tests at 5% rejection region.

experiments considered, and there is no clear indication that its size performance would improve with larger sample sizes. The observation that $\hat{T}^*_{ac}$ performs better than $\hat{T}^*_c$ is expected, because $\hat{T}^*_{ac}$ explicitly imposes the additive structure, which is missing in $\hat{T}^*_c$. Overall, our test captures the size fairly well relative to the other two tests.

On the other hand, the power of each test increases quickly as n doubles, or as the deviation from the null (δ_1) increases, indicating that all tests are consistent. The level of power, and the rate at which it increases with δ_1, are very different across the tests. To visualize the difference, we plot the power of four tests in Figure 1 by setting $\delta_1 = (0, 0.1, 0.2, \ldots, 1.0)$. When the DGP deviates slightly from the null, that is, δ_1 is very small, our tests $\hat{T}^*_c$ and $\hat{T}^*_{ac}$ exhibit less local power than $\hat{T}^*_w$ and more power than $\hat{T}^*_s$ (except in GDP_2). However, as δ_1 increases to be

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

larger than or equal to 0.3, $\hat{T}_c^*$ and $\hat{T}_{ac}^*$ outperform the other two tests in terms of power across different DGP's for $n = 50$ and 100. The pattern is clear in the large sample $n = 200$, with $\hat{T}_w^*$ being most powerful, followed by $\hat{T}_c^*$ or $\hat{T}_{ac}^*$, and generally, $\hat{T}_s$ is the least powerful test. This observation is expected. When $d = 2$, the convergence rate of $\hat{T}_w^*$ is nh (Thm. 3 in WC), faster than the rate of nh^3 in our test $\hat{T}_c^*$ and nh^5 in $\hat{T}_s^*$ (Thm. 9 in STY). With $n = 200$, $\hat{T}_c^*$, $\hat{T}_{ac}^*$, and $\hat{T}_w^*$ have power one when there is a moderate deviation from the null ($\delta_1 > 0.4$ in DGP_{1-2} and $\delta_1 > 0.2$ in DGP_3), and the power of $\hat{T}_s^*$ increases but remains distinguishably less than one. The observation that our test can be more powerful than $\hat{T}_w^*$ with a small sample and large deviation from the null is consistent with our local power comparison in Section 3.3. $\hat{T}_c^*$ is uniformly more or equally powerful relative to $\hat{T}_{ac}^*$, consistent with our conjecture that $\hat{T}_c^*$ exhibits adequate power but less accuracy in size relative to $\hat{T}_{ac}^*$. Again, the discrepancy becomes indistinguishable as the sample size increases. In all, our test, with a different strategy of approximating the null distribution, exhibits a reasonable performance of size and power, presenting a viable, perhaps more balanced, alternative to the other two additivity tests.

6. EMPIRICAL APPLICATION

The landmark study of Rosen (1974) pioneers the hedonic price model (HPM) in which residential house prices are influenced by both internal and external characteristics of houses being sold. Ekeland et al. (2004) show that the HPM involves a complex, nonlinear relationship between house prices and their various attributes, suitable for nonparametric rather than parametric modeling. With this inspiration, Bontemps et al. (2008) investigate the negative impact of pollution due to intensive agricultural activities on house prices in France through a semiparametric HPM:

$$P_i = Q_i'\beta + m(TMEAD_i, NITRO_i) + \epsilon_i,\ i = 1, \ldots, 2,092, \tag{10}$$

where P_i is the log of the house sales price paid by the buyers, $TMEAD_i$ is an aggregate indicator of agricultural pollution that measures the proportion of permanent grassland converted into cultivated grassland, and $NITRO_i$ is another indicator of pollution that captures the nitrogen emissions from livestock farming per hectare of arable land in the rural district where the residential house is located. Q_i is a vector of control variables, including four physical characteristics of the house (age of the house (*AGE*), state of repair (*REPAIR*), number of rooms (*ROOMS*), and lot size (*SIZE*)) and four characteristics of the surrounding districts (population of the district (*POPULATION*), average taxable family income (*INCOME*), proportion of vacant houses (*VACANT*), and whether the surrounding district is located in an urbanized department of Ille et Vilaine (*COUNTY*)). By controlling Q linearly, one can investigate the potentially negative

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

impacts of *TMEAD* or *NITRO* on housing prices through the nonparametric function $m(\cdot)$.[6]

After estimating β with $\hat{\beta}$ in Robinson (1988), Bontemps et al. (2008) assess the structure of $m(\cdot)$ through individually testing the parametric or semiparametric null models against a fully nonparametric alternative for $m(\cdot)$. They concluded that an additive model is a reasonable specification, that is, $m^0(TMEAD, NITRO) = m_1(TMEAD) + m_2(NITRO)$. We could apply our test to infer model structures on $m(\cdot)$ in (10). However, the strong assumptions on the linearity in Q and the separability between Q and $(TMEAD, NITRO)$ might invalidate the testing results. So we probe the validity of these assumptions with nonparametric functional form tests already available. To alleviate the *curse of dimensionality* in the nonparametric tests, we restrict the discrete variables in Q, denoted by Q_d $(AGE, REPAIR, ROOMS, COUNTY, VACANT)$, to be linear and focus on exploring the nonlinearity arising from the continuous variables in Q, denoted by Q_c $(SIZE, POPULATION, INCOME)$, and $(TMEAD, NITRO)$. With the additivity test by Gozalo and Linton (2001), we are not able to reject the null of additivity between Q_c and $(TMEAD, NITRO)$, but we reject a further additive specification of each variable in Q_c. Employing the linearity test in an additive model by Chen et al. (2011), we reject the linear specification of Q_c, and variants of the linear specification which include quadratic terms and interaction terms. Guided by the testing results, we proceed to infer model structures of $m(\cdot)$ in

$$P_i = Q_{di}'\beta_1 + g(Q_{ci}) + m(TMEAD_i, NITRO_i) + \epsilon_i, \tag{11}$$

where β_1 are the coefficients of Q_d, and $g(\cdot)$ is a multivariate smooth function of Q_c. Notably, (11) nests (10) as a special case. Below, we focus on (11) to infer model structures on $m(\cdot)$ with our proposed test.

We estimate β_1 in (11) by $\hat{\beta}_1$ in Robinson (1988), with the bandwidth selected using a cross-validation criterion (CVLS). Because the elements of $\hat{\beta}_1$ are numerically similar to those in Tab. II of Bontemps et al. (2008), we omit their presentation here. Then we apply the two-step estimator in Wang and Yang (2009) on $P_i - Q_{di}'\hat{\beta}_1$ to estimate the additive functions $\hat{g}(Q_{ci})$ and $\hat{m}(TMEAD_i, NITRO_i)$. With a new dependent variable $\tilde{P}_i = P_i - Q_{di}'\hat{\beta}_1 - \hat{g}(Q_{ci})$, we apply our bootstrap tests $\hat{T}_c^*$ and $\hat{T}^*$ in the following nonparametric model:

$$\tilde{P}_i = m(TMEAD_i, NITRO_i) + \epsilon_i. \tag{12}$$

With $X = TMEAD$, $Z = NITRO$, we test the significant variables in $g_X(\cdot)$ with three null hypotheses: 1) H_{01} : *TMEAD* is linear and additively separable from *NITRO* $(W^r = (X, Z))$, 2) H_{02} : *TMEAD* is linear $(W^r = X)$, and 3) H_{03} : *TMEAD* is separable from *NITRO* $(W^r = Z)$. We perform 399 bootstrap and report p-values

[6]The data are directly from Bontemps et al. (2008); thus, we omit the data description and summary for brevity.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

TABLE 2. Tests for significant variables in gradients of $m(TMEAD, NITRO)$ in HPM

Panel A		p-value of $\hat{T}_c^*$ ($\hat{T}^*$) with $X = TMEAD$		
	i	1	2	3
	H_{0i}	*TMEAD* is linear and separable from *NITRO*	*TMEAD* is linear	*TMEAD* is separable from *NITRO*
	W^r	*TMEAD*, *NITRO*	*TMEAD*	*NITRO*
	W^c	—	*NITRO*	—
	$C_h = 0.5$	0.0000 (0.0000)	0.0000 (0.0000)	0.2950 (0.2725)
	$C_h = 1.0$	0.0000 (0.0000)	0.0200 (0.0275)	0.2175 (0.1975)
	$C_h = 1.5$	0.0000 (0.0000)	0.0325 (0.0375)	0.1475 (0.1625)
	CVLS	0.0000 (0.0000)	0.0025 (0.0050)	0.5175 (0.4950)
Panel B		p-value of $\hat{T}_c^*$ ($\hat{T}^*$) with $X = NITRO$		
	i	1	2	3
	H_{0i}	*NITRO* is linear and separable from *TMEAD*	*NITRO* is linear	*NITRO* is separable from *TMEAD*
	W^r	*TMEAD*, *NITRO*	*NITRO*	*TMEAD*
	W^c	—	*TMEAD*	*NITRO*
	$C_h = 0.5$	0.0275 (0.0350)	0.0000 (0.0000)	0.2600 (0.2425)
	$C_h = 1.0$	0.0500 (0.0600)	0.0000 (0.0000)	0.2200 (0.1900)
	$C_h = 1.5$	0.0550 (0.0650)	0.0100 (0.0150)	0.1625 (0.1750)
	CVLS	0.0225 (0.0325)	0.0000 (0.0000)	0.2950 (0.3100)

under H_{0i}, $i = 1, 2, 3$. Following the simulations, we select bandwidths with the *rule-of-thumb* (with $C_h = (0.5, 1.0, 1.5)$) and CVLS.[7]

We report the results in Panel A of Table 2. Across different choices of bandwidth, H_{01} is rejected at the 1% level, indicating that a linear and separable specification of *TMEAD* is inappropriate in modeling the HPM. Thus, *TMEAD* can be nonlinear, interactive with *NITRO*, or both. The tests of H_{02} and H_{03} reveal that *TMEAD* is significantly nonlinear at the 5% level but separable from *NITRO*. We perform a similar exercise to test the significance $X = NITRO$ in $g_X(\cdot)$ with H_{01-03}, reported in Panel B of Table 2. We reach the same conclusion regarding the nonlinearity and additivity of *NITRO* in (12).

[7]Notably, with Q_d modeled linearly, our testing results also reached the same conclusion regarding the inferred additive structure of $m(\cdot)$, whether we model Q_c linearly or additively through $\sum_{j=1}^{3} g_j(Q_{cj})$. See Appendix 3 in the online supplement for simulation results of our bootstrap test using bandwidths selected with CVLS in the PLM.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

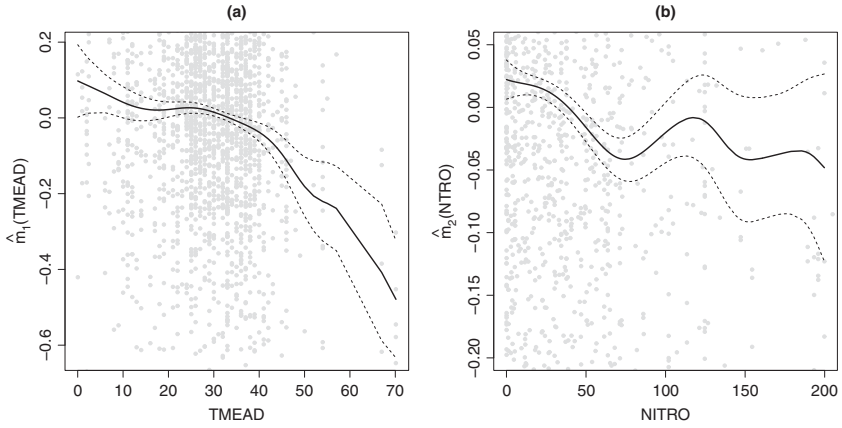


FIGURE 2. Estimates (solid lines) of inferred additive functions $\hat{m}_1(TMEAD)$ in (a) and $\hat{m}_2(NITRO)$ in (b) with their 95% confidence bands (dashed lines) and data points (gray circles).

Consistent with Bontemps et al. (2008), we infer a nonlinear additive structure in (12) as

$$\tilde{P}_i = m_1(TMEAD_i) + m_2(NITRO_i) + \epsilon_i.$$

To illustrate our point more vividly, we plot the centered estimates $\hat{m}_1(TMEAD)$ in panel (a) and $\hat{m}_2(NITRO)$ in panel (b) of Figure 2. The function is estimated with the marginal integration using a bandwidth selected through CVLS. We superimpose in the plot the 95% confidence bands in dashed lines constructed with the wild bootstrap method using 399 repetitions, and data points in gray circle of $\tilde{P} = P - Q_d'\hat{\beta}_1 - \hat{g}(Q_c)$ against *TMEAD* in (a) and *NITRO* in (b). To display the inferred structures clearly, we only plot the data points that are relatively close to the estimated functions. Clearly, both functions exhibit nonlinear effects on the house sales price after controlling for various housing characteristics in Q. To provide further evidence, we follow Li (2000) to estimate the interactive effect of *TMEAD* and *NITRO* through a partially linear additive model:

$$\tilde{P}_i = (TMEAD_i \times NITRO_i)\gamma + m_1(TMEAD_i) + m_2(NITRO_i) + \epsilon_i.$$

We estimate γ using the cubic B-spline estimator and select the number of evenly spaced interior knots through CVLS. We obtain an insignificant estimate $\hat{\gamma} = -0.000045$, and its standard error is 0.00014, lending further support of the inferred additive structure.

7. CONCLUSION

Given a set of significant regressors in a conditional mean function, we propose a variation-based test for significant variables in the gradient function. We

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

construct the test by estimating the gradient with a local polynomial estimator, obtain the asymptotic null distribution of the test, and demonstrate that the test is consistent. Based on the analysis of the asymptotic null distribution, we propose a wild bootstrap test, which is easy to implement. We demonstrate the validity of the bootstrap because its asymptotic distribution is the same as the asymptotic null, whether the null hypothesis is valid or not. The Monte Carlo studies illustrate encouraging finite sample properties of our test, with the size being close to the nominal level and the power approaching one quickly as the sample size increases. The test can be used to infer certain notable aspects of model structures, including linearity and additivity implied by some well-known semiparametric models like the additive, varying coefficient, and the PLM. An empirical application in the HPM reveals that pollution, from intensive agricultural activities, has a significantly nonlinear but noninteractive impact on housing prices.

SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit: http://doi.org/10.1017/S0266466620000407

APPENDIX

We state two Lemmas and provide proofs for Theorems 1 and 2 below. We relegate the proof of Lemma 1 and Theorem 3 to Appendix 2, and additional simulation results to Appendix 3, all collected in the online supplement.

LEMMA 1. *Uniformly for all $w \in \mathcal{W}$, the support of W, which is a compact subset of $\Re^d$,*

$$\hat{g}_X(w) - g_X(w) = \frac{1}{nh^{d+1}f(w)} \sum_{i=1}^{n} SK\left(\frac{W_i - w}{h}\right) \times \left(\sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(w + \lambda(W_i - w)) \left(\frac{W_i - w}{h}\right)^{\mathbf{k}} + \epsilon_i \right) (1 + o_p(1)).$$

In our proof, we have made repeated use of the following Lemma 2, which is the same as Thm. 1 in Yao and Martins-Filho (2015). Let $\{Q_i\}_{i=1}^n$ be a sequence of independent and identically distributed (i.i.d.) random variables and $\phi_n(Q_1, \ldots, Q_k)$ be a symmetric function with $k < n$. We call $\phi_n(Q_1, \ldots, Q_k)$ a kernel function that depends on n and a U-statistic u_n of degree k is defined as

$$u_n = \binom{n}{k}^{-1} \sum_{(n,k)} \phi_n(Q_{i_1}, \ldots, Q_{i_k}), \tag{A.1}$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

where $\sum_{(n,k)}$ denotes the sum over all subsets $1 \le i_1 < i_2 < \cdots < i_k \le n$ of $\{1,2,\ldots,n\}$. Now, let $\phi_{cn}(q_1,\ldots,q_c) = E(\phi_n(Q_1,\ldots,Q_c,Q_{c+1},\ldots,Q_k)|Q_1 = q_1, Q_2 = q_2,\ldots,Q_c = q_c)$, $\sigma_{cn}^2 = Var(\phi_{cn}(Q_1,\ldots,Q_c))$ and $\theta_n = E(\phi_n(Q_1,\ldots,Q_k))$. In addition, recursively define $h_n^{(1)}(q_1) = \phi_{1n}(q_1) - \theta_n,\ldots,h_n^{(c)}(q_1,\ldots,q_c) = \phi_{cn}(q_1,\ldots,q_c) - \sum_{j=1}^{c-1}\sum_{(c,j)} h_n^{(j)}(q_{i_1},\ldots,q_{i_j}) - \theta_n$ for $c = 2,\ldots,k$, where the sum $\sum_{(c,j)}$ is over all subsets $1 \le i_1 < \cdots < i_j \le c$ of $\{1,\ldots,c\}$. By Hoeffding's H-decomposition we have

$$u_n = \theta_n + \binom{n}{k}^{-1} \sum_{j=1}^{k} \binom{n-j}{k-j} \sum_{(n,j)} h_n^{(j)}(Q_{v_1},\ldots,Q_{v_j})$$

$$= \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(Q_{v_1},\ldots,Q_{v_j}),$$

where $H_n^{(j)}(Q_{v_1},\cdots,Q_{v_j}) = \binom{n}{j}^{-1}\sum_{(n,j)} h_n^{(j)}(Q_{v_1},\ldots,Q_{v_j})$. Since u_n can be written as a finite sum of $H_n^{(j)}$, its magnitude can be determined by studying $H_n^{(j)}$. The following result shows that the magnitude of $H_n^{(j)}$ is determined by n and the leading variance σ_{jn}^2 defined above.

LEMMA 2. *Let $\{Q_i\}_{i=1}^n$ be an i.i.d. sequence and u_n be defined as in equation (A.1) such that*

$$u_n = \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(Q_{v_1},\ldots,Q_{v_j}).$$

Then, (a) $Var(H_n^{(j)}) = O(n^{-j}\sum_{c=1}^{j}\sigma_{cn}^2) = O(n^{-j}\sigma_{jn}^2)$ *and* $H_n^{(j)} = O_p((n^{-j}\sigma_{jn}^2)^{1/2})$;
(b) *for* $1 \le c \le c' \le k$, *we have* $\frac{\sigma_{cn}^2}{c} \le \frac{\sigma_{c'n}^2}{c'}$.

Proof of Theorem 1. It is easy to see that

$$T_{2n} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_X(W_i) - g_X(W_i)\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}(\hat{g}_X(W_l^r;W_i^c) - g_X(W_l^r;W_i^c))\right]^2$$

$$- \frac{2}{n}\sum_{i=1}^{n}\left[\hat{g}_X(W_i) - g_X(W_i)\right]\left[\frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}(\hat{g}_X(W_l^r;W_i^c) - g_X(W_l^r;W_i^c))\right]$$

$$= T_{21} + T_{22} + T_{23}.$$

Lemma 1 shows that uniformly for all $W_i \in \mathcal{W}$, and for $j \neq i$,

$$\hat{g}_X(W_i) - g_X(W_i)$$

$$= \frac{1}{nh^{d+1}}\sum_{j=1}^{n}\hat{SK}\left(\frac{W_j - W_i}{h}\right)$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$\times\left(\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}+\epsilon_j\right)$$

$$=\frac{1}{nh^{d+1}f(W_i)}\sum_{j=1}^{n}SK\left(\frac{W_j-W_i}{h}\right)$$

$$\times\left(\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}+\epsilon_j\right)(1+o_p(1)). \quad \textbf{(A.2)}$$

We use equation (A.2) to perform the expansion and we show (1)–(3) below, which lead to the stated result (6) in the theorem.

(1) $nh^{2+\frac{d}{2}}\left(T_{21}-\frac{1}{nh^{2+d}}B_{1n}\right)\xrightarrow{d}\mathcal{N}(0,\Omega)$.

(2) $T_{23}=-\frac{2}{nh^{2+d-d_1}}B_{3n}+o_p\left(\left(nh^{2+\frac{d}{2}}\right)^{-1}\right)$.

(3) $T_{22}=\frac{1}{nh^{2+d-d_1}}B_{2n}+o_p\left(\left(nh^{2+\frac{d}{2}}\right)^{-1}\right)$.

Proof of (1): By the first equality in equation (A.2),

$$T_{21}=\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq j,\,i\neq t}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\frac{1}{h^{2d+2}}\hat{SK}\left(\frac{W_j-W_i}{h}\right)\hat{SK}\left(\frac{W_t-W_i}{h}\right)$$

$$\times\left[\epsilon_j\epsilon_t+\epsilon_j\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_t-W_i))\left(\frac{W_t-W_i}{h}\right)^{\mathbf{k}}\right.$$

$$+\epsilon_t\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}$$

$$+\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_t-W_i))\left(\frac{W_t-W_i}{h}\right)^{\mathbf{k}}$$

$$\left.\times\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}(D^{\mathbf{k}'}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}'}\right]$$

$$=T_{211n}+T_{212n}+T_{213n}+T_{214n}.$$

Below, we demonstrate that the asymptotic behavior of T_{21} above is totally determined by T_{211n}. Thus, the expressions based on the bias of estimation $(D^{\mathbf{k}}m)$ do not play a role asymptotically. Specifically,

$$nh^{2+\frac{d}{2}}\left(T_{211n}-\frac{1}{nh^{2+d}}B_{1n}\right)$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$= nh^{2+\frac{d}{2}} \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n} \sum_{j=1}^{n} \frac{\epsilon_i \epsilon_j}{h^{2+2d}}$$

$$\times \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t (1 + o_p(1))$$

$$\xrightarrow{d} \mathcal{N}(0, \Omega),$$

which determines the asymptotic distribution. We give the details below.

(a) $T_{211n} = \frac{1}{n^3} \sum_{\substack{i=1 \\ i\neq j, i\neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \frac{1}{h^{2d+2}} \hat{SK}\left(\frac{W_j - W_i}{h}\right) \hat{SK}\left(\frac{W_t - W_i}{h}\right) \epsilon_j \epsilon_t.$

(i) Consider the case that $i \neq j \neq t$. By the second equality in equation (A.2),

$$T_{211n} = \frac{1}{n^3} \sum_{\substack{i=1 \\ i\neq j\neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \frac{1}{f(W_i)^2 h^{2d+2}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right) \epsilon_j \epsilon_t (1 + o_p(1))$$

$$= \frac{1}{n^3} \sum_{\substack{i=1 \\ i\neq j\neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \psi_{nijt} (1 + o_p(1))$$

$$= T_{211} (1 + o_p(1)),$$

and notice ψ_{nijt} is symmetric in indices j and t.

Let $\phi_{nijt} = \psi_{nijt} + \psi_{njit} + \psi_{ntji}$, and it is symmetric in all of its indices, then

$$T_{211} = \frac{1}{3n^3} \sum_{\substack{i=1 \\ i\neq j\neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \phi_{nijt} = \frac{2}{n^3} \sum_{\substack{i=1 \\ i<j<t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \phi_{nijt}$$

$$= \frac{1}{3} \left[O(n^{-4}) + \binom{n}{3}^{-1} \right] \sum_{\substack{i=1 \\ i<j<t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \phi_{nijt}$$

$$= \frac{1}{3} \binom{n}{3}^{-1} \sum_{\substack{i=1 \\ i<j<t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \phi_{nijt} (1 + o_p(1)).$$

Note that it is a U-statistic. We apply Hoeffding's H-decomposition for the U-statistic with a sample size-dependent kernel in Lemma 2, for $\phi_n(Q_i, Q_j, Q_t) \equiv \phi_{nijt}$, to obtain

$$T_{211} = \frac{1}{3} \left[\theta_n + \sum_{j=1}^{3} \binom{3}{j} H_n^{(j)}(Q_{v_1}, \ldots, Q_{v_j}) \right] (1 + o_p(1)),$$

where

$$H_n^{(j)}(Q_{v_1}, \ldots, Q_{v_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(Q_{v_1}, \ldots, Q_{v_j}), \quad Q_i = (W_i', \epsilon_i)'.$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Here, $\theta_n = E(\phi_{nijt}) = 0$.

$$H_n^{(1)}(Q_{v_1}) = \binom{n}{1}^{-1} \sum_{i=1}^{n} h_n^{(1)}(Q_i) = \frac{1}{n}\sum_{i=1}^{n}[E(\phi_{nijt}|Q_i) - \theta_n] = 0,$$

since $E(\epsilon_i|W_i) = 0$.

$$\begin{aligned}
&\frac{1}{3}\binom{3}{2} H_n^{(2)}(Q_{v_1}, Q_{v_2}) \\
&\quad = \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} h_n^{(2)}(Q_i, Q_j) \\
&\quad = \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} [E(\phi_{nijt}|Q_i, Q_j) - h_n^{(1)}(Q_i) - h_n^{(1)}(Q_j) - \theta_n] \\
&\quad = \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} E(\psi_{nijt} + \psi_{njit} + \psi_{ntji}|Q_i, Q_j) = \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} E(\psi_{ntji}|Q_i, Q_j) \\
&\quad = \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \frac{\epsilon_i\epsilon_j}{h^{2d+2}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t.
\end{aligned}$$

$V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})) = O_p(n^{-3}\sigma_{3n}^2) = O_p(n^{-3}h^{-2d-4})$, because by A2, A3(1),

$$\begin{aligned}
\sigma_{3n}^2 &= E\phi_{nijt}^2 \le CE\psi_{nijt}^2 \\
&= C\frac{1}{h^{4d+4}} E\left[\frac{1}{f^4(W_i)} SK^2\left(\frac{W_j - W_i}{h}\right) SK^2\left(\frac{W_t - W_i}{h}\right) \sigma^2(W_j)\sigma^2(W_t)\right] \\
&= O_p(h^{-2d-4}).
\end{aligned}$$

So we obtain $H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}h^{-d-2}) = O_p(n^{-1}h^{-\frac{d}{2}-2}n^{-\frac{1}{2}}h^{-\frac{d}{2}}) = o_p((nh^{2+\frac{d}{2}})^{-1})$. So in all, for $i \neq j \neq t$,

$$\begin{aligned}
T_{211} &\equiv T_{211e} \\
&= \binom{n}{2}^{-1} \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \frac{\epsilon_i\epsilon_j}{h^{2d+2}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t (1 + o_p(1)).
\end{aligned}$$

(ii) $$T_{211e} = \binom{n}{2}^{-1} h^{-2-\frac{d}{2}} \times \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \underbrace{\frac{\epsilon_i\epsilon_j}{h^{\frac{3d}{2}}} \int SK\left(\frac{W_j - W_t}{h}\right) SK\left(\frac{W_i - W_t}{h}\right) \frac{1}{f(W_t)} dW_t}_{\phi_{nij}} (1 + o_p(1)).$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Note that $\phi_{nij} = \phi_n(Q_i, Q_j)$ is symmetric in Q_i and Q_j, and $E(\phi_{nij}|Q_i) = 0$. By Theorem 1 in Hall (1984), if $E\phi_{nij}^2 < \infty$, and $(EG_n^2(Q_i, Q_j) + (1/n)E\phi_{nij}^4)/([E\phi_{nij}^2]^2) \to 0$, then

$$nh^{2+\frac{d}{2}} T_{211e} = n\binom{n}{2}^{-1} \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i<j}}^{n} \phi_{nij} \xrightarrow{d} \mathcal{N}(0, 2E\phi_{nij}^2).$$

Recall the definition of $SK(\cdot)$ in equation (4), and let $(W_i - W_t)/h = \Psi$, $(W_j - W_i)/h = \Psi_1$, then $(W_j - W_t)/h = \Psi_1 + \Psi$, $W_t = W_i - h\Psi$, and $W_j = W_i + h\Psi_1$. Then

$$\begin{aligned}
&2E\phi_{nij}^2 \\
&\quad = 2E\left[\sigma^2(W_i)\sigma^2(W_j)\frac{1}{h^d}\left(\frac{1}{h^d}\int SK\left(\frac{W_j - W_t}{h}\right)SK\left(\frac{W_i - W_t}{h}\right)\frac{1}{f(W_t)}dW_t\right)^2\right] \\
&\quad = 2\int f(W_i)f(W_i + h\Psi_1)\sigma^2(W_i)\sigma^2(W_i + h\Psi_1)\left[\int \sum_{0\le|\mathbf{j}|\le p}\sum_{l=1}^{N_{|\mathbf{j}|}}[S^{-1}]_{1+d,\ \sum_{l'=0}^{|\mathbf{j}|-1} N_{l'}+l}\right. \\
&\qquad \times K(\Psi_1 + \Psi)(\Psi_1 + \Psi)^{G_{|\mathbf{j}|}(l)} \sum_{0\le|\mathbf{j}'|\le p}\sum_{\tau=1}^{N_{|\mathbf{j}'|}}[S^{-1}]_{1+d,\ \sum_{\tau'=0}^{|\mathbf{j}'|-1} N_{\tau'}+\tau} K(\Psi)\Psi^{G_{|\mathbf{j}'|}(\tau)} \\
&\qquad \left.\times\frac{1}{f(W_i - h\Psi)}d\Psi\right]^2 dW_i d\Psi_1 \\
&\quad \to 2\int \sigma^4(W_i)dW_i\left[\int\left(\int SK(\Psi_1 + \Psi)SK(\Psi)d\Psi\right)^2 d\Psi_1\right] = \Omega.
\end{aligned}$$

$E\phi_{nij}^4 = h^{-d}\int E(\epsilon_i^4|W_i)E(\epsilon_j^4|W_i + h\Psi_1)[\int SK(\Psi_1 + \Psi)SK(\Psi)\frac{1}{f(W_i - h\Psi)}d\Psi]^4 f(W_i)f(W_i + h\Psi_1)dW_i d\Psi_1$, thus, with A1, A2, and A3(1), $\frac{1}{n}E\phi_{nij}^4 \to 0$ as $nh^d \to \infty$. With a similar change of variables, that is, $\Psi = \frac{W_i - W_t}{h}$, $\Psi_1 = \frac{W_j - W_t'}{h}$, $\Psi_2 = \frac{W_l - W_i}{h}$, $\Psi_3 = \frac{W_i - W_j}{h}$, we have

$$\begin{aligned}
&EG_n^2(Q_i, Q_j) \\
&\quad = E[E^2(\phi_{nli}\phi_{nlj}|Q_i, Q_j)] \\
&\quad = h^{-2d}E\left\{\epsilon_i^2\epsilon_j^2\left[\int \sigma^2(W_l)\left(\frac{1}{h^d}\int SK\left(\frac{W_i - W_t}{h}\right)SK\left(\frac{W_l - W_t}{h}\right)\frac{1}{f(W_t)}dW_t\right)\right.\right. \\
&\qquad \left.\left.\times\left(\frac{1}{h^d}\int SK\left(\frac{W_j - W_t'}{h}\right)SK\left(\frac{W_l - W_t'}{h}\right)\frac{1}{f(W_t')}dW_t'\right)f(W_l)dW_l\right]^2\right\} \\
&\quad = h^d\int \sigma^2(W_i)\sigma^2(W_i - h\psi_3)f(W_i)f(W_i - h\psi_3) \\
&\qquad \times\left[\int \sigma^2(W_i + h\psi_2)\left(\int SK(\Psi)SK(\Psi_2 + \Psi)\frac{1}{f(W_i - h\Psi)}d\Psi\right]\right. \\
&\qquad \left.\times\left(\int SK(\Psi_1)SK(\Psi_2 + \Psi_1 + \Psi_3)\frac{1}{f(W_i - h\Psi_3 - h\Psi_1)}d\Psi_1\right)\right.
\end{aligned}$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$\times f(W_i + h\Psi_2) d\,\Psi_2]^2 d\Psi_3 dW_i$$
$$= O(h^d),$$

so the conditions for the asymptotic normality are verified and $nh^{2+\frac{d}{2}} T_{211e} \xrightarrow{d} \mathcal{N}(0, \Omega)$.

(iii) Since $i \neq j$ and $i \neq t$, the only other case is for $i \neq j = t$. In this case,

$$T_{211n} \equiv T_{211B} = \frac{1}{nh^{2+d}} \frac{1}{n^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \frac{\epsilon_j^2}{h^d} \hat{SK}^2 \left(\frac{W_j - W_i}{h} \right) = \frac{1}{nh^{2+d}} B_{1n}.$$

Combining above (i)–(iii), we obtain that $nh^{2+\frac{d}{2}} (T_{211n} - \frac{1}{nh^{2+d}} B_{1n}) \xrightarrow{d} \mathcal{N}(0, \Omega)$. (b) By the second equality in equation (A.2),

$$\begin{aligned} T_{212n} &= \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq j, i \neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \frac{1}{h^{2d+2}} \hat{SK} \left(\frac{W_j - W_i}{h} \right) \hat{SK} \left(\frac{W_t - W_i}{h} \right) \epsilon_j \\ &\quad \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h} \right)^{\mathbf{k}} \\ &= \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq j, i \neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \frac{1}{f^2(W_i) h^{2d+2}} SK \left(\frac{W_j - W_i}{h} \right) SK \left(\frac{W_t - W_i}{h} \right) \epsilon_j \\ &\quad \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_i + \lambda(W_t - W_i)) \left(\frac{W_t - W_i}{h} \right)^{\mathbf{k}} \\ &\quad \times (1 + o_p(1)) \\ &= \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq j, i \neq t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \psi_{nijt} (1 + o_p(1)) = T_{212} (1 + o_p(1)). \end{aligned}$$

(i) Consider the case $i \neq j \neq t$. With $\phi_{nijt} = \psi_{nijt} + \psi_{nitj} + \psi_{njit} + \psi_{njti} + \psi_{ntij} + \psi_{ntji}$, we apply Lemma 2 to obtain

$$\begin{aligned} T_{212} &= \frac{1}{6} \begin{pmatrix} n \\ 3 \end{pmatrix}^{-1} \sum_{\substack{i=1 \\ i<j<t}}^{n} \sum_{j=1}^{n} \sum_{t=1}^{n} \phi_{nijt} (1 + o_p(1)) \\ &= \frac{1}{6} \left[\theta_n + \sum_{j=1}^{3} \begin{pmatrix} 3 \\ j \end{pmatrix} H_n^{(j)} (Q_{v_1}, \ldots, Q_{v_j}) \right] (1 + o_p(1)). \end{aligned}$$

Here $\theta_n = 0$ as $E(\epsilon_j | W_j) = 0$. Because $(D^{\mathbf{k}} m)(W)$ is bounded and uniformly continuous for $|\mathbf{k}| = p + 1$ as in A3(2),

$$H_n^{(1)} (Q_{v_1}) = n^{-1} \sum_{i=1}^{n} E(\phi_{nijt} | Q_i)$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{2\epsilon_i}{h^{2d+2}}\int \frac{1}{f^2(W_j)} SK\left(\frac{W_i - W_j}{h}\right) SK\left(\frac{W_t - W_j}{h}\right)$$

$$\times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_j + \lambda(W_t - W_j))\left(\frac{W_t - W_j}{h}\right)^{\mathbf{k}}$$

$$\times f(W_j)f(W_t)dW_j dW_t$$

$$= O_p(n^{-\frac{1}{2}}h^{p-1}).$$

Given that $\sigma_{2n}^2 = E(\phi_{2n}^2(Q_i, Q_j))$, $\phi_{2n}(Q_i, Q_j) = E(\phi_{nijt}|Q_i, Q_j) = E(\psi_{nijt} + \psi_{njit} + \psi_{ntij} + \psi_{ntji}|Q_i, Q_j)$,

$$E(\psi_{nijt}|Q_i, Q_j) = \frac{\epsilon_j}{f^2(W_i)h^{2d+2}}\int SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right)$$

$$\times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i + \lambda(W_t - W_i))\left(\frac{W_t - W_i}{h}\right)^{\mathbf{k}} f(W_t)dW_t.$$

We obtain $E[E^2(\psi_{nijt}|Q_i, Q_j)] = O(h^{2p+2-4-d}) = O(h^{2p-d-2})$. Since similar arguments apply to the other terms, we have $\sigma_{2n}^2 = O(h^{2p-d-2})$. Since $V(H_n^{(2)}(Q_{v_1}, Q_{v_2})) = O(n^{-2}\sigma_{2n}^2)$, we conclude that $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}h^{p-\frac{d}{2}-1}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$. $\sigma_{3n}^2 = E(\phi_{nijt}^2) = O(h^{2(p-d)-2})$, $V(H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-3}\sigma_{3n}^2)$, and $H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3})$ $= O_p(n^{-\frac{3}{2}}h^{(p-d)-1}) = O_p((n^{-1}h^{-\frac{d}{2}-2})(n^{-\frac{1}{2}}h^{(p-\frac{d}{2})+1})) = o_p(n^{-1}h^{-\frac{d}{2}-2})$ since $nh^d \to \infty$. So $T_{212} = O_p(n^{-\frac{1}{2}}h^{p-1}) + o_p(n^{-1}h^{-\frac{d}{2}-2}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$ with $nh^{2p+d+2} \to 0$ as in A4(2).

(ii) Consider the case $i \neq j = t$.

$$T_{212} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\frac{1}{f^2(W_i)h^{2d+2}} SK^2\left(\frac{W_j - W_i}{h}\right)\epsilon_j$$

$$\times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i))\left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}}.$$

We again apply Lemma 2 to obtain with similar arguments that $T_{212} = O_p(n^{-2}h^{p-\frac{3}{2}d-1}) +$ $O_p(n^{-\frac{3}{2}}h^{p-d-1}) = O_p(n^{-\frac{3}{2}}h^{p-d-1}) = O_p(n^{-1}h^{-\frac{d}{2}-2}n^{-\frac{1}{2}}h^{p-\frac{d}{2}+1}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$.

(i)-(ii) above give the results that $T_{212} = o_p(n^{-1}h^{-\frac{d}{2}-2})$. So, $T_{212n} = o_p(n^{-1}h^{-\frac{d}{2}-2})$.

(c) $T_{213n} = o_p(n^{-1}h^{-\frac{d}{2}-2})$ follows from the same arguments in (b) for T_{212n}.

(d) By the second equality in equation (A.2),

$$T_{214n} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j,\, i\neq t}}^{n}\sum_{t=1}^{n}\frac{1}{f^2(W_i)h^{2d+2}} SK\left(\frac{W_j - W_i}{h}\right) SK\left(\frac{W_t - W_i}{h}\right)$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$\times \sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}$$
$$\times \sum_{|\mathbf{k}'|=p+1} \frac{h^{p+1}}{\mathbf{k}'!}(D^{\mathbf{k}'}m)(W_i+\lambda(W_t-W_i))\left(\frac{W_t-W_i}{h}\right)^{\mathbf{k}'}(1+o_p(1))$$
$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq j,\, i\neq t}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\psi_{nijt}(1+o_p(1)) = T_{214}(1+o_p(1)).$$

(i) Consider the case $i \neq j \neq t$. Note that ψ_{nijt} is symmetric in j,t. So we let $\phi_{nijt} = \psi_{nijt} + \psi_{njit} + \psi_{ntji}$ and apply Lemma 2 to have

$$T_{214} = \frac{1}{3}\binom{n}{3}^{-1}\sum_{\substack{i=1\\ i\neq j\neq t}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\phi_{nijt}(1+o_p(1))$$
$$= (1+o_p(1))\frac{1}{3}\left[\theta_n + \sum_{j=1}^{3}\binom{3}{j}H_n^{(j)}(Q_{v_1},\ldots,Q_{v_j})\right].$$

Recall the definition of $SK(\cdot)$ in equation (4),

$$\frac{1}{3}\theta_n = E\psi_{nijt}$$
$$= h^{2p}\sum_{0\leq|\mathbf{j}|\leq p}\sum_{l=1}^{N_{|\mathbf{j}|}}\sum_{0\leq|\mathbf{j}'|\leq p}\sum_{\tau=1}^{N_{|\mathbf{j}'|}}[S^{-1}]_{1+d,\,\sum_{l'=0}^{|\mathbf{j}|-1}N_{l'}+l}[S^{-1}]_{1+d,\,\sum_{\tau'=0}^{|\mathbf{j}'|-1}N_{\tau'}+\tau}\int\frac{1}{f^2(W_i)}K(\Psi_1)$$
$$\times\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda h\Psi_1)(\Psi_1)^{\mathbf{k}+G_{|\mathbf{j}|}(l)}K(\Psi_2)$$
$$\times\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}(D^{\mathbf{k}'}m)(W_i+\lambda h\Psi_2)(\Psi_2)^{\mathbf{k}'+G_{|\mathbf{j}'|}(\tau)}$$
$$\times f(W_i)f(W_i+h\Psi_1)f(W_i+h\Psi_2)dW_i d\Psi_1 d\Psi_2$$
$$= O(h^{2p}) = o(n^{-1}h^{-\frac{d}{2}-2}) \text{ since } nh^{\frac{d}{2}+2}h^{2p}\to 0 \text{ in A4(2).}$$

$H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}}h^{2p}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$ as $V(H_n^{(1)}(Q_{v_1})) = O(n^{-1}\sigma_{1n}^2)$, and $\sigma_{1n}^2 = O(h^{4p})$. $H_n^{(2)}(Q_{v_1},Q_{v_2}) = O_p(n^{-1}h^{2p-\frac{d}{2}}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$ as $V(H_n^{(2)}(Q_{v_1},Q_{v_2})) = O(n^{-2}\sigma_{2n}^2)$, and $\sigma_{2n}^2 = O(h^{4p-d})$. $H_n^{(3)}(Q_{v_1},Q_{v_2},Q_{v_3}) = O_p(n^{-\frac{3}{2}}h^{2p-d}) = o_p(n^{-1}h^{-\frac{d}{2}-2})$ as $V(H_n^{(3)}(Q_{v_1},Q_{v_2},Q_{v_3})) = O(n^{-3}\sigma_{3n}^2)$, and $\sigma_{3n}^2 = O(h^{4p-2d})$. So when $i\neq j\neq t$, $T_{214} = o_p(n^{-1}h^{-\frac{d}{2}-2})$.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

(ii) Consider the case $i \neq j = t$.

$$T_{214} = \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \frac{1}{f^2(W_i) h^{2d+2}} SK^2 \left(\frac{W_j - W_i}{h} \right) \times \left(\sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_i + \lambda(W_j - W_i)) \left(\frac{W_j - W_i}{h} \right)^{\mathbf{k}} \right)^2.$$

Since $E|T_{214}| = O(n^{-1} h^{2(p+1)-d-2}) = O(n^{-1} h^{2p-d}) = O_p(n^{-1} h^{-\frac{d}{2}-2} h^{2p-\frac{d}{2}+2}) = o_p(n^{-1} h^{-\frac{d}{2}-2})$, where we need $p+1 > \frac{d}{4}$ in A4(2), we obtain that $T_{214} = o_p(n^{-1} h^{-\frac{d}{2}-2})$.

In all, $T_{214} = o_p(n^{-1} h^{-\frac{d}{2}-2})$. So, $T_{214n} = o_p(n^{-1} h^{-\frac{d}{2}-2})$. Combining results in (1)(a)-(d), we obtain the claim in (1).

Proof of (2):

$$\begin{aligned}
&\hat{g}_X(W_l^r; W_i^c) - g_X(W_l^r; W_i^c) \\
&= \frac{1}{nh^{d+1}} \sum_{\substack{j=1 \\ j \neq l, j \neq i}}^{n} \hat{SK} \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) \\
&\times \left[\sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_l^r + \lambda(W_j^r - W_l^r); W_i^c + \lambda(W_j^c - W_i^c)) \right. \\
&\left. \times \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right)^{\mathbf{k}} + \epsilon_j \right] \\
&= \frac{1}{f(W_l^r; W_i^c) nh^{d+1}} \sum_{\substack{j=1 \\ j \neq l, j \neq i}}^{n} SK \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) \\
&\times \left[\sum_{|\mathbf{k}|=p+1} \frac{h^{p+1}}{\mathbf{k}!} (D^{\mathbf{k}} m)(W_l^r + \lambda(W_j^r - W_l^r); W_i^c + \lambda(W_j^c - W_i^c)) \right. \\
&\left. \times \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right)^{\mathbf{k}} + \epsilon_j \right] (1 + o_p(1)), \qquad \textbf{(A.3)}
\end{aligned}$$

where we follow equations (4) and (5) to define

$$\begin{aligned}
SK \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) &= \sum_{0 \leq |\mathbf{k}| \leq p} \sum_{m=1}^{N_{|\mathbf{k}|}} [S^{-1}]_{1+d, \sum_{m'=0}^{|\mathbf{k}|-1} N_{m'} + m} K \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right) \\
&\times \left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h} \right)^{G_{|\mathbf{k}|}(m)}, \qquad \textbf{(A.4)}
\end{aligned}$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$\hat{SK}\left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h}\right) = \sum_{0 \le |\mathbf{k}| \le pm=1}^{N_{|\mathbf{k}|}} [S_n(W_l^r; W_i^c)^{-1}]_{1+d, \sum_{m'=0}^{|\mathbf{k}|-1} N_{m'}+m} \times K\left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h}\right)\left(\frac{W_j^r - W_l^r}{h}; \frac{W_j^c - W_i^c}{h}\right)^{G_{|\mathbf{k}|}(m)}. \tag{A.5}$$

Here we write $(W_j - W_i/h) \equiv (W_j^r - W_i^r/h; W_j^c - W_i^c/h)$ to denote an equivalent rearrangement, highlighting the variables W^r in W to be tested for significance. With equations (4), (5), (A.2), (A.3), (A.4) and (A.5), and the fact that $\frac{1}{n-1} - \frac{1}{n} = O(n^{-2})$, we have

$$\begin{aligned}
T_{23} = &-\frac{2}{n}\sum_{i=1}^{n}\left[\frac{1}{nh^{d+1}}\sum_{\substack{j=1\\ j\neq i}}^{n}\hat{SK}\left(\frac{W_j - W_i}{h}\right)\right. \\
&\left.\times\left(\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i))\left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}} + \epsilon_j\right)\right] \\
&\times\frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}\frac{1}{nh^{d+1}}\sum_{\substack{t=1\\ t\neq l, t\neq i}}^{n}\hat{SK}\left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h}\right) \\
&\times\left[\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_l^r + \lambda(W_t^r - W_l^r); W_i^c + \lambda(W_t^c - W_i^c))\right. \\
&\left.\times\left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h}\right)^{\mathbf{k}} + \epsilon_t\right] \\
= &-\frac{2}{n^3(n-1)}\sum_{\substack{i=1\\ i\neq j, t\neq l, t\neq i, l\neq i}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h^{2d+2}}\hat{SK}\left(\frac{W_j - W_i}{h}\right)\hat{SK}\left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h}\right) \\
&\times\left[\epsilon_j\epsilon_t + \epsilon_t\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i + \lambda(W_j - W_i))\left(\frac{W_j - W_i}{h}\right)^{\mathbf{k}}\right. \\
&+\epsilon_j\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_l^r + \lambda(W_t^r - W_l^r); W_i^c + \lambda(W_t^c - W_i^c)) \\
&\left.\times\left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h}\right)^{\mathbf{k}}\right.
\end{aligned}$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$+\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}$$
$$\times\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}(D^{\mathbf{k}'}m)(W_l^r+\lambda(W_t^r-W_l^r);W_i^c+\lambda(W_t^c-W_i^c))$$
$$\times\left(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h}\right)^{\mathbf{k}'}\Bigg]$$
$$=T_{231n}+T_{232n}+T_{233n}+T_{234n}.$$

In the online supplement, we show that (a) $T_{231n}=-2/(nh^{2+d-d_1})B_{3n}+o_p(n^{-1}h^{-\frac{d}{2}-2})$. (b) $T_{232n}=o_p(n^{-1}h^{-\frac{d}{2}-2})$. (c) $T_{233n}=o_p(n^{-1}h^{-\frac{d}{2}-2})$. (d) $T_{234n}=o_p(n^{-1}h^{-\frac{d}{2}-2})$. We omit the details here to save space. The claim in (2) follows from (a) to (d) above.

Proof of (3): We omit the details here as the argument is similar to (2). Details are provided in the online supplement.

Theorem 1 is established based on (1)–(3). □

Proof of Theorem 2. (i) $\hat{T}_c=(nh^{2+\frac{d}{2}}\hat{T}-h^{-\frac{d}{2}}\hat{B}_{1n}-h^{d_1-\frac{d}{2}}(\hat{B}_{2n}-2\hat{B}_{3n}))/\sqrt{\hat{\Omega}}$ in equation (7). We observe from Corollary 1 that $h^{-\frac{d}{2}}\hat{B}_{1n}=h^{-\frac{d}{2}}B_{1n}+o_p(1)$. Since $(nh^{2+\frac{d}{2}})^{-1}h^{-\frac{d}{2}}B_{1n}=n^{-1}h^{-2-d}B_{1n}\xrightarrow{p}0$ for our choice of bandwidth h when $p>1$, $(nh^{2+\frac{d}{2}})^{-1}h^{-\frac{d}{2}}\hat{B}_{1n}=o_p(1)$. Similarly, from Corollary 1, $h^{d_1-\frac{d}{2}}\hat{B}_{jn}=h^{d_1-\frac{d}{2}}B_{jn}+o_p(1)$ for $j=2,3$, and $\hat{\Omega}=\Omega+o_p(1)$. We have $(nh^{2+\frac{d}{2}})^{-1}h^{d_1-\frac{d}{2}}\hat{B}_{jn}=o_p(1)$ for $j=2,3$. For the claim in (i), we only need to show that $\hat{T}=E[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]^2+o_p(1)$ under the alternative H_A that $g_X(W^r;W^c)$ depends on W^r. Then $(nh^{2+\frac{d}{2}})^{-1}\hat{T}_c\xrightarrow{p}E[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]^2/\sqrt{\Omega}$. Thus, as $n\to\infty$, $P(\hat{T}_c>c_n)\to 1$ for any nonstochastic, positive sequence $c_n=o(nh^{2+\frac{d}{2}})$.

Given that $\hat{T}=T_{1n}+T_{2n}+T_{3n}$, and Theorem 1 implies that $T_{2n}=o_p(1)$, we show below that

(1) $T_{1n}=E[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]^2+o_p(1)$.
(2) $T_{3n}=o_p(1)$.

Proof of (1):

$$T_{1n}=\frac{1}{n}\sum_{i=1}^{n}g_X^2(W_i^r;W_i^c)-\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}g_X(W_i^r;W_i^c)g_X(W_l^r;W_i^c)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{\substack{l=1\\ i\neq l}}^{n}g_X(W_l^r;W_i^c)\right]^2$$
$$=T_{11}+T_{12}+T_{13}.$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

(a) T_{11} : with $Eg_X^2(W_i^r;W_i^c) < C$ as implied by A3(2) and A1(2), we easily have $T_{11} \xrightarrow{p} Eg_X^2(W_i^r;W_i^c)$.

$$\text{(b)} \quad T_{12} = -\frac{1}{n(n-1)}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}[g_X(W_i^r;W_i^c)g_X(W_l^r;W_i^c)+g_X(W_l^r;W_l^c)g_X(W_i^r;W_l^c)]$$

$$= -\frac{1}{n(n-1)}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\phi_{il}.$$

Since $i \neq l$, we apply Lemma 2 to obtain

$$T_{12} = -\frac{2}{n(n-1)}\sum_{\substack{i=1\\ i<l}}^{n}\sum_{l=1}^{n}\phi_{il}$$

$$= -\left\{\frac{2}{n}\sum_{i=1}^{n}[E(\phi_{il}|W_i)-E\phi_{il}]+E\phi_{il}+O_p(n^{-1}(E\phi_{il}^2)^{\frac{1}{2}})\right\}.$$

Since $E\phi_{il}^2 \le CEg_X^2(W_i^r;W_i^c)g_X^2(W_l^r;W_i^c) \le C[Eg_X^4(W_i^r;W_i^c)]^{\frac{1}{2}}[Eg_X^4(W_l^r;W_i^c)]^{\frac{1}{2}} < C$ by the Cauchy–Schwarz inequality, A3(2) and A1(2), $O_p(n^{-1}(E\phi_{il}^2)^{\frac{1}{2}}) = O_p(n^{-1})$. Furthermore, $\frac{2}{n}\sum_{i=1}^{n}[E(\phi_{il}|W_i)-E\phi_{il}] = O_p(n^{-\frac{1}{2}})$, since $E[E(\phi_{il}|W_i)-E\phi_{il}]^2 \le CE\phi_{il}^2 = O(1)$. Clearly, $E\phi_{il} = 2Eg_X(W_i^r;W_i^c)g_X(W_l^r;W_i^c)$, thus we obtain $T_{12} = -2Eg_X(W_i^r;W_i^c)g_X(W_l^r;W_i^c)+o_p(1)$.

(c) $T_{13} = \frac{1}{n(n-1)^2}\sum_{\substack{i=1\\ i\neq l, i\neq j}}^{n}\sum_{l=1}^{n}\sum_{j=1}^{n}g_X(W_l^r;W_i^c)g_X(W_j^r;W_i^c)$.

(I) Consider the case $i \neq j \neq l$. We apply Lemma 2 to write it as a U-statistic, perform Hoeffding's H-decomposition, analyze each term and obtain $T_{13} = Eg_X(W_l^r;W_i^c)g_X(W_j^r;W_i^c)+o_p(1)$.

(II) Consider the case $j = l \neq i$. $T_{13} = O_p(n^{-1})$ since $E|T_{13}| = O(n^{-1})$. So in all, we obtain $T_{13} = Eg_X(W_l^r;W_i^c)g_X(W_j^r;W_i^c)+o_p(1)$.

Combining (1)(a)-(c) above, we obtain

$$T_{1n} = Eg_X^2(W_i^r;W_i^c)-2Eg_X(W_i^r;W_i^c)g_X(W_l^r;W_i^c)+Eg_X(W_l^r;W_i^c)g_X(W_j^r;W_i^c)+o_p(1)$$

$$= E[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]^2+o_p(1).$$

Proof of (2):

$$T_{3n} = \frac{2}{n}\sum_{i=1}^{n}\left[g_X(W_i)-\frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}g_X(W_l^r;W_i^c)\right][\hat{g}_X(W_i)-g_X(W_i)]$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$-\frac{2}{n}\sum_{i=1}^{n}\left[g_X(W_i)-\frac{1}{n-1}\sum_{\substack{l=1\\l\neq i}}^{n}g_X(W_l^r;W_i^c)\right]\frac{1}{n-1}\sum_{\substack{l=1\\i\neq l}}^{n}[\hat{g}_X(W_l^r;W_i^c)-g_X(W_l^r;W_i^c)]$$
$$=T_{31}+T_{32}.$$

Given that

$$\frac{1}{n-1}\sum_{\substack{l=1\\l\neq i}}^{n}g_X(W_l^r;W_i^c)=E_{W_l^r}(g_X(W_l^r;W_i^c))+o_p(1)$$

uniformly for W_i^c, we denote

$$D^{\mathbf{k}}m_{ji}\equiv(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))\left(\frac{W_j-W_i}{h}\right)^{\mathbf{k}}.$$

Then

(*a*) T_{31}

$$=\frac{2}{n}\sum_{i=1}^{n}[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]\left[\frac{1}{f(W_i)nh^{d+1}}\sum_{\substack{j=1\\j\neq i}}^{n}SK\left(\frac{W_j-W_i}{h}\right)\right.$$
$$\left.\left(\epsilon_j+\sum_{0\leq|\mathbf{k}|\leq p}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{ji}\right)\right](1+o_p(1))$$
$$=(T_{311}+T_{312})(1+o_p(1)).$$

(I) $$T_{311}=\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}[g_X(W_i)-E_{W_l^r}(g_X(W_l^r;W_i^c))]\frac{1}{f(W_i)h^{d+1}}SK\left(\frac{W_j-W_i}{h}\right)\epsilon_j$$
$$=\frac{2}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\psi_{nij}.$$

We apply Lemma 2 to write $\phi_{nij}=\psi_{nij}+\psi_{nji}$, and

$$T_{311}=\binom{n}{2}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_{nij}(1+o_p(1))=\theta_n+\sum_{j=1}^{2}\binom{2}{j}H_n^{(j)}(Q_{v_1},\ldots,Q_{v_j}).$$

$\theta_n=0$, $H_n^{(1)}(Q_{v_1})=O_p(n^{-\frac{1}{2}}\sigma_{1n})=O_p(n^{-\frac{1}{2}}h^{-1})$ since $\sigma_{1n}^2=O(h^{-2})$, $H_n^{(2)}(Q_{v_1},Q_{v_2})=O_p(n^{-1}\sigma_{2n})=O_p(n^{-1}h^{-1-\frac{d}{2}})$ since $\sigma_{2n}^2=O(h^{-2-d})$. Thus, $T_{311}=O_p(n^{-\frac{1}{2}}h^{-1})$.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$\text{(II)}\quad T_{312} = \frac{2}{n^2}\sum_{\substack{i=1}}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}[g_X(W_i) - E_{W_l^r}(g_X(W_l^r; W_i^c))]\frac{1}{f(W_i)h^{d+1}} SK\left(\frac{W_j - W_i}{h}\right)$$
$$\times \sum_{0\le|\mathbf{k}|\le p}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{ji}(1+o_p(1)) = O_p(h^p),$$

since $E|T_{312}| = O(h^p)$. So we obtain $T_{31} = O_p(n^{-\frac{1}{2}}h^{-1}) + O_p(h^p) = o_p(1)$.

(b) With similar arguments, we show that

$$T_{32} = -\frac{2}{n}\sum_{i=1}^{n}[g_X(W_i) - E_{W_l^r}(g_X(W_l^r; W_i^c))]\frac{1}{n-1}\sum_{\substack{l=1\\ i\neq l}}^{n}\frac{1}{f(W_l^r; W_i^c)nh^{d+1}}$$
$$\times . \sum_{\substack{t=1\\ t\neq l, t\neq i}}^{n} SK\left(\frac{W_t^r - W_l^r}{h}; \frac{W_t^c - W_i^c}{h}\right)$$
$$\times \left[\epsilon_t + \sum_{0\le|\mathbf{k}|\le p}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{tl;ti}\right](1+o_p(1))$$
$$= o_p(1),$$

where $D^{\mathbf{k}}m_{tl;ti} = (D^{\mathbf{k}}m)(W_l^r + \lambda(W_t^r - W_l^r); W_i^c + \lambda(W_t^c - W_i^c))((W_t^r - W_l^r)/h; (W_t^c - W_i^c)/h)^{\mathbf{k}}$. The claim in (2) follows from (2)(a)–(b) above. Thus, we obtain $T_{3n} = o_p(1)$ and Theorem 2(i) is established based on (1) and (2).

(ii) The analysis of $\hat{B}_{jn}$ for $j = 1, 2,$ and 3 and $\hat{\Omega}$ are the same as in Corollary 1. By equation (7) and Corollary 1,

$$\hat{T}_c = \frac{(1+o_p(1))}{\sqrt{\Omega}}\left[nh^{2+\frac{d}{2}}\hat{T} - h^{-\frac{d}{2}}B_{1n} - h^{d_1-\frac{d}{2}}(B_{2n} - 2B_{3n}) + o_p(1)\right].$$

$\hat{T} = T_{1n} + T_{2n} + T_{3n}$. Since the analysis of T_{2n} is the same as in Theorem 1, we only need to show that under $H_A(\delta_n)$,

(1) $T_{1n} = \delta_n^2 E[D(W_i^r; W_i^c) - E_{W_l^r}(D(W_l^r; W_i^c))]^2(1+o_p(1))$.

(2) $T_{3n} = O_p(\delta_n(n^{-\frac{1}{2}}h^{-1} + h^p))$.

With $\delta_n = n^{-\frac{1}{2}}h^{-1-\frac{d}{4}}$, $nh^{2+\frac{d}{2}}T_{1n} \xrightarrow{p} E[D(W_i^r; W_i^c) - E_{W_l^r}(D(W_l^r; W_i^c))]^2$. With assumption A4(2), $T_{3n} = o_p(\delta_n^2)$; thus, $nh^{2+\frac{d}{2}}T_{3n} \xrightarrow{p} 0$. With Theorem 1 and Corollary 1, we obtain the claimed result in (ii).

Proof of (1): given $H_A(\delta_n)$, with $ED^4(W_i^r; W_j^c) < C$, we can easily show that

$$T_{1n} = \frac{1}{n}\sum_{i=1}^{n}\left[g_X^0(W_i^c) + \delta_n D(W_i) - \frac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^{n}(g_X^0(W_i^c) + \delta_n D(W_l^r; W_i^c))\right]^2$$

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

$$= \delta_n^2 \left[\frac{1}{n}\sum_{i=1}^{n} D^2(W_i) - \frac{2}{n(n-1)} \sum_{i=1}^{n}\sum_{\substack{l=1 \\ l\neq i}}^{n} D(W_i) D(W_l^r; W_i^c) \right.$$
$$\left. + \frac{1}{n}\sum_{i=1}^{n} \left[\frac{1}{n-1} \sum_{\substack{l=1 \\ l\neq i}}^{n} D(W_l^r; W_i^c) \right]^2 \right]$$
$$= \delta_n^2 [ED^2(W_i) - 2E(D(W_i^r; W_i^c)(D(W_l^r; W_i^c) + ED(W_l^r; W_i^c)D(W_j^r; W_i^c))](1 + o_p(1))$$
$$= \delta_n^2 E[D(W_i^r; W_i^c) - E_{W_l^r}(D(W_l^r; W_i^c))]^2 (1 + o_p(1)).$$

Proof of (2): with similar arguments and the additional assumption that $D(w_i^r; w_j^c) \in C^1$, we apply Lemma 2 to obtain

$$T_{3n} = 2\delta_n \left\{ \frac{1}{n}\sum_{i=1}^{n} \left[D(W_i) - \frac{1}{n-1} \sum_{\substack{l=1 \\ l\neq i}}^{n} D(W_l^r; W_i^c) \right] [\hat{g}_X(W_i) - g_X(W_i)] \right.$$
$$\left. - \frac{1}{n}\sum_{i=1}^{n} \left[D(W_i) - \frac{1}{n-1} \sum_{\substack{l=1 \\ l\neq i}}^{n} D(W_l^r; W_i^c) \right] \frac{1}{n-1} \sum_{\substack{l=1 \\ l\neq i}}^{n} (\hat{g}_X(W_l^r; W_i^c) - g_X(W_l^r; W_i^c)) \right\}$$
$$= O_p(\delta_n(n^{-\frac{1}{2}} h^{-1} + h^p)). \qquad \square$$

REFERENCES

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

Altonji, J. G., H. Ichimura, & T. Otsu (2012) Estimating derivatives in nonseparable models with limited dependent variables. *Econometrica* 80(4), 1701–1719.

Azzalini, A. & A. W. Bowman (1993) On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society Series B (Methodological)* 55, 549–557.

Banerjee, A. N. (2007) A method of estimating the average derivative. *Journal of Econometrics* 136(1), 65–88.

Bierens, H. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometric Theory* 65, 1129–1151.

Bontemps, C., M. Simioni, & Y. Surry (2008) Semiparametric hedonic price models: Assessing the effects of agricultural nonpoint source pollution. *Journal of Applied Econometrics* 23(6), 825–842.

Calonico, S., M. D. Cattaneo, & M. H. Farrell (2018) On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113(522), 767–779.

Calonico, S., M. D. Cattaneo, & M. H. Farrell (2019) Coverage Error Optimal Confidence Intervals for Local Polynomial Regression. Working Paper. https://arxiv.org/abs/1808.01398.

Calonico, S., M. D. Cattaneo, & R. Titiunik (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295–2326.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Cattaneo, M. D., R. K. Crump, & M. Jansson (2014) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30(1), 176–200.
Cattaneo, M. D., M. H. Farrell, & Y. Feng (2019a) Large sample properties of partitioning-based series estimators. *Annals of Statistics* 48, 1718.
Cattaneo, M. D. & M. Jansson (2018) Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica* 86(3), 955–995.
Cattaneo, M. D., M. Jansson, & X. Ma (2019b) Two-step estimation and inference with possibly many included covariates. *Review of Economic Studies* 86(3), 1095–1122.
Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.
Chen, R., H. Liang, & J. Wang (2011) Determination of linear components in additive models. *Journal of Nonparametric Statistics* 23(2), 367–383.
Chen, X. & Y. Fan (1999) Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.
Coppejans, M. & H. Sieg (2005) Kernel estimation of average derivatives and differences. *Journal of Business and Economic Statistics* 23(2), 211–225.
Delgado, M. & W. G. Manteiga (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29, 1469–1507.
Dette, H. (1999) A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics* 27, 1012–1040.
Ekeland, I., J. J. Heckman, & L. Nesheim (2004) Identification and estimation of hedonic models. *Journal of Political Economy*, 112(S1), S60–S109.
Eubank, R. L. & C. H. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* 85, 387–392.
Fan, J. & I. Gijbels (1996) *Local Polynomial Modeling and its Applications*. Chapman & Hall.
Fan, J. & L. Huang (2001) Goodness-of-Fit tests for parametric regression models. *Journal of the American Satistical Association* 96, 640–652.
Fan, Y. & E. Guerre (2016) Multivariate local polynomial estimators: Uniform boundary properties and asymptotic linear representation. *In Essays in Honor of Aman Ullah, Emerald Group Publishing Limited* 36, 489–537.
Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistic of absolutely regular processes with applications to model specification tests. *Journal of Nonparametric Statistics* 10, 245–271.
Fan, Y. & Q. Li (2002) A consistent model specification test based on the kernel sum of squares of residuals. *Econometric Reviews* 21, 337–352.
Fan, Y., Q. Li, & I. Min (2006) A nonparametric bootstrap test of conditional distributions. *Econometric Theory* 22, 587–613.
Fan, Y., Q. Li, & A. Weersink (1996) Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14, 460–468.
Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25(6), 1869–1892.
Gozalo, P. L. (1993) A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451–477.
Gozalo, P. L. & O. B. Linton (2001) Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104(1), 1–48.
Gu, J., D. Li, & D. Liu (2007) Bootstrap non-parametric significance test. *Journal of Nonparametric Statistics* 19, 215–230.
Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.
Hall, P. & S. R. Wilson (1991) Two guidelines for bootstrap hypothesis testing. *Biometrics* 47(2), 757–762.
Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Härdle, W. & T. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84(408), 986–995.
Hart, J. D. (1997) *Nonparametric Smoothing and Lack-of-Fit Test*. Springer.
Hjellvik, V. & D. Tjøstheim (1995) Nonparametric tests of linearity for time series. *Biometrika* 82, 351–368.
Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.
Hong, Y. & Y. J. Lee (2013) A loss function approach to model specification testing and its relative efficiency. *Annals of Statistics* 41, 1166–1203.
Hong, Y. M. & H. L. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
Horowitz, J. (1997) Bootstrap methods in econometrics: Theory and numerical performance. In D. Kreps and K. W. Wallis (eds.), *Advances in Economics and Econometrics, 7th World Congress*, vol. 3, pp. 188–222. Cambridge University Press.
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.
Hsiao, C., Q. Li, & J. Racine (2007) A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics* 140, 802–826.
Kaido, H. (2017) Asymptotically efficient estimation of weighted average derivatives with an interval censored variable. *Econometric Theory* 33 (5), 1218–1241.
Korolev, I. (2018) Consistent LM Type Specification Tests for Semiparametric Models. Working Paper.
Kress, J. P., M. H. Neumann, & Q. Yao (2008) Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and Its Interface* 1(2), 367–380.
Lavergne, P. & Q. Vuong (2000) Nonparametric significance testing. *Econometric Theory* 16(4), 576–601.
Lewbel, A. (1995) Consistent nonparametric hypothesis tests with an application to Slutsky symmetry. *Journal of Econometrics* 67(2), 379–401.
Li, H., Q. Li, & R. Liu (2016) Consistent model specification tests based on k-nearest-neighbor estimation method. *Journal of Econometrics* 194(1), 187–202.
Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
Li, Q. (2000) Efficient estimation of additive partially linear models. *International Economic Review*, 41(4), 1073–1092.
Li, Q., C. Hsiao, & J. Zinn (2003a). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.
Li, Q., X. Lu, & A. Ullah (2003b) Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics* 15(4–5), 607–624.
Li, Q. & J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, 145–165.
Lin, Z., Q. Li, & Y. Sun (2014) A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics* 178, 167–179.
Linton, O. B. & J. P. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.
Lu, Z. Q. (1996) Multivariate locally weighted polynomial fitting and partial derivative estimation. *Journal of Multivariate Analysis* 59 (2), 187–205.
Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.
Powell, J. L., J. H. Stock, T. M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.
Racine, J. (1997) Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* 15, 369–378.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press

Robinson, P. M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.
Rosen, S. (1974) Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82(1), 34–55.
Ruppert, D. & M. P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.
Sperlich, S., D. Tjøstheim, & L. Yang (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* 18, 197–251.
Staniswalis, J. G., & T. A. Severini (1991) Diagnostics for assessing regression models. *Journal of the American Satistical Association* 86, 684–692.
Stengos, T. & Y. Sun (2001) A consistent model specification test for a regression function based on nonparametric wavelet estimation. *Econometric Reviews* 20(1), 41–60.
Su, L., & X. Lu (2013) Nonparametric dynamic panel data models: Kernel estimation and specification testing. *Journal of Econometrics* 176(2), 112–133.
Su, L. & A. Ullah (2008) Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144, 193–218.
Su, L. & A. Ullah (2013) A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory* 29(1), 187–212.
Ullah, A. (1985) Specification analysis of econometric models. *Journal of Quantitative Economics* 1, 187–209.
Wang, J. & L. Yang (2009) Efficient and fast spline-backfitted kernel smoothing of additive models. *Annals of the Institute of Statistical Mathematics* 61(3), 663–690.
Wang, X. & K. C. Carriere (2011) Assessing additivity in nonparametric models – a kernel-based method. *Canadian Journal of Statistics* 39(4), 632–655.
Whang, Y. J. (2000) Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics* 98(1), 27–46.
Wooldridge, J. M. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* 8, 452–475.
Yao, F. & C. Martins-Filho (2015) An asymptotic characterization of finite degree U-statistics with sample size-dependent kernels: Applications to nonparametric estimators and test statistics. *Communications in Statistics–Theory and Methods* 44(15), 3251–3265.
Yao, F. & A. Ullah (2013) A nonparametric R-square test for the presence of relevant variables. *Journal of Statistical Planning and Inference* 143(9), 1527–1547.
Yao, F., F. Zhang, & S. C. Kumbhakar (2019) Semiparametric smooth coefficient stochastic frontier model with panel data. *Journal of Business and Economic Statistics* 37(3), 556–572.
Yatchew, A. (1992) Nonparametric regression tests based on least squares. *Econometric Theory* 8(4), 435–451.
Zheng, J. X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

https://doi.org/10.1017/S0266466620000407 Published online by Cambridge University Press