A significance test for covariates in nonparametric regression

Pascal Lavergne

Toulouse School of Economics, France
e-mail: pascal.lavergne@ut-capitole.fr

Samuel Maistre

CREST (Ensai), IRMAR
and
Université de Lyon
Université Claude Bernard Lyon 1
CNRS UMR 5208, Institut Camille Jordan
43 blvd. du 11 novembre 1918
F-69622 Villeurbanne cedex, France
e-mail: maistre@math.univ-lyon1.fr

and

Valentin Patilea*

CREST (Ensai), IRMAR, France
and
Bucharest University of Economic Studies, Romania
e-mail: patilea@ensai.fr

Abstract: We consider testing the significance of a subset of covariates in a nonparametric regression. These covariates can be continuous and/or discrete. We propose a new kernel-based test that smoothes only over the covariates appearing under the null hypothesis, so that the curse of dimensionality is mitigated. The test statistic is asymptotically pivotal and the rate of which the test detects local alternatives depends only on the dimension of the covariates under the null hypothesis. We show the validity of wild bootstrap for the test. In small samples, our test is competitive compared to existing procedures.

Contents

1. Introduction

Testing the significance of covariates is common in applied regression analysis. Sound parametric inference hinges on the correct functional specification of the regression function, but the likelihood of misspecification in a parametric framework cannot be ignored, especially as applied researchers tend to choose functional forms on the basis of parsimony and tractability. Significance testing in a nonparametric framework has therefore obvious appeal as it requires much less restrictive assumptions. Fan (1996); Fan and Li (1996); Racine (1997); Chen and Fan (1999); Lavergne and Vuong (2000); Ait-Sahalia, Bickel and Stoker (2001), and Delgado and González Manteiga (2001) proposed tests of significance for continuous variables in nonparametric regression models. A related issue is testing for the equality of nonparametric regression functions across subsamples, see Delgado (1993); Dette and Neumeyer (2001); Lavergne (2001); Neumeyer and Dette (2003); Racine, Hart and Li (2006). When the division into subsamples is governed by a discrete random variable, this is akin to testing the significance of this discrete variable, see Lavergne (2001). Volgushev et al. (2013) considered significance testing in nonparametric quantile regression. For each test, one needs first to estimate the model without the covariates under test, that is under the null hypothesis. The result is then used to check the significance of extra covariates. Two competing approaches are then possible. In the "smoothing approach," one regresses the residuals onto the whole set of covariates nonparametrically, while in the "empirical process approach" one uses the empirical process of residuals marked by a function of all covariates.

In this work, we adopt an hybrid approach to develop a new significance test of a subset of covariates in a nonparametric regression. Our new test has three specific features. First, it does not require smoothing with respect to the covariates under test as in the "empirical process approach." This allows to mitigate the curse of dimensionality that appears with nonparametric smoothing, hence improving the power properties of the test. Our simulation results show that indeed our test can be more powerful than such competitors under different

kind of alternatives. Second, the test statistic is asymptotically pivotal as in the "smoothing approach," while wild bootstrap can be used to obtain small samples critical values of the test. This yields a test whose level is well controlled by bootstrapping independently of the smoothing parameter choice, as shown in our simulations. Third, our test equally applies whether the covariates under test are continuous or discrete, showing that there is no need of a specific tailored procedure for each situation. Hence our test alos applies to comparing two or more nonparametric regressions that are characterized by different values of a discrete random variable.

The paper is organized as follows. In Section 2, we present our testing procedure. In Section 3, we study its asymptotic properties under a sequence of local alternatives and we establish the validity of wild bootstrap. In Section 4, we compare the small sample behavior of our test to some existing procedures. Section 6 gathers our proofs.

2. Testing framework and procedure

2.1. Testing principle

We want to assess the significance of $X \in \mathbb{R}^q$ in the nonparametric regression of $Y \in \mathbb{R}$ on $W \in \mathbb{R}^p$ and X. Formally, this corresponds to the null hypothesis

$$H_0 \, : \, \mathbb{E}\left[Y \mid W, X\right] = \mathbb{E}\left[Y \mid W\right] \quad \text{a.s.}$$

which is equivalent to

$$H_0 \, : \, \mathbb{E}\left[u \mid W, X\right] = 0 \quad \text{a.s.} \tag{1}$$

where $u = Y - \mathbb{E}[Y \mid W]$. The corresponding alternative hypothesis is

$$H_1 \, : \, \mathbb{P}\left\{\mathbb{E}\left[u \mid W, X\right] = 0\right\} < 1.$$

The following result is the cornerstone of our approach. It characterizes the null hypothesis H_0 using a suitable unconditional moment equation.

Lemma 1. *Let (W_1, X_1, u_1) and (W_2, X_2, u_2) be two independent draws of (W, X, u), $\nu(W)$ a strictly positive function on the support of W such that $\mathbb{E}[u^2\nu^2(W)] < \infty$, and $K(\cdot)$ and $\psi(\cdot)$ even functions with (almost everywhere) positive Fourier integrable transforms. Define*

$$I\left(h\right) = \mathbb{E}\left[u_1 u_2 \nu\left(W_1\right) \nu\left(W_2\right) h^{-p} K\left(\left(W_1 - W_2\right)/h\right) \psi\left(X_1 - X_2\right)\right].$$

Then for any $h > 0$, $I(h) \geq 0$ and

$$\mathbb{E}\left[u \mid W, X\right] = 0 \; a.s. \Leftrightarrow I(h) = 0 \Leftrightarrow \lim_{h \to 0} I(h) = 0.$$

Proof. Let $\langle \cdot, \cdot \rangle$ denote the standard inner product. Using Fourier Inversion Theorem, change of variables, and elementary properties of conditional expectation,

$$I\left(h\right) = \mathbb{E}\left[u_1 u_2 \nu\left(W_1\right) \nu\left(W_2\right) \int_{\mathbb{R}^p} e^{2\pi i \langle t, \, W_1 - W_2 \rangle} \mathcal{F}\left[K\right]\left(th\right) \, dt\right.$$

$$\times \int_{\mathbb{R}^q} e^{2\pi i \langle s,\, X_1 - X_2 \rangle} \mathcal{F}[\psi](s)\, ds \Bigg]$$

$$= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left| \mathbb{E}\left[\mathbb{E}[u \mid W, X] \nu(W) e^{2\pi i \{\langle t, W \rangle + \langle s, X \rangle\}} \right] \right|^2$$

$$\times \mathcal{F}[K](th)\, \mathcal{F}[\psi](s)\, dt ds.$$

Since the Fourier transforms $\mathcal{F}[K]$ and $\mathcal{F}[\psi]$ are strictly positive, $I(h) \geq 0$. Moreover $I(h) = 0$ iff

$$\mathbb{E}\left[\mathbb{E}[u \mid W, X] \nu(W) e^{2\pi i \{\langle t, W \rangle + \langle s, X \rangle\}} \right] = 0 \qquad \forall t, s.$$

This is equivalent to $\mathbb{E}[u \mid W, X]\nu(W) = 0$ a.s., which by our assumption on $\nu(\cdot)$ is equivalent to H_0.

When $h \to 0$, $\mathcal{F}[K](th) \to \mathcal{F}[K](0)$, so that by Lebesgue dominated convergence $\lim_{h \to 0} I(h) \geq 0$, and $= 0$ iff H_0 holds, by the same reasoning as above. □

2.2. The test

Lemma 1 holds whether the covariates W and X are continuous or discrete. For now, we assume W is continuously distributed, and we later comment on how to modify our procedure in the case where some of its components are discrete. We however do not restrict X to be continuous. Since it is sufficient to test whether $I(h) = 0$ for any arbitrary h, we can choose h to obtain desirable properties. So we consider a sequence of h decreasing to zero when the sample size increases, which is one of the ingredient that allows to obtain a tractable asymptotic distribution for the test statistic.

Assume we have at hand a random sample (Y_i, W_i, X_i), $1 \leq i \leq n$, from (Y, W, X). In what follows, $f(\cdot)$ denotes the density of W, $r(\cdot) = \mathbb{E}[Y \mid W = \cdot]$, $u = Y - r(W)$, and f_i, r_i, u_i respectively denote $f(W_i)$, $r(W_i)$, and $Y_i - r(W_i)$. Since nonparametric estimation should be entertained to approximate u_i, we consider usual kernel estimators based on kernel $L(\cdot)$ and bandwidth g. With $L_{nik} = \frac{1}{g^p} L(\frac{W_i - W_k}{g})$, let

$$\hat{f}_i = (n-1)^{-1} \sum_{k \neq i, k=1}^{n} L_{nik},$$

$$\hat{r}_i = \frac{1}{\hat{f}_i} \frac{1}{(n-1)} \sum_{k \neq i, k=1}^{n} Y_k L_{nik}$$

so that $\quad \hat{u}_i = Y_i - \hat{r}_i = \frac{1}{\hat{f}_i} \frac{1}{(n-1)} \sum_{k \neq i, k=1}^{n} (Y_i - Y_k) L_{nik}.$

Denote by $n^{(m)}$ the number of arrangements of m distinct elements among n, and by $[1/n^{(m)}] \sum_a$, the average over these arrangements. In order to avoid random denominators, we choose $\nu(W) = f(W)$, which fulfills the assumption

of Lemma 1. Then we can estimate $I(h)$ by the U-statistic

$$
\begin{aligned}
\widehat{I}_n &= \frac{1}{n^{(2)}} \sum_a \hat{u}_i \hat{f}_i \hat{u}_j \hat{f}_j K_{nij} \psi_{ij} \\
&= \frac{1}{n^{(2)} (n-1)^2} \sum_a \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k)(Y_j - Y_l) L_{nik} L_{njl} K_{nij} \psi_{ij},
\end{aligned}
$$

with $K_{nij} = \frac{1}{h^p} K(\frac{W_i - W_j}{h})$ and $\psi_{ij} = \psi(X_i - X_j)$. We also consider the alternative statistic

$$
\tilde{I}_n = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k)(Y_j - Y_l) L_{nik} L_{njl} K_{nij} \psi_{ij}.
$$

It is clear that $\tilde{I}_n$ is obtained from $\widehat{I}_n$ by removing asymptotically negligible "diagonal" terms. Under the null hypothesis, both statistics will have the same asymptotic normal distribution, but removing diagonal terms reduces the bias of the statistic under H_0. Our statistics $\tilde{I}_n$ and $\widehat{I}_n$ are respectively similar to the ones of Fan and Li (1996) and Lavergne and Vuong (2000), with the fundamental difference that there is no smoothing relative to the covariates X. Indeed these authors used a multidimensional smoothing kernel over (W, X), that is $h^{-(p+q)} \tilde{K}((W_i - W_j)/h, (X_i - X_j)/h)$, while we use $K_{nij} \psi_{ij}$. For I_n being either $\tilde{I}_n$ or $\widehat{I}_n$, we will show that $nh^{p/2} I_n$ converges to $\mathcal{N}(0, \omega^2)$ in distribution under H_0 and $nh^{p/2} I_n \xrightarrow{p} \infty$ under H_1. By contrast, the statistics of Fan and Li (1996) and Lavergne and Vuong (2000) exhibit a $nh^{(p+q)/2}$ rate of convergence. The alternative test of Delgado and González Manteiga (2001) uses the kernel residuals $\hat{u}_i$ and the empirical process approach of Stute (1997). This avoids extra smoothing, but at the cost of a test statistic with a non pivotal asymptotic law under H_0. Hence, our proposal is an hybrid approach that combines the advantages of existing procedures, namely smoothing only for the variables W appearing under the null hypothesis but with an asymptotic normal distribution for the statistic. Given a consistent estimator ω_n^2 of ω^2, as provided in the next section, we obtain an asymptotic α-level test of H_0 as

$$
\text{Reject } H_0 \text{ if } nh^{p/2} I_n / \omega_n > z_{1-\alpha},
$$

where $z_{1-\alpha}$ is the $(1 - \alpha)$-th quantile of the standard normal distribution. In small samples, we will show the validity of a wild bootstrap scheme to obtain critical values.

The test applies whether X is continuous or has some discrete components. When X is discrete, the null hypothesis under test is that the regression functions $\mathbb{E}[Y \mid W, X]$ are equal across possible values of X. The procedure is also easily adapted to some discrete components of W. In that case, one would replace kernel smoothing by cells' indicators for the discrete components, so that for W composed of continuous W_c of dimension p_c and discrete W_d, one would use $h^{-p_c} K(\frac{W_{ic} - W_{jc}}{h}) \mathbb{I}(W_{id} = W_{jd})$ instead of K_{nij}. It would also be possible to smooth on the discrete components, as proposed by Racine and Li (2004). To obtain scale invariance, we recommend that observations on covariates should be scaled, say by their sample standard deviation as is customary in nonpara-

metric estimation. It is equally important to scale the X_i before they are used as arguments of $\psi(\cdot)$ to preserve such invariance.

The outcome of the test may depend on the choice of the kernels $K(\cdot)$ and $L(\cdot)$, while this influence is expected to be limited as usual in nonparametric estimation. The choice of the function $\psi(\cdot)$ might be more important, but our simulations reveal that it is not so. From our theoretical study, this function, as well as $K(\cdot)$ should possess an almost everywhere positive and integrable Fourier transform. This is true for (products of) the triangular, normal, Laplace, and logistic densities, see Johnson, Kotz and Balakrishnan (1995), and for a Student density, see Hurst (1995). Alternatively, one can choose $\psi(x)$ as a univariate density applied to some transformation of x, such as its norm. This yields $\psi(x) = g(\|x\|)$ where $g(\cdot)$ is any of the above univariate densities. This is the form we will consider in our simulations to study the influence of $\psi(\cdot)$.

3. Theoretical properties

We here give the asymptotic properties of our test statistics under H_0 and some local alternatives. To do so in a compact way, we consider the sequence of hypotheses

$$H_{1n} \,:\, \mathbb{E}\left[Y \mid W, X\right] = r\left(W\right) + \delta_n d\left(W, X\right), \qquad n \geq 1,$$

where $d(\cdot)$ is a fixed integrable function. Since $r(W) = \mathbb{E}[Y \mid W]$, our setup imposes $\mathbb{E}[d(W, X) \mid W] = 0$. The null hypothesis corresponds to the case $\delta_n \equiv 0$, while considering a sequence $\delta_n \to 0$ yields local Pitman-like alternatives.

3.1. Assumptions

We begin by some useful definitions.

Definition 1. (i) $\mathcal{U}^p$ is the class of integrable uniformly continuous functions from $\mathbb{R}^p$ to $\mathbb{R}$;

(ii) $\mathcal{D}_s^p$ is the class of m-times differentiable functions from $\mathbb{R}^p$ to $\mathbb{R}$, with derivatives of order $\lfloor s \rfloor$ that are uniformly Lipschitz continuous of order $s - \lfloor s \rfloor$, where $\lfloor s \rfloor$ denotes the integer such that $\lfloor s \rfloor \leq s < \lfloor s \rfloor + 1$.

Note that a function belonging to $\mathcal{U}^p$ is necessarily bounded.

Definition 2. $\mathcal{K}_m^p$, $m \geq 2$, is the class of even integrable functions $K \,:\, \mathbb{R}^p \to \mathbb{R}$ with compact support satisfying $\int K(t)dt = 1$ and, if $t = (t_1, \ldots, t_p)$,

$$\int_{\mathbb{R}^p} t_1^{\alpha_1} \ldots t_p^{\alpha_p} K\left(t\right) dt = 0 \quad \text{for} \quad 0 < \sum_{i=1}^p \alpha_i \leq m - 1,\, \alpha_i \in \mathbb{N} \quad \forall i$$

This definition of higher-order kernels is standard in nonparametric estimation. The compact support assumption is made for simplicity and could be relaxed at the expense of technical conditions on the rate of decrease of the kernels at infinity, see e.g. Definition 1 in Fan and Li (1996). In particular, the gaussian kernel could be allowed for. We are now ready to list our assumptions.

Assumption 1. *(i) For any $x \in \mathbb{R}^q$ in the support of X, the vector W admits a conditional density given $X = x$ with respect to the Lebesgue measure in $\mathbb{R}^p$, denoted by $\pi(\cdot \mid x)$. Moreover, $\mathbb{E}[Y^8] < \infty$. (ii) The observations (W_i, X_i, Y_i), $i = 1, \ldots, n$ are independent and identically distributed as (W, X, Y).*

The existence of the conditional density given $X = x$ for all $x \in \mathbb{R}^q$ in the support of X implies that W admits a density with respect to the Lebesgue measure on $\mathbb{R}^p$. As noted above, our results easily generalizes to some discrete components of W, but for the sake of simplicity we do not formally consider this in our theoretical analysis.

Assumption 2. (i) *$f(\cdot)$ and $r(\cdot)f(\cdot)$ belong to $\mathcal{U}^p \cap \mathcal{D}_s^p$, $s \geq 2$;*
(ii) *$\mathbb{E}[u^2 \mid W = \cdot]f(\cdot)$, $\mathbb{E}[u^4 \mid W = \cdot]f^4(\cdot)$ belong to $\mathcal{U}^p$;*
(iii) *the function $\psi(\cdot)$ is bounded and has an almost everywhere positive and integrable Fourier transform;*
(iv) *$K(\cdot) \in \mathcal{K}_2^p$ and has an almost everywhere positive and integrable Fourier transform, while $L(\cdot) \in \mathcal{K}_{\lfloor s \rfloor}^p$ and is of bounded variation;*
(v) *let $\sigma^2(w, x) = \mathbb{E}[u^2 \mid W = w, X = x]$, then $\sigma^2(\cdot, x)f^2(\cdot)\pi(\cdot \mid x)$ belongs to $\mathcal{U}^p$ for any x in the support of X, has integrable Fourier transform, and $\mathbb{E}[\sigma^4(W, X)f^4(W)\pi(W \mid X)] < \infty$;*
(vi) *$\mathbb{E}[d^2(W, X) \mid W = \cdot]f^2(\cdot)$ belongs to $\mathcal{U}^p$, $d(\cdot, x)f(\cdot)\pi(\cdot \mid x)$ is integrable and squared integrable for any x in the support of X, and $\mathbb{E}[d^2(W, X)f^2(W)\pi(W \mid X)] < \infty$.*

Standard regularity conditions are assumed for various functions. A higher-order kernel $L(\cdot)$ is used in conjunction with the differentiability conditions in (i) to ensure that the bias in nonparametric estimation is small enough.

3.2. Asymptotic analysis

The following result characterizes the behavior of our statistics under the null hypothesis and a sequence of local alternatives.

Theorem 1. *Let I_n be any of the statistics $\widehat{I}_n$ or $\tilde{I}_n$. Under Assumptions 1 and 2, and if as $n \to \infty$ (i) $g, h \to 0$, (ii) $ng^{4p/3}/\ln n, nh^p \to \infty$, (iii) $nh^{p/2}g^{2s} \to 0$, and (iv) $h/g \to 0$ if $I_n = \tilde{I}_n$ or $h/g^2 \to 0$ if $I_n = \widehat{I}_n$, then*

(i) *If $\delta_n^2 nh^{p/2} \to C$ with $0 \leq C < \infty$, $nh^{p/2}I_n \xrightarrow{d} \mathcal{N}(C\mu, \omega^2)$ where*

$$\mu = \mathbb{E}\left[\int d(w, X_1) d(w, X_2) f^2(w) \pi(w \mid X_1) \pi(w \mid X_2) \psi(X_1 - X_2) dw\right] > 0$$

$$and \quad \omega^2 = 2 \int K^2(s) ds$$

$$\mathbb{E}\left[\int \sigma^2(w, X_1) \sigma^2(w, X_2) f^4(w) \pi(w \mid X_1) \pi(w \mid X_2) \psi^2(X_1 - X_2) dw\right].$$

(ii) *If $\delta_n^2 nh^{p/2} \to \infty$, $nh^{p/2}I_n \xrightarrow{p} \infty$.*

The rate of convergence of the test statistic depends only on the dimension of W, the covariates present under the null hypothesis, but not on the dimension of X, the covariates under test. Similarly, the rate of local alternatives that are detected by the test depends only on the dimension of W. As shown in our simulations, this yields some gain in power compared to competing tests based on the smoothing approach that smooth over th whole set of covariates and whose rate of convergence depends on the dimension of W and X. Conditions (i) to (iv) together require that $s > p/2$ for $I_n = \tilde{I}_n$ and $s > p/4$ for $I_n = \widehat{I}_n$, so removing diagonal terms in $\widehat{I}_n$ allows to weaken the restrictions on the bandwidths. Condition (ii) could be slightly weakened to $ng^p \to \infty$ at the price of handling high order U-statistics in the proofs, but allows for a shorter argument based on empirical processes, see Lemma 3 in the proofs section.

To estimate ω^2, we can either mimic Lavergne and Vuong (2000) to consider

$$\tilde{\omega}_n^2 = \frac{2h^p}{n^{(6)}} \sum_a \left(Y_i - Y_k\right)\left(Y_i - Y_{k'}\right)\left(Y_j - Y_l\right)\left(Y_j - Y_{l'}\right) L_{nik}L_{nik'}L_{njl}L_{njl'} K_{nij}^2 \psi_{ij}^2,$$

or generalize the variance estimator of Fan and Li (1996) as

$$\widehat{\omega}_n^2 = \frac{2h^p}{n^{(2)}} \sum_a \hat{u}_i^2 \hat{f}_i^2 \hat{u}_j^2 \hat{f}_j^2 K_{nij}^2 \psi_{ij}^2.$$

The first one is consistent for ω^2 under both the null and alternative hypothesis, but the latter is faster to compute.

Corollary 1. *Let I_n be any of the statistics $\widehat{I}_n$ or $\tilde{I}_n$ and let ω_n denote any of $\widehat{\omega}_n$ or $\tilde{\omega}_n$. Under the assumptions of Theorem 1, the test that rejects H_0 when $nh^{p/2}I_n/\omega_n > z_{1-\alpha}$ is of asymptotic level α under H_0 and is consistent under the sequence of local alternatives H_{1n} provided $\delta_n^2 nh^{p/2} \to \infty$.*

3.3. Bootstrap critical values

It is known that asymptotic theory may be inaccurate for small and moderate samples when using smoothing methods. Hence, as in Härdle and Mammen (1993) or Delgado and González Manteiga (2001), we consider a wild bootstrap procedure to approximate the quantiles of our test statistic. Resamples are obtained from $Y_i^* = \hat{r}_i + u_i^*$, where $u_i^* = \eta_i \hat{u}_i$ and η_i are i.i.d. variables independent of the initial sample with $\mathbb{E}\eta_i = 0$ and $\mathbb{E}\eta_i^2 = \mathbb{E}\eta_i^3 = 1$, $1 \leq i \leq n$. The η_i could for instance follow the two-point law of Mammen (1993). With at hand a bootstrap sample (Y_i^*, W_i, X_i), $1 \leq i \leq n$, we obtain a bootstrapped statistic $nh^{p/2}I_n^*/\omega_n^*$ with bootstrapped observations Y_i^* in place of original observations Y_i. When the scheme is repeated many times, the bootstrap critical value $z_{1-\alpha,n}^\star$ at level α is the empirical $(1-\alpha)$-th quantile of the bootstrapped test statistics. The asymptotic validity of this bootstrap procedure is guaranteed by the following result.

Theorem 2. *Suppose Assumptions 1, 2, and Conditions (i) to (iii) of Theorem 1 hold. Moreover, assume $\inf_{w \in \mathcal{S}_W} f(w) > 0$ and $h/g^2 \to 0$. Then for I_n^**

equal to any of $\widehat{I}_n^*$ and $\tilde{I}_n^*$,

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left[nh^{p/2} I_n^* / \omega_n^* \le z \mid Y_1, W_1, X_1, \ldots, Y_n, W_n, X_n \right] - \Phi(z) \right| \xrightarrow{p} 0,$$

where $\Phi(\cdot)$ is the standard normal distribution function.

4. Monte Carlo study

We investigated the small sample behavior of our test and studied its performances relative to alternative tests. We generated data through

$$Y = (W'\theta)^3 - W'\theta + \delta d(X) + \varepsilon$$

where W follows a two-dimensional standard normal, X independently follows a q-variate standard normal, $\varepsilon \sim \mathcal{N}(0, 4)$, and we set $\theta = (1, -1)'/\sqrt{2}$. The null hypothesis corresponds to $\delta = 0$, and we considered various forms for $d(\cdot)$ to investigate power. We only considered the test based on $\tilde{I}_n/\hat{\omega}_n$, labelled LMP, as preliminary simulation results showed that the tests based on $\tilde{I}_n$ had similar or better performances than the test based on $\widehat{I}_n$, and $\hat{\omega}_n$ is computationally less demanding than $\tilde{\omega}_n$. We compared it to the test of Lavergne and Vuong (2000) (hereafter LV) and the test of Delgado and González Manteiga (2001) (hereafter DGM). The latter is based on the Cramer-von-Mises statistic

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{n} \widehat{u}_j \widehat{f}_j \mathbf{1}\{W_j \le W_i\} \mathbf{1}\{X_j \le X_i\} \right]^2.$$

As its distribution under H_0 is not asymptotically pivotal, critical values must be obtained by wild bootstrapping. To compute bootstrap critical values, we used 199 bootstrap replications and the two-point distribution

$$\mathbb{P}\left(\eta_i = \frac{1 - \sqrt{5}}{2} \right) = \frac{5 + \sqrt{5}}{10}, \ \mathbb{P}\left(\eta_i = \frac{1 + \sqrt{5}}{2} \right) = \frac{5 - \sqrt{5}}{10}.$$

We also applied this method to obtain small sample critical values for LV test as well as our test. For all tests, each time a kernel appears, we used the normal density. The bandwidth parameters are set to $g = C_1 n^{-1/6}$ and $h = C_2 n^{-2.1/6}$, and we let C_1 and C_2 vary to investigate the sensitivity of our results to the smoothing parameters choice. To study the influence of $\psi(\cdot)$ on our test, we considered $\psi(x) = l(\|x\|)$, where $l(\cdot)$ is a triangular or normal density, each with a second moment equal to one.

Figures 1 and 2 reports the empirical level of the various tests for $n = 100$ based on 5000 replications when we let C_1, C_2, and q vary. For our test, bootstrapping yields more accurate rejection levels than the asymptotic normal critical values for any bandwidth factors C_1 and C_2 and dimension q. The choice of $\psi(\cdot)$ does not influence the results. The empirical level of LV test is much more sensitive to the bandwidth and the dimension. The empirical level of the
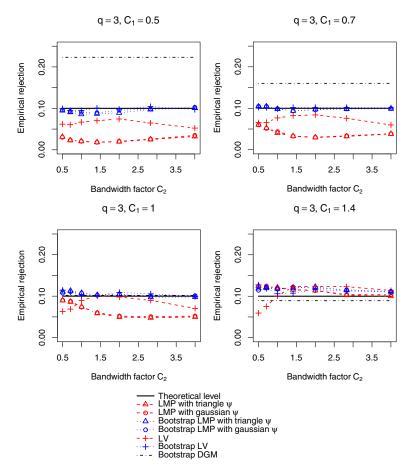
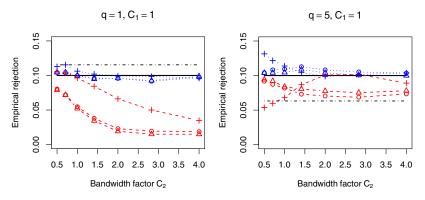FIG 1. *Empirical rejections under H_0 as a function of the bandwidth, $n = 100$.*

FIG 2. *Empirical rejections under H_0 as a function of the bandwidth, $n = 100$.*
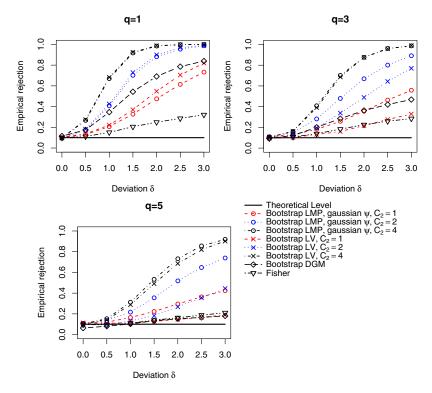
FIG 3. *Empirical power curves for a quadratic alternative, $n = 100$.*

DGM test is close to the nominal one for a low dimension q and large enough bandwidth g, but decreases with increasing q and increases with decreasing bandwidth g.

We also investigated how power varies with different alternatives as specified by $d(\cdot)$. We first focus on a quadratic alternative, where $d(X) = (X'\beta - 1)^2/\sqrt{2}$, with $\beta = (1, , 1, , \dots)'/\sqrt{q}$. Figure 3 reports power curves of the different tests for the quadratic alternative, $n = 100$, and a nominal level of 10% based on 2500 replications. We also report the power of a Fisher test based on a linear specification in the components of X. The power of our test, as well as the one of LV test, increases when the bandwidth factor C_2 increases, in line with theoretical findings. Our test always perform as well or better than LV test, but for $q = 1$ and $C_2 = 1$. These tests, as well as DGM test, perform better than the Fisher test. For a low dimension, i.e. $q = 1$, there is no clear ranking of the DGM and our test depending on the bandwidth. For larger dimensions, $q = 3$ or 5, our test dominates DGM test. In Figure 4, we let n vary for a dimension $q = 5$. The power of all tests improve, but our main qualitative findings are not affected. It is noteworthy that the power advantage of our test compared to LV test become more pronounced as n increases. In Figure 5, we considered a sine alternative defined as $d(X) = \sum_{j=1}^{q} \sin(\gamma X_j)$ for $\gamma = 1.5$ or 3. The ranking
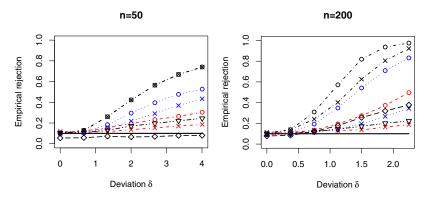
n=50 **n=200**

FIG 4. *Empirical power curves for a quadratic alternative, $q = 5$.*

of the different tests heavily depends on the frequency and the dimension of X. For a low frequency alternative, i.e. $\gamma = 1.5$, our main findings remain unchanged. While DGM test is best for a low dimension, its power decreases a lot when increasing the dimension of variables under test and/or the frequency of the alternative. The power of our test is affected when increasing q, but to a lesser extent. For a high frequency alternative, i.e. $\gamma = 3$, the LV test always dominates, as smoothing on the variables under test allows to adapt better to such alternatives.

We also considered the case of a discrete X. We generated data following

$$Y = W^3 - W + \delta d\left(W\right) \mathbf{1}\left\{X = 1\right\} + \varepsilon$$

where W follows a univariate standard normal, $\varepsilon \sim \mathcal{N}(0, 4)$, and X is Bernoulli with probability of success 0.6. In this setup, testing the significance of X is equivalent to testing the equality of the regression curves on Y on W when X equals 0 or 1. We considered two competitors to our test: the test proposed by Lavergne (2001), which is similar to our test with the main difference that $\psi(\cdot)$ is the indicator function, i.e. $\psi(X_i - X_j) = \mathbf{1}\{X_i = X_j\}$; and the test of Neumeyer and Dette (2003) (hereafter ND), which is similar in spirit to the DGM test. Other details of the simulations are similar to above. Figure 6 show that bootstrap yield accurate rejection levels.

Figure 7 report power of the tests against the quadratic alternative $d(W) = W^2 - 1$ and the sine alternative $d(W) = \sin(1.5\,W)$. Our test and Lavergne's one have basically the same power in both cases. Under a quadratic alternative, ND test is more powerful, while under a sine alternative, our test outperforms ND test for a moderate ($C_2 = 2$) or large ($C_2 = 4$) bandwidth.

5. Conclusion

We have developed a testing procedure for the significance of covariates in a nonparametric regression. Smoothing is entertained only for the covariates under
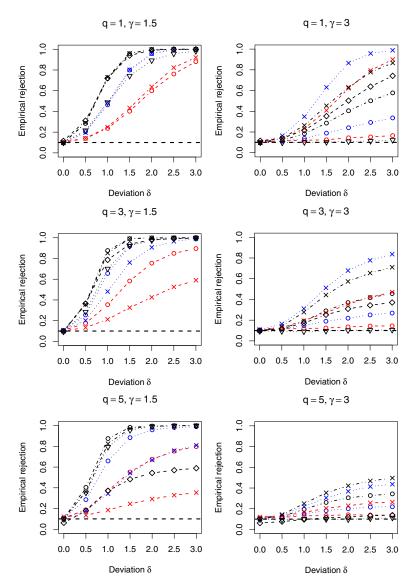
FIG 5. *Empirical power curves for sine alternative, $n = 100$.*

the null hypothesis. The resulting test statistic is asymptotically pivotal, and wild bootstrap can be used to obtain critical values in small and moderate samples. The test is versatile, as it applies whether the covariates under test are continuous and/or discrete. Simulations reveal that our test outperforms its competitors in many situations, and especially when the dimension of covariates is large.
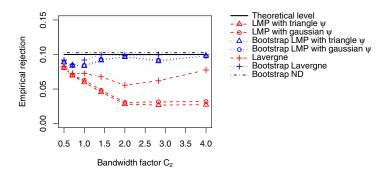
FIG 6. *Empirical rejection under H_0 as a function of the bandwidth, X Bernoulli and $n = 100$.*
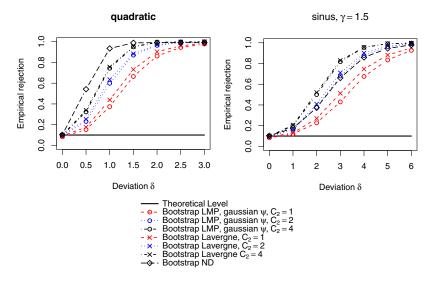

FIG 7. *Empirical power curves, X Bernoulli and $n = 100$.*

6. Proofs

We here provide the proofs of the main results. Technical lemmas are relegated to the Appendix. We let $\mathcal{F}_X[\delta](u) = \mathbb{E}[e^{-2\pi i \langle X, u \rangle} \delta(X)]$, $u \in \mathbb{R}^q$, for any integrable function $\delta(X)$. Moreover, for any index set I not containing i with cardinality $|I|$, let

$$\widehat{f_i}^I = (n - |I| - 1)^{-1} \sum_{k \neq i, k \notin I} L_{nik}.$$

This is consistent with the notation $\widehat{f_i}$, in which case I is the empty set.

6.1. Proof of Theorem 1

We first consider the case $I_n = \tilde{I}_n$. Next, we study the difference between $\tilde{I}_n$ and $\hat{I}_n$ and hence deduce the result for $I_n = \hat{I}_n$.

Case $I_n = \tilde{I}_n$. Consider the decomposition

$$
\begin{aligned}
I_n &= \frac{1}{n^{(4)}} \sum_a (u_i - u_k)(u_j - u_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{2}{n^{(4)}} \sum_a (u_i - u_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a (r_i - r_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_1 + 2I_2 + I_3,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i u_j f_i f_j K_{nij} \psi_{ij} + \frac{2(n-2)}{n-3} \frac{1}{n^{(2)}} \sum_a u_i \big(\hat{f}_i^j - f_i\big) u_j f_j K_{nij} \psi_{ij} \\
&\quad + \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i \big(\hat{f}_i^j - f_i\big) u_j \big(\hat{f}_j^i - f_j\big) K_{nij} \psi_{ij} - \frac{2}{n^{(3)}} \sum_a u_i f_i u_l L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n^{(3)}} \sum_a u_i \big(\hat{f}_i^{j,l} - f_i\big) u_l L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(4)}} \sum_a u_k u_l L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{n^{(4)}} \sum_a u_i u_j L_{nik} L_{njk} K_{nij} \psi_{ij} \\
&= \frac{n-2}{n-3} [I_{0n} + 2I_{1,1} + I_{1,2}] - 2I_{1,3} - 2I_{1,4} + I_{1,5} - I_{1,6},
\end{aligned}
$$

and

$$
\begin{aligned}
I_2 &= \frac{1}{n^{(3)}} \sum_a u_i f_i (r_j - r_l) L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(3)}} \sum_a u_i \big(\hat{f}_i^{j,l} - f_i\big)(r_j - r_l) L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{n^{(4)}} \sum_a u_k (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} = I_{2,1} + I_{2,2} - I_{2,3}.
\end{aligned}
$$

In Proposition 1 we prove that, under H_0, I_{0n} is asymptotically centered Gaussian with variance ω^2, while in Proposition 2 we prove that, under H_{1n}, I_{0n} is asymptotically Gaussian with mean μ and variance ω^2 provided $\delta_n^2 n h^{p/2}$ converges to some positive real number. In Propositions 3 and 4 we show that all remaining terms in the decomposition of I_n are asymptotically negligible.

Proposition 1. *Under the conditions of Theorem 1, $nh^{p/2} I_{0n} \xrightarrow{d} \mathcal{N}(0, \omega^2)$ under H_0.*

Proof. Let us define the martingale array $\{S_{n,m}, \mathcal{F}_{n,m}, 1 \leq m \leq n, n \geq 1\}$ where $S_{n,1} = 0$, and

$$S_{n,m} = \sum_{i=1}^{m} G_{n,i} \quad \text{with} \quad G_{n,i} = \frac{2h^{p/2}}{n-1} u_i f_i \sum_{j=1}^{i-1} u_j f_j K_{nij} \psi_{ij}, \qquad 2 \le i, m \le n,$$

and $\mathcal{F}_{n,m}$ is the σ-field generated by $\{W_1, \ldots, W_n, X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$. Thus $nh^{p/2}I_{0n} = S_{n,n}$. Also define

$$V_n^2 = \sum_{i=2}^{n} E\left[G_{n,i}^2 \mid \mathcal{F}_{n,i-1}\right] = \frac{4h^p}{(n-1)^2} \sum_{i=2}^{n} \sigma_i^2 f_i^2 \left(\sum_{j=1}^{i-1} u_j f_j K_{nij} \psi_{ij}\right)^2$$

where $\sigma_i^2 = \sigma^2(W_i, X_i)$. We can decompose V_n^2 as

$$V_n^2 = \frac{4h^p}{(n-1)^2} \sum_{i=2}^{n} \sigma_i^2 f_i^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} u_j f_j u_k f_k K_{nij} K_{nik} \psi_{ij} \psi_{ik}$$

$$= \frac{4h^p}{(n-1)^2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \sigma_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2$$

$$+ \frac{8h^p}{(n-1)^2} \sum_{i=3}^{n} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \sigma_i^2 f_i^2 u_j f_j u_k f_k K_{nij} K_{nik} \psi_{ij} \psi_{ik} = A_n + B_n.$$

The result follows from the Central Limit Theorem for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). The conditions required for Corollary 3.1 of Hall and Heyde (1980), among which $V_n^2 \xrightarrow{P} \omega^2$, are checked in Lemma 2 below. Its proof is provided in the Appendix.

Lemma 2. *Under the conditions of Proposition 1,*

1. *$A_n \xrightarrow{p} \omega^2$,*
2. *$B_n \xrightarrow{p} 0$,*
3. *the martingale difference array $\{G_{n,i}, \mathcal{F}_{n,i}, 1 \le i \le n\}$ satisfies the Lindeberg condition*

$$\forall \varepsilon > 0, \quad \sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^2 I\left(|G_{n,i}| > \varepsilon\right) \mid \mathcal{F}_{n,i-1}\right] \xrightarrow{p} 0. \qquad \square$$

Proposition 2. *Under the conditions of Theorem 1 and H_{1n}, if $\delta_n^2 nh^{p/2} \to C$ with $0 < C < \infty$, $nh^{p/2}I_{0n} \xrightarrow{d} \mathcal{N}(C\mu, \omega^2)$.*

Proof. Let $\varepsilon_i = Y_i - \mathbb{E}[Y_i \mid W_i, X_i]$ and let us decompose

$$nh^{p/2}I_{0n} = \frac{h^{p/2}}{n-1} \sum_{i=1}^{n} \sum_{j \ne i} u_i f_i u_j f_j K_{nij} \psi_{ij}$$

$$= \frac{h^{p/2}}{n-1} \sum_{i=1}^{n} \sum_{j \ne i} (\delta_n d_i + \varepsilon_i) f_i (\delta_n d_j + \varepsilon_j) f_j K_{nij} \psi_{ij}$$

$$= \frac{h^{p/2}}{n-1} \sum_{i=1}^{n} \sum_{j \neq i} \varepsilon_i f_i \varepsilon_j f_j K_{nij} \psi_{ij}$$

$$+ \frac{\delta_n h^{p/2}}{n-1} \sum_{i=1}^{n} \sum_{j \neq i} d_i f_i \left(\delta_n d_j + 2\varepsilon_j \right) f_j K_{nij} \psi_{ij}$$

$$= C_{0n} + C_n.$$

By Proposition 1, $C_{0n} \xrightarrow{d} \mathcal{N}(0, \omega^2)$. As for C_n, we have

$$\mathbb{E}\left[C_n\right] = \delta_n^2 n h^{p/2} \mathbb{E}\left[d_i f_i d_j f_j K_{nij} \psi_{ij}\right] = \delta_n^2 n h^{p/2} \mu_n.$$

By repeated application of Fubini's Theorem, Fourier Inverse formula, Dominated Convergence Theorem, and Parseval's identity, we obtain

$$\mu_n = \mathbb{E}\left[d_1 f_2 d_1 f_2 K_{n12} \psi_{12}\right]$$

$$= \mathbb{E}\left[\iint \varrho\left(w_1, X_1\right) f\left(w_1\right) \varrho\left(w_2, X_2\right) f\left(w_2\right)\right.$$

$$\left. \times h^{-p} K\left(\frac{w_1 - w_2}{h}\right) dw_1 dw_2 \psi(X_1 - X_2)\right]$$

$$= \mathbb{E}\left[\int \mathcal{F}[\varrho\left(\cdot, X_1\right) f(\cdot)](t) \, \mathcal{F}[\varrho\left(\cdot, X_2\right) f(\cdot)](-t) \, \mathcal{F}\left[K\right](ht) \, dt \, \psi(X_1 - X_2)\right]$$

$$\to \mathbb{E}\left[\left[\int \mathcal{F}[\varrho\left(\cdot, X_1\right) f(\cdot)](t) \, \mathcal{F}[\varrho\left(\cdot, X_2\right) f(\cdot)](-t) \, dt\right] \psi(X_1 - X_2)\right]$$

$$= \mathbb{E}\left[\int \varrho\left(w, X_1\right) \varrho\left(w, X_2\right) f^2\left(w\right) \psi(X_1 - X_2) \, dw\right]$$

$$= \int \left[\int \mathcal{F}_X\left[\varrho\left(w, \cdot\right)\right](u) \mathcal{F}_X\left[\varrho\left(w, \cdot\right)\right](-u) \mathcal{F}[\psi](u) du\right] f^2\left(w\right) dw$$

$$= \iint \left|\mathcal{F}_X\left[d\left(w, \cdot\right) \pi\left(w \mid \cdot\right)\right](u)\right|^2 \mathcal{F}[\psi](u) f^2\left(w\right) dudw = \mu,$$

where $\varrho(w, x) = d(w, x)\pi(w|x)$. Moreover,

$$\text{Var}\left[C_n\right] \leq \frac{4\delta_n^4 h^p}{(n-1)^2} \sum_a \mathbb{E}\left[d_i^2 f_i^2 d_k d_l f_k f_l K_{nik} K_{nil} \psi_{ik} \psi_{il}\right]$$

$$+ \frac{2\delta_n^4 h^p}{(n-1)^2} \sum_a \mathbb{E}\left[d_i^2 f_i^2 d_k^2 f_k^2 K_{nik}^2 \psi_{ik}^2\right]$$

$$+ \frac{4\delta_n^2 h^p}{(n-1)^2} \sum_a \mathbb{E}\left[d_i f_i d_j f_j \varepsilon_k^2 f_k^2 K_{nik} K_{njk} \psi_{ik} \psi_{jk}\right]$$

$$+ \frac{4\delta_n^2 h^p}{(n-1)^2} \sum_a \mathbb{E}\left[d_i^2 f_i^2 \varepsilon_k^2 f_k^2 K_{nik}^2 \psi_{ik}^2\right]$$

$$= O\left(\delta_n^4 n h^p\right) + O\left(\delta_n^4\right) + O\left(\delta_n^2 n h^p\right) + O\left(\delta_n^2\right).$$

Therefore $C_n = C\mu_n + O_p(\delta_n n^{1/2} h^{p/2}) \xrightarrow{p} C\mu$, and the desired result follows. $\quad\square$

Proposition 3. *Under the conditions of Theorem [1],*

(i) $nh^{p/2}I_{1,3} = \delta_n\sqrt{n}h^{p/2}O_p(1) + o_p(1),$
(ii) $nh^{p/2}I_{1,5} = o_p(1),$
(iii) $nh^{p/2}I_{1,6} = \delta_n^2 nh^{p/2}o_p(1) + o_p(1),$
(iv) $nh^{p/2}I_{2,1} = \delta_n\sqrt{n}h^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}g^s O_p(1) + o_p(1),$
(v) $nh^{p/2}I_{2,3} = o_p(1),$
(vi) $nh^{p/2}I_3 = nh^{p/2}O_p(g^{2s}) + o_p(1).$

Proposition 4. *Under the conditions of Theorem [1],*

(i) $nh^{p/2}I_{1,1} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1),$
(ii) $nh^{p/2}I_{1,2} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1),$
(iii) $nh^{p/2}I_{1,4} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + (ng^p)^{-1/2}o_p(1) + o_p(1),$
(iv) $nh^{p/2}I_{2,2} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1).$

The proofs of the above propositions follow the ones in Lavergne and Vuong (2000)). For illustration, we provide in the Appendix the proofs of the first statements of each proposition.

Case $I_n = \widehat{I}_n$**.**   We have the following decomposition

$$n^{(4)}\tilde{I}_n = n(n-1)^3\,\widehat{I}_n - n^{(3)}V_{1n} - 2n^{(3)}V_{2n} + n^{(2)}V_{3n} \tag{2}$$

$$\text{where}\quad V_{1n} = \frac{1}{n^{(3)}}\sum_a (Y_i - Y_k)(Y_j - Y_k)\,L_{nik}L_{njk}K_{nij}\psi_{ij},$$

$$V_{2n} = \frac{1}{n^{(3)}}\sum_a (Y_i - Y_j)(Y_j - Y_k)\,L_{nij}L_{njk}K_{nij}\psi_{ij},$$

$$\text{and}\quad V_{3n} = \frac{1}{n^{(2)}}\sum_a (Y_i - Y_j)^2\,L_{nij}^2 K_{nij}\psi_{ij}.$$

Hence, to show that $\widehat{I}_n$ has the same asymptotic distribution as $\tilde{I}_n$, it is sufficient to investigate the behavior of V_{1n} to V_{3n}. Using $Y_i = r_i + u_i$, it is straightforward to see that the dominating terms in V_{1n}, V_{2n} and V_{3n} are

$$V_{13} = \frac{1}{n^{(3)}}\sum_a (r_i - r_k)(r_j - r_k)\,L_{nik}L_{njk}K_{nij}\psi_{ij},$$

$$V_{23} = \frac{1}{n^{(3)}}\sum_a (r_i - r_j)(r_j - r_k)\,L_{nij}L_{njk}K_{nij}\psi_{ij},$$

$$V_{33} = \frac{1}{n^{(2)}}\sum_a (r_i - r_j)^2\,L_{nij}^2 K_{nij}\psi_{ij},$$

respectively. Now

$$\begin{aligned}
\mathbb{E}\left[|V_{13}|\right] &= \mathbb{E}\left[|\,(r_i - r_k)(r_j - r_k)\,L_{nik}L_{njk}K_{nij}|\right]\\
&= O\left(g^{-p}\right)\mathbb{E}\left[|r_i - r_k|\,\mathbf{L}_{nik}\mathbb{E}\left[|r_j - r_k|\,\mathbf{K}_{nij} \mid Z_i, Z_k\right]\right] = O\left(g^{-p}\right),
\end{aligned}$$

$$\mathbb{E}\left[|V_{23}|\right] = \mathbb{E}\left[\left|\left(r_i - r_j\right)\left(r_j - r_k\right) L_{nij} L_{njk} K_{nij}\right|\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[|r_j - r_k|\, \mathbf{L}_{njk} \mid Z_j\right] |r_i - r_j|\, \mathbf{L}_{nij}\mathbf{K}_{nij}\right]$$
$$= o\left(1\right)\mathbb{E}\left[|r_i - r_j|\, \mathbf{L}_{nij}\mathbf{K}_{nij}\right] = o\left(g^{-p}\right)$$
$$\mathbb{E}\left[|V_{33}|\right] = \mathbb{E}\left[\left(r_i - r_j\right)^2 L_{nij}^2 |K_{nij}|\right]$$
$$= O\left(g^{-2p}\right)\mathbb{E}\left[\left(r_i - r_j\right)^2 \mathbf{K}_{nij}\right] = o\left(g^{-2p}\right).$$

It then follows that $nh^{p/2}(\tilde{I}_n - \widehat{I}_n) = O_p(h^{p/2}g^{-p})$ which is negligible if $h/g^2 \to 0$. The asymptotic irrelevance of the above diagonal terms thus require more restrictive relationships between the bandwidths h and g. For the sake of comparison, recall that Fan and Li (1996) impose $h^{(p+q)}g^{-2p} \to 0$ while Lavergne and Vuong (2000) require only $h^{p+q}g^{-p} \to 0$. Since we do not smooth the covariates X, we are able to further relax the restriction between the two bandwidths.

6.2. Proof of Corollary 1

It suffices to prove $\omega_n^2 - \omega^2 = o_p(1)$ with ω_n^2 any of $\widehat{\omega}_n^2$ or $\tilde{\omega}_n^2$. First we consider the case $\omega_n^2 = \widehat{\omega}_n^2$. A direct approach would consist in replacing the definition of $\hat{u}_i\hat{f}_i$ and $\hat{u}_j\hat{f}_j$, writing $\widehat{\omega}_n^2$ as a U-statistic of order 6, and studying its mean and variance. A shorter approach is based on empirical process tools. The price to pay is the stronger condition $n^{7/8}g^p/\ln n \to \infty$ instead of $ng^p \to \infty$. Let $\Delta\hat{f}_i = \hat{f}_i - f_i$, $\Delta\hat{r}_i\hat{f}_i = \hat{r}_i\hat{f}_i - r_if_i$, and write

$$\hat{u}_i\hat{f}_i = u_if_i + Y_i\Delta\hat{f}_i - \Delta\hat{r}_i\hat{f}_i. \tag{3}$$

Lemma 3. *Under Assumption 1, if $r(\cdot)f(\cdot) \in \mathcal{U}^p$, $L(\cdot)$ is a function of bounded variation, $g \to 0$, and $n^{7/8}g^p/\ln n \to \infty$, then*

$$\sup_{1\leq i\leq n} \left\{|\Delta\hat{r}_i\hat{f}_i| + |\Delta\hat{f}_i|\right\} = o_p(1).$$

The proof relies on the uniform convergence of empirical processes and is provided in the Appendix. Now proceed as follows: square Equation (3), replace $\hat{u}_i^2\hat{f}_i^2$ in the definition of $\widehat{\omega}_n^2$, and use Lemma 3 to deduce that

$$\widehat{\omega}_n^2 = \frac{2h^p}{n^{(2)}} \sum_{a(2)} u_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 + o_p(1).$$

Elementary calculations of mean and variance yield

$$\frac{2h^p}{n^{(2)}} \sum_{a(2)} u_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 - \omega^2 = o_p(1),$$

and thus $\widehat{\omega}_n^2 - \omega^2 = o_p(1)$.

To deal with $\tilde{\omega}_n^2$, note that $\tilde{\omega}_n^2 - \widehat{\omega}_n^2$ consists of "diagonal" terms plus a term which is $O(n^{-1}\tilde{\omega}_n^2)$. By tedious but rather straightforward calculations, one can check that such diagonal terms are each of the form $n^{-1}g^{-p}$ times a U-statistic which is bounded in probability. Hence $\tilde{\omega}_n^2 - \widehat{\omega}_n^2 = o_p(1)$.

6.3. Proof of Theorem 2

Let $\overline{Z}$ denote the sample (Y_i, W_i, X_i), $1 \leq i \leq n$. Since the limit distribution is continuous, it suffices to prove the result pointwise by Polya's theorem. Hence we show that $\forall t \in \mathbb{R}$, $\mathbb{P}[nh^{p/2}I_n^*/\omega_n^* \leq t \mid \overline{Z}] - \Phi(t) = o_p(1)$.

First, we consider the case $I_n^* = \tilde{I}_n$. Consider

$$
\begin{aligned}
I_{n,LV}^* &= \frac{1}{n^{(4)}} \sum_a \left(\eta_i \hat{u}_i - \eta_k \hat{u}_k\right)\left(\eta_j \hat{u}_j - \eta_l \hat{u}_l\right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&+ \frac{2}{n^{(4)}} \sum_a \left(\eta_i \hat{u}_i - \eta_k \hat{u}_k\right)\left(\hat{r}_j - \hat{r}_l\right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&+ \frac{1}{n^{(4)}} \sum_a \left(\hat{r}_i - \hat{r}_k\right)\left(\hat{r}_j - r_l\right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_1^* + 2I_2^* + I_3^*
\end{aligned}
$$

where we can further decompose

$$
\begin{aligned}
I_1^* &= \frac{1}{n^{(4)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&- \frac{2}{n^{(4)}} \sum_a \eta_j \hat{u}_j \eta_k \hat{u}_k L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&+ \frac{1}{n^{(4)}} \sum_a \eta_k \hat{u}_k \eta_l \hat{u}_l L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{1,1}^* + I_{1,2}^* + I_{1,3}^*
\end{aligned}
$$

with

$$
\begin{aligned}
I_{1,1}^* &= \frac{(n-1)^2}{(n-3)(n-4)} \times \frac{1}{n^{(2)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j \hat{f}_i \hat{f}_j K_{nij} \psi_{ij} \\
&- \frac{2}{n-4} \times \frac{1}{n^{(3)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{nij} K_{nij} \psi_{ij} \\
&- \frac{1}{n-4} \times \frac{1}{n^{(3)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{njk} K_{nij} \psi_{ij} \\
&- \frac{1}{(n-3)(n-4)} \times \frac{1}{n^{(2)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nij}^2 K_{nij} \psi_{ij} \\
&= I_{0n}^* - \frac{2}{n-4} I_{1,1,1}^* - \frac{1}{n-4} I_{1,1,2}^* - \frac{1}{(n-3)(n-4)} I_{1,1,3}^*.
\end{aligned}
$$

Now let $D_n^* = \tilde{I}_n^* - I_{0n}^*$ and write

$$
\mathbb{P}\left(\frac{nh^{p/2}\tilde{I}_n^*}{\tilde{\omega}_n^*} \leq t \mid \overline{Z}\right) = \mathbb{P}\left(\frac{nh^{p/2}\left(I_{0n}^* + D_n^*\right)}{\tilde{\omega}_n^*} \leq t \mid \overline{Z}\right)
$$

$$= \mathbb{P}\left(\frac{nh^{p/2}I_{0n}^*}{\widehat{\omega}_n} + \frac{nh^{p/2}D_n^*}{\widehat{\omega}_n} \right.$$
$$\left. + \frac{nh^{p/2}\left(I_{0n}^* + D_n^*\right)}{\widehat{\omega}_n}\left(\frac{\widetilde{\omega}_n}{\widehat{\omega}_n^*} - 1\right) \le t \mid \overline{Z} \right).$$

It thus suffices to prove that

$$\mathbb{P}\left(\frac{nh^{p/2}I_{0n}^*}{\hat{\omega}_{n,FL}} \le t \mid \overline{Z} \right) - \Phi\left(t\right) \xrightarrow{p} 0 \qquad \forall t \in \mathbb{R},$$

$$\frac{nh^{p/2}D_n^*}{\hat{\omega}_{n,FL}} = o_p(1), \qquad \text{and} \qquad \frac{nh^{p/2}\left(I_{0n}^* + D_n^*\right)}{\hat{\omega}_{n,FL}}\left(\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,LV}^*} - 1\right) = o_p\left(1\right).$$
$$(4)$$

The first result is stated below.

Proposition 5. *Under the conditions of Theorem 2, conditionally on the observed sample, the statistic $nh^{p/2}I_{0n}^*/\hat{\omega}_{n,FL}$ converges in law to a standard normal distribution.*

Proof. We proceed as in the proof of Proposition 1 and check the conditions for a CLT for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). Define the martingale array $\{S_{n,m}^*, \mathcal{F}_{n,m}^*, 1 \le m \le n, n \ge 1\}$ where $\mathcal{F}_{n,m}^*$ is the σ-field generated by $\{\overline{Z}, \eta_1, \ldots, \eta_m\}$, $S_{n,1}^* = 0$, and $S_{n,m}^* = \sum_{i=1}^m G_{n,i}^*$ with

$$G_{n,i}^* = \frac{2h^{p/2}}{n-1}\eta_i\hat{u}_i\sum_{j=1}^{i-1}\eta_j\hat{u}_j\hat{f}_i\hat{f}_jK_{nij}\psi_{ij}.$$

Then

$$I_{0n}^* = \frac{(n-1)^2}{(n-3)(n-4)} \times \frac{1}{n^{(2)}}\sum_a \eta_i\hat{u}_i\eta_j\hat{u}_j\hat{f}_i\hat{f}_jK_{nij}\psi_{ij} = \frac{(n-1)^2}{(n-3)(n-4)}S_{n,n}^*.$$

Now consider

$$V_n^{2*} = \sum_{i=2}^n \mathbb{E}\left[G_{n,i}^{2*} \mid \mathcal{F}_{n,i-1}^* \right]$$
$$= \frac{4h^p}{(n-1)^2}\sum_{i=2}^n\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}\hat{u}_i^2\eta_j\eta_k\hat{u}_j\hat{u}_k\hat{f}_i^2\hat{f}_j\hat{f}_kK_{nij}K_{nik}\psi_{ij}\psi_{ik}$$
$$= \frac{4h^p}{(n-1)^2}\sum_{i=2}^n\sum_{j=1}^{i-1}\hat{u}_i^2\eta_j^2\hat{u}_j^2\hat{f}_i^2\hat{f}_j^2K_{nij}^2\psi_{ij}^2$$
$$+ \frac{8h^p}{(n-1)^2}\sum_{i=3}^n\sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\hat{u}_i^2\eta_j\eta_k\hat{u}_j\hat{u}_k\hat{f}_i^2\hat{f}_j\hat{f}_kK_{nij}K_{nik}\psi_{ij}\psi_{ik}$$
$$= A_n^* + B_n^*.$$

Note that $\mathbb{E}[A_n^* \mid \overline{Z}] = [n/(n-1)]\mathbb{E}[\widehat{\omega}_n^2]$ and that

$$
\begin{aligned}
\mathrm{Var}\left[\tilde{A}_n^* \mid \overline{Z}\right] &\leq \frac{16h^{2p}\mathbb{E}\left[\eta^4\right]}{(n-1)^4} \sum_{i=2}^{n}\sum_{i'=2}^{n}\sum_{j=1}^{i\wedge i'-1} \hat{u}_i^2\hat{u}_{i'}^2\hat{u}_j^4\hat{f}_i^2\hat{f}_{i'}^2\hat{f}_j^4 K_{nij}^2 K_{ni'j}^2 \psi_{ij}^2\psi_{i'j}^2 \\
&\leq \frac{16h^{2p}\mathbb{E}\left[\eta^4\right]}{(n-1)^4} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \hat{u}_i^4\hat{u}_j^4\hat{f}_i^4\hat{f}_j^4 K_{nij}^4 \psi_{ij}^4 \\
&\quad+ \frac{32h^{2p}\mathbb{E}\left[\eta^4\right]}{(n-1)^4} \sum_{i=3}^{n}\sum_{i'=2}^{i-1}\sum_{j=1}^{i'-1} \hat{u}_i^2\hat{u}_{i'}^2\hat{u}_j^4\hat{f}_i^2\hat{f}_{i'}^2\hat{f}_j^4 K_{nij}^2 K_{ni'j}^2 \psi_{ij}^2\psi_{i'j}^2 \\
&= Q_{1n} + Q_{2n}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\mathbb{E}[B_n^{*2} \mid \overline{Z}] \\
&= \frac{64h^{2p}}{(n-1)^4}\sum_{i=3}^{n}\sum_{i'=3}^{n}\sum_{j=2}^{i\wedge i'-1}\sum_{k=1}^{j-1} \hat{u}_i^2\hat{u}_{i'}^2\hat{u}_j^2\hat{u}_k^2\hat{f}_i^2\hat{f}_{i'}^2\hat{f}_j^2\hat{f}_k^2 K_{nij}K_{ni'j}K_{nik}K_{ni'k}\psi_{ij}\psi_{i'j}\psi_{ik}\psi_{i'k} \\
&= \frac{64h^{2p}}{(n-1)^4}\sum_{i=3}^{n}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1} \hat{u}_i^4\hat{u}_j^2\hat{u}_k^2\hat{f}_i^4\hat{f}_j^2\hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2\psi_{ik}^2 \\
&\quad+ \frac{128h^{2p}}{(n-1)^4}\sum_{i=4}^{n}\sum_{i'=3}^{i-1}\sum_{j=2}^{i'-1}\sum_{k=1}^{j-1} \hat{u}_i^2\hat{u}_{i'}^2\hat{u}_j^2\hat{u}_k^2\hat{f}_i^2\hat{f}_{i'}^2\hat{f}_j^2\hat{f}_k^2 K_{nij}K_{ni'j}K_{nik}K_{ni'k}\psi_{ij}\psi_{i'j}\psi_{ik}\psi_{i'k} \\
&= Q_{3n} + Q_{4n}.
\end{aligned}
$$

Finally the Lindeberg condition involves

$$
\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\left[G_{n,i}^{2*}I\left(\left|G_{n,i}^*\right| > \varepsilon\right) \mid \mathcal{F}_{n,i-1}^*\right] \\
&\leq \frac{1}{\varepsilon^4}\sum_{i=1}^{n}\mathbb{E}\left[G_{n,i}^{4*} \mid \mathcal{F}_{n,i-1}^*\right] \\
&\leq \frac{16h^{2p}\mathbb{E}\left[\eta^4\right]}{\varepsilon^4(n-1)^4}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \hat{u}_i^4\hat{u}_j^2\hat{u}_k^2\hat{f}_i^4\hat{f}_j^2\hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2\psi_{ik}^2 \\
&\leq \frac{16h^{2p}\mathbb{E}\left[\eta^4\right]}{\varepsilon^4(n-1)^4}\sum_{i=2}^{n}\sum_{j=1}^{i-1} \hat{u}_i^4\hat{u}_j^4\hat{f}_i^4\hat{f}_j^4 K_{nij}^4 \psi_{ij}^4 \\
&\quad+ \frac{32h^{2p}\mathbb{E}\left[\eta^4\right]}{\varepsilon^4(n-1)^4}\sum_{i=3}^{n}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1} \hat{u}_i^4\hat{u}_j^2\hat{u}_k^2\hat{f}_i^4\hat{f}_j^2\hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2\psi_{ik}^2 \\
&= Q_{5n} + Q_{6n}.
\end{aligned}
$$

It thus suffices to show that $Q_{jn} = o_p(1)$, $j = 1,\ldots 6$. Now, there exist positive random variables $\tilde{\gamma}_{1n}$ and $\tilde{\gamma}_{2n}$ such that $\tilde{\gamma}_{1n} + \tilde{\gamma}_{2n} = o_p(1)$ and $\forall 1 \leq i \leq n$ and

$\forall k = 1, 2 \in \{1, 2\}$,

$$\hat{u}_i^{2k} \hat{f}_i^{2k} \leq 3^{2k-1} \left(u_i^{2k} f_i^{2k} + Y_i^{2k} f_i^{2k} \tilde{\gamma}_{1n}^{2k} + \tilde{\gamma}_{2n}^{2k} \right).$$

Indeed, $\hat{u}_i \hat{f}_i = u_i f_i + Y_i f_i f_i^{-1} (\hat{f}_i - f_i) + [\hat{r}_i \hat{f}_i - r_i f_i] = u_i f_i + Y_i f_i \gamma_{1i} - \gamma_{2i}$, where $\sup_{1 \leq i \leq n} |\gamma_{ji}| \leq \tilde{\gamma}_j$ and $\tilde{\gamma}_j = o_p(1)$ by Lemma 3. Hence

$$\hat{u}_i^2 \hat{f}_i^2 \leq 3 \left(u_i^2 f_i^2 + Y_i^2 f_i^2 \tilde{\gamma}_{1n}^2 + \tilde{\gamma}_{2n}^2 \right).$$

The inequality for $k = 2$ is obtained similarly. Using these inequalities, one can bound the expectations of $|Q_{1n}|$ to $|Q_{6n}|$ and thus show that $|Q_{1n}| + \cdots + |Q_{6n}| = o_p(1)$. $\square$

Next we show (4). First we need the following.

Proposition 6. *Under the conditions of Theorem 2*, $\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,FL}^*} \xrightarrow{p} 1$ *and* $\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,LV}^*} \xrightarrow{p} 1$.

The proof uses the following result, which is proved in the Appendix.

Lemma 4. *Under the conditions of Theorem 2*, $\sup_{1 \leq i \leq n} |\hat{u}_i^* \hat{f}_i - u_i^* \hat{f}_i| = o_p(1)$, *where* $u_i^* = \eta_i \hat{u}_i$ *and*

$$\hat{u}_i^* = Y_i^* - \frac{\sum_{k \neq i} Y_k^* L_{nik}}{\sum_{k \neq i} L_{nik}}.$$

Proof. Using Lemma 4, we have

$$\hat{\omega}_{n,FL}^{*2} = \omega_n^{*2} + o_p(1)$$

where $\omega_n^{*2} = \frac{2h^p}{n^{(2)}} \sum_a u_i^{*2} u_j^{*2} \hat{f}_i^2 \hat{f}_j^2 K_{nij}^2 \psi_{ij}^2$. Notice that $\mathbb{E}[\omega_n^{*2} \mid \overline{Z}] = \hat{\omega}_{n,FL}^2$ and that

$$\text{Var}\left(\omega_n^{*2} - \hat{\omega}_{n,FL}^2 \right) = \text{Var}\left(\mathbb{E}\left[\omega_n^{*2} - \hat{\omega}_{n,FL}^2 \mid \overline{Z} \right] \right) + \mathbb{E}\left[\text{Var}\left(\omega_n^{*2} \mid \overline{Z} \right) \right]$$

where the first term is zero and

$$\text{Var}\left(\omega_n^{*2} \mid \overline{Z} \right) = \frac{8h^{2p} \text{Var}\left(\eta^2 \right)}{\left\{ n^{(2)} \right\}^2} \sum_a \hat{u}_i^4 \hat{u}_j^4 \hat{f}_i^4 \hat{f}_j^4 K_{nij}^4 \psi_{ij}^4.$$

Then,

$$\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,FL}^*} = 1 + \frac{\hat{\omega}_{n,FL} - \hat{\omega}_{n,FL}^*}{\hat{\omega}_{n,FL}^*} = 1 + \frac{o_p(1)}{\omega^2 [1 + o_p(1)]} = 1 + o_p(1).$$

Since $\hat{\omega}_{n,LV}^* - \hat{\omega}_{n,FL}^*$ contains only diagonal terms, we deduce that $\hat{\omega}_{n,FL} / \hat{\omega}_{n,LV}^* \xrightarrow{p} 1$. $\square$

We next have to bound $D_n^* = I_{n,LV}^* - I_{0n}^*$. For this, let us decompose

$$\hat{r}_i - \hat{r}_k = (\hat{r}_i - r_i) - (\hat{r}_k - r_k) + (r_i - r_k)$$

and replace all such differences appearing in the definition of D_n^*. First, let us look at I_3^* which does not contain any bootstrap variable η. We obtain

$$
\begin{aligned}
I_3^* &= \frac{1}{n^{(4)}} \sum_a \left(\hat{r}_i - \hat{r}_k \right) \left(\hat{r}_j - \hat{r}_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= \frac{1}{n^{(4)}} \sum_a \left(r_i - r_k \right) \left(r_j - r_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a \left(\hat{r}_i - r_i \right) \left(\hat{r}_j - r_j \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a \left(\hat{r}_k - r_k \right) \left(\hat{r}_l - r_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{2}{n^{(4)}} \sum_a \left(\hat{r}_i - r_i \right) \left(r_j - r_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n^{(4)}} \sum_a \left(\hat{r}_k - r_k \right) \left(r_j - r_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n^{(4)}} \sum_a \left(\hat{r}_k - r_k \right) \left(\hat{r}_j - r_j \right) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{3,1}^* + I_{3,2}^* + I_{3,3}^* + 2I_{3,4}^* - 2I_{3,5}^* - 2I_{3,6}^*.
\end{aligned}
$$

Next, use the fact that

$$
\begin{aligned}
\hat{r}_i - r_i &= (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} \left(Y_{i'} - r_i \right) L_{nii'} \\
&= (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} \left(r_{i'} - r_i \right) L_{nii'} + (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} u_{i'} L_{nii'} \quad (5)
\end{aligned}
$$

and further replace terms like $\hat{r}_i - r_i$. Among the terms $I_{3,1}^*$ to $I_{3,6}^*$, the term $I_{3,1}^*$ could be easily handled with existing results in Lavergne and Vuong (2000). Namely $nh^{p/2}I_{3,1}^* = nh^{p/2}O_p(g^{2s}) + o_p(1)$ by Proposition 7 of Lavergne and Vuong (2000). For the other five terms we have to control the density estimates appearing in the denominators. For this purpose, let us introduce the notation $\Delta(f_i^I)^{-1} = (\hat{f}_i^I)^{-1} - f_i^{-1}$ and write

$$
\begin{aligned}
\frac{n - |I|}{n - 1} \times \hat{f}_i^{-1} &= \left(\frac{(n - |I|) \hat{f}_i^I}{(n-1) \hat{f}_i} - 1 \right) \left(\hat{f}_i^I \right)^{-1} + \left(\hat{f}_i^I \right)^{-1} \\
&= \frac{\sum_{k \in I} L_{nik}}{(n-1) \hat{f}_i \hat{f}_i^I} + \Delta \left(f_i^I \right)^{-1} + f_i^{-1}. \quad (6)
\end{aligned}
$$

Then we obtain for instance

$$
I_{3,5}^* = \frac{1}{n^{(4)}} \sum_a \left(\hat{r}_k - r_k \right) \left(r_j - r_l \right) L_{nik} L_{njl} K_{nij} \psi_{ij}
$$

$$= \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} f_k^{-1} \left(r_{k'} - r_k \right) \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \Delta \left(f_k^{i,j,l,k'} \right)^{-1} \left(r_{k'} - r_k \right) \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{(n-1) \, n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \left(\hat{f}_k \hat{f}_k^{i,j,l,k'} \right)^{-1} \left(L_{nik} + L_{njk} + L_{nlk} + L_{nk'k} \right)$$

$$\times \left(r_{k'} - r_k \right) \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} f_k^{-1} u_{k'} \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \Delta f_k^{-1} u_{k'} \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{(n-1) \, n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \left(\hat{f}_k \hat{f}_k^{i,j,l,k'} \right)^{-1} \left(L_{nik} + L_{njk} + L_{nlk} + L_{nk'k} \right)$$

$$\times u_{k'} \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$= I_{3,5,1}^* + I_{3,5,2}^* + I_{3,5,3}^* + I_{3,5,4}^* + I_{3,5,5}^* + I_{3,5,6}^*.$$

Next, if we consider for instance $I_{3,5,1}^*$ that contains only terms like f_i^{-1} appearing from the decomposition 6, we obtain

$$I_{3,5,1}^* = \frac{1}{n^{(5)}} \sum_{a(5)} f_k^{-1} \left(r_{k'} - r_k \right) \left(r_j - r_l \right) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} \left(r_i - r_k \right) \left(r_j - r_l \right) L_{nik}^2 L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} \left(r_j - r_k \right) \left(r_j - r_l \right) L_{njk} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$+ \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} \left(r_l - r_k \right) \left(r_j - r_l \right) L_{nlk} L_{nik} L_{njl} K_{nij} \psi_{ij}$$

$$= I_{3,5,1,1}^* + I_{3,5,1,2}^* + I_{3,5,1,3}^* + I_{3,5,1,4}^*$$

where the terms $I_{3,5,1,2}^*$ to $I_{3,5,1,4}^*$ are called "diagonal terms". Such terms require more restrictions on the bandwidths. Next, the terms containing terms like $\Delta (f_i^I)^{-1}$ produced by the decomposition (6) can be treated like in the Propositions 8 to 11 of Lavergne and Vuong (2000). Finally, given that I is finite and with fixed cardinal

$$(n-1)^{-1} \hat{f}_i^{-1} \left(\hat{f}_i^I \right)^{-1} \sum_{k \in I} L_{nik} = O_p \left(n^{-1} g^{-p} \right) = o_p(1)$$

given that $\| f^{-1} \|_\infty < \infty$. Therefore we can be easily handle the terms containing $(n-1)^{-1} \hat{f}_i^{-1} (\hat{f}_i^I)^{-1} \sum_{k \in I} L_{nik}$ by taking absolute values. Now let us investigate

the diagonal term $I_{3,5,1,2}^*$. We have

$$
\begin{aligned}
\mathbb{E}\left[|I_{3,5,1,2}^*|\right] &= O\left(n^{-1}\right)\mathbb{E}\left[f_k^{-1}\left|r_j - r_k\right|\left|r_j - r_l\right||L_{njk}||L_{nik}||L_{njl}||K_{nij}|\right] \\
&= O\left(n^{-1}g^{-p}\right)\mathbb{E}\left[f_k^{-1}\left|r_j - r_k\right|\left|r_j - r_l\right||L_{njl}||L_{njl}||K_{nij}|\right] \\
&= O\left(n^{-1}g^{-p}\right)\mathbb{E}\left[f_k^{-1}\left|r_j - r_k\right||L_{njk}|\mathbb{E}\left[\left|r_j - r_l\right||L_{njl}|\mid Z_j\right]|K_{nij}|\right] \\
&= o\left(n^{-1}g^{-p}\right)\mathbb{E}\left[f_k^{-1}\left|r_j - r_k\right||L_{njk}||K_{nij}|\right] \\
&= o\left(n^{-1}g^{-p}\right).
\end{aligned}
$$

To prove that he term $I_{3,5,1,2}^* = o_p(nh^{p/2})$ it suffices to prove $\mathbb{E}[|I_{3,5,1,2}^*|] = o(nh^{p/2})$ and this latter rate is implied by the condition $h/g^2 = o(1)$. This additional condition on the bandwidths is not surprising as the bootstrapped statistic introduced "diagonal" terms as in Fan and Li (1996) which indeed require the condition $h/g^2 \to 0$.

Let us now consider a term in the decomposition of D_n^* that involve bootstrap variables η, namely we investigate I_2^*. The arguments for the other terms are similar. Consider

$$
\begin{aligned}
I_2^* &= \frac{1}{n^{(4)}}\sum_a \left(\eta_i\hat{u}_i - \eta_k\hat{u}_k\right)\left(\hat{r}_j - \hat{r}_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&= \frac{1}{n^{(4)}}\sum_a \eta_i\hat{u}_i\left(r_j - r_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&\quad + \frac{1}{n^{(4)}}\sum_a \eta_i\hat{u}_i\left(\hat{r}_j - r_j\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&\quad - \frac{1}{n^{(4)}}\sum_a \eta_i\hat{u}_i\left(\hat{r}_l - r_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&\quad - \frac{1}{n^{(4)}}\sum_a \eta_k\hat{u}_k\left(r_j - r_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&\quad - \frac{1}{n^{(4)}}\sum_a \eta_k\hat{u}_k\left(\hat{r}_j - r_j\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&\quad + \frac{1}{n^{(4)}}\sum_a \eta_k\hat{u}_k\left(\hat{r}_l - r_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij} \\
&= I_{2,1}^* + I_{2,2}^* - I_{2,3}^* - I_{2,4}^* - I_{2,5}^* + I_{2,6}^*.
\end{aligned}
$$

Next it suffices to use the fact that

$$
\hat{u}_i = u_i - \hat{f}_i^{-1}\sum_{i'\neq i}u_{i'}L_{nii'} + \hat{f}_i^{-1}\sum_{i'\neq i}\left(r_i - r_{i'}\right)L_{nii'}.
$$

For instance, using this identity with $I_{2,1}^*$ we can write

$$
I_{2,1}^* = \frac{1}{n^{(4)}}\sum_a \eta_i u_i\left(r_j - r_l\right)L_{nik}L_{njl}K_{nij}\psi_{ij}
$$

$$-\frac{1}{(n-1)\,n^{(4)}}\sum_a\sum_{i'\neq i}\hat{f}_i^{-1}\eta_i u_{i'}\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$+\frac{1}{(n-1)\,n^{(4)}}\sum_a\sum_{i'\neq i}\hat{f}_i^{-1}\eta_i\,(r_i-r_{i'})\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$=\frac{1}{n^{(3)}}\sum_a \eta_i u_i f_i\,(r_j-r_l)\,L_{njl}K_{nij}\psi_{ij}$$

$$+\frac{1}{n^{(4)}}\sum_a \eta_i u_i\,(r_j-r_l)\,\Delta f_i^{j,l}L_{njl}K_{nij}\psi_{ij}$$

$$-\frac{1}{(n-1)\,n^{(4)}}\sum_a\sum_{i'\neq i}f_i^{-1}\eta_i u_{i'}\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$-\frac{1}{n^{(5)}}\sum_a \Delta\left(f_i^{j,k,l,i'}\right)^{-1}\eta_i u_{i'}\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$-\frac{1}{(n-1)\,n^{(4)}}\sum_a\sum_{i'\neq i}\left(\hat{f}_i\hat{f}_i^{j,k,l,i'}\right)^{-1}\eta_i u_{i'}\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$+\frac{1}{(n-1)\,n^{(4)}}\sum_a\sum_{i'\neq i}f_i^{-1}\eta_i\,(r_i-r_{i'})\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$+\frac{1}{n^{(5)}}\sum_a \Delta\left(f_i^{j,k,l,i'}\right)^{-1}\eta_i\,(r_i-r_{i'})\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$+\frac{1}{n^{(5)}}\sum_a \left(\hat{f}_i\hat{f}_i^{j,k,l,i'}\right)^{-1}\eta_i\,(r_i-r_{i'})\,(r_j-r_l)\,L_{nik}L_{njl}K_{nij}\psi_{ij}$$

$$=I_{2,1,1}^*+I_{2,1,2}^*+I_{2,1,3}^*+I_{2,1,4}^*+I_{2,1,5}^*+I_{2,1,6}^*+I_{2,1,7}^*+I_{2,1,8}^*$$

Handling one problem at a time, let us notice that $I_{2,1,1}^*$ is a zero-mean U-statistic of order three with kernel $H_n(Z_i^*,Z_j^*,Z_l^*)=\eta_i u_i f_i(r_j-r_l)L_{njl}K_{nij}\psi_{ij}$ where $Z_i^*=(Y_i,W_i,X_i,\eta_i)$. Using the Hoeffding decomposition of $I_{2,1,1}^*$ in degenerate U-statistics, it is easy to check that the third and second order projections are small. For the first order degenerate U-statistic it suffices to note that $\mathbb{E}[H_n\mid Z_j^*]=\mathbb{E}[H_n\mid Z_l^*]=0$ and $\mathbb{E}[H_n\mid Z_i^*]=\eta_i u_i f_i\mathbb{E}[(r_j-r_l)L_{njl}K_{nij}\psi_{ij}\mid Z_i]$ so that

$$\mathbb{E}\left[\mathbb{E}^2\left[H_n\mid Z_i^*\right]\right]=\mathbb{E}\left[\eta_i^2 u_i^2 f_i^2\mathbb{E}^2\left[(r_j-r_l)\,L_{njl}K_{nij}\psi_{ij}\mid Z_i\right]\right]$$
$$=\mathbb{E}\left[u_i^2 f_i^2\mathbb{E}^2\left[(r_j-r_l)\,L_{njl}K_{nij}\psi_{ij}\mid Z_i\right]\right]$$

which, given that $\|\psi\|_\infty<\infty$, is similar to the term ξ_1 bounded in the proof of Proposition 5 of Lavergne and Vuong (2000).

Finally, let us briefly consider the case $I_n^*=\tilde{I}_n$. Like in the decomposition (2), we have

$$n\,(n-1)^3\,I_{n,FL}^*=n^{(4)}I_{n,LV}^*+n^{(3)}V_{1n}^*+2n^{(3)}V_{2n}^*-n^{(2)}V_{3n}^*$$

where $\forall j\in\{1,2,3\}$, the V_{jn}^*s are obtained by replacing the Y_is by the Y_i^*s in the V_{jn}s. All these terms could be handled by arguments similar to the ones detailed above. The proof of Theorem 2 is now complete.

Appendix: Additional proofs

We here provide proofs of technical lemmas and additional details for the proofs in the manuscript. We define $Z_i = (Y_i, W_i, X_i)$, $\|\psi\|_\infty = \sup_{x \in \mathbb{R}^q} |\psi(x)|$,

$$\mathbf{K}_{nij} = |K_{nij}| = \frac{1}{h^p} \left| K \left(\frac{W_i - W_j}{h} \right) \right|,$$

and

$$\mathbf{L}_{nij} = |L_{nij}| = \frac{1}{g^p} \left| L \left(\frac{W_i - W_j}{g} \right) \right|.$$

Proof of Lemma 2. 1. We have

$$\mathbb{E}\left[A_n\right] = \frac{4h^p}{(n-1)^2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2\right] = \frac{2nh^p}{n-1} \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2\right],$$

and

$$\begin{aligned}
\mathrm{Var}\left[A_n\right] &\leq \frac{64 h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^{n} \sum_{j=2}^{i-1} \sum_{j'=1}^{j-1} \mathbb{E}\left[\sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_{j'}^2 f_{j'}^2 K_{nij}^2 K_{nij'}^2\right] \\
&\quad + \frac{32 h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^{n} \sum_{i'=1}^{i-1} \sum_{j=2}^{i'-1} \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j^4 f_j^4 K_{nij}^2 K_{ni'j}^2\right] \\
&\quad + \frac{16 h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[\sigma_i^4 f_i^4 u_j^4 f_j^4 K_{nij}^4\right] \\
&= O\left(n^{-1}\right) \mathbb{E}\left[\sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij} \mathbf{K}_{nik}\right] \\
&\quad + O\left(n^{-1}\right) \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j^4 f_j^4 \mathbf{K}_{nij} \mathbf{K}_{ni'j}\right] \\
&\quad + O\left(n^{-2}h^{-p}\right) \mathbb{E}\left[\sigma_i^4 f_i^4 u_j^4 f_j^4 \mathbf{K}_{nij}\right] \\
&= O\left(n^{-1}\right) + O\left(n^{-2}h^{-p}\right).
\end{aligned}$$

Deduce that $\mathrm{Var}[A_n] \to 0$, and hence remains to show that $\mathbb{E}[A_n] \to \omega^2$. We have

$$h^p \, \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2\right] = \mathbb{E}\left[\int \varphi_{X_i}(t) \, \varphi_{X_j}(-t) \, \mathcal{F}\left[K^2\right](ht) \, \psi^2 \left(X_i - X_j\right) dt\right]$$

where $\varphi_x(t) = \mathcal{F}[\sigma^2(\cdot, x) f^2(\cdot) \pi(\cdot \mid x)](t)$. Let us note that

$$\begin{aligned}
\mathbb{E}\left[\int \left|\varphi_{X_i}(t) \, \varphi_{X_j}(-t)\right| \psi^2(X_i - X_j) \, dt\right] &\leq \|\psi\|_\infty \, \mathbb{E}\left[\int \left|\varphi_X(t)\right|^2 dt\right] \\
&= \|\psi\|_\infty \, \mathbb{E}\left[\sigma^4(W, X) f^4(W) \pi(W \mid X)\right],
\end{aligned}$$

by Plancherel Theorem. Moreover, $\mathcal{F}[K^2](ht)$ is bounded and converges pointwise to $\int K^2(s) ds$ as $h \to 0$. Then by Lebesgue's dominated convergence theo-

rem,

$$h^p \, \mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2\right] \to \mathbb{E}\left[\int \varphi_{X_i}(t)\,\varphi_{X_j}(-t)\,\psi^2\,(X_i - X_j)\,dt\right]\int K^2\,(s)\,ds$$
$$= \omega^2,$$

by Parseval's Theorem.

2. By elementary calculations,

$$\mathbb{E}\left[B_n^2\right]$$

$$= \frac{64h^{2p}}{(n-1)^4}\sum_{i=3}^{n}\sum_{i'=3}^{n}\sum_{j=2}^{i-1}\sum_{j'=2}^{i'-1}\sum_{k=1}^{j-1}\sum_{k'=1}^{j'-1}\mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j f_j u_{j'} f_{j'} u_k f_k u_{k'} f_{k'}\right.$$

$$\left. \times K_{nij}K_{ni'j'}K_{nik}K_{ni'k'}\psi_{ij}\psi_{i'j'}\psi_{ik}\psi_{i'k'}\right]$$

$$\leq \frac{64h^{2p}\left\|\psi\right\|_\infty^4}{(n-1)^4}\sum_{i=3}^{n}\sum_{i'=3}^{n}\sum_{j=2}^{i\wedge i'-1}\sum_{k=1}^{j-1}\mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij}K_{ni'j}K_{nik}K_{ni'k}\right]$$

$$= \frac{64h^{2p}\left\|\psi\right\|_\infty^4}{(n-1)^4}\sum_{i=3}^{n}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1}\mathbb{E}\left[\sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij}^2 K_{nik}^2\right]$$

$$+ \frac{128h^{2p}\left\|\psi\right\|_\infty^4}{(n-1)^4}\sum_{i=3}^{n}\sum_{i'=3}^{i-1}\sum_{j=2}^{i'-1}\sum_{k=1}^{j-1}\mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij}K_{ni'j}K_{nik}K_{ni'k}\right]$$

$$= O\left(n^{-1}\right)\mathbb{E}\left[\sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij}\mathbf{K}_{nik}\right]$$

$$+ O\left(h^p\right)\mathbb{E}\left[\sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij}\mathbf{K}_{ni'j}\mathbf{K}_{nik}\right]$$

$$= O\left(n^{-1}\right) + O\left(h^p\right) = o(1).$$

3. We have $\forall \varepsilon > 0$, $\forall n \geq 1$, and $1 < i \leq n$,

$$\mathbb{E}\left[G_{n,i}^2 I\left(|G_{n,i}| > \varepsilon\right)|\mathcal{F}_{n,i-1}\right] \leq \mathbb{E}^{1/2}\left[G_{n,i}^4|\mathcal{F}_{n,i-1}\right]\mathbb{E}^{1/2}\left[I\left(|G_{n,i}| > \varepsilon\right)|\mathcal{F}_{n,i-1}\right]$$
$$\leq \frac{\mathbb{E}\left[G_{n,i}^4|\mathcal{F}_{n,i-1}\right]}{\varepsilon^2}.$$

Then

$$\sum_{i=2}^{n}\mathbb{E}\left[G_{n,i}^2 I\left(|G_{n,i}| > \varepsilon\right)|\mathcal{F}_{n,i-1}\right]$$

$$\leq \frac{1}{\varepsilon^2}\sum_{i=2}^{n}\mathbb{E}\left[G_{n,i}^4|\mathcal{F}_{n,i-1}\right]$$

$$\leq \frac{1}{\varepsilon^2}\frac{16h^{2p}}{(n-1)^4}\sum_{i=2}^{n}\mathbb{E}\left[u_i^4 f_i^4|W_i,\,X_i\right]\left(\sum_{j=1}^{i-1}u_j K_{nij}\psi_{ij}\right)^4$$

$$\leq \frac{1}{\varepsilon^2} \frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^{n} \left(\sum_{j=1}^{i-1} u_j K_{nij} \psi_{ij} \right)^4,$$

where κ_4 is any constant that bounds $\mathbb{E}[u^4 f^4 \,|\, W, X]$. The last expression that multiplies ε^{-2} is positive and has expectation

$$\frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^{n} \sum_{j_1=1}^{i-1} \sum_{j_2=1}^{i-1} \sum_{j_3=1}^{i-1} \sum_{j_4=1}^{i-1} \mathbb{E}\left[u_{j_1} f_{j_1} u_{j_2} f_{j_2} u_{j_3} j_{j_3} u_{j_4} f_{j_4} \right.$$
$$\left. \times K_{nij_1} K_{nij_2} K_{nij_3} K_{nij_4} \psi_{ij_1} \psi_{ij_2} \psi_{ij_3} \psi_{ij_4} \right]$$

$$= \frac{96\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=3}^{n} \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \mathbb{E}\left[u_j^2 f_j^2 u_k^2 f_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2 \psi_{ik}^2 \right]$$

$$+ \frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[u_j^4 f_j^4 K_{nij}^4 \psi_{ij}^4 \right]$$

$$= O\left(n^{-1}\right) \mathbb{E}\left[u_j^2 f_j^2 u_k^2 f_k^2 \mathbf{K}_{nij} \mathbf{K}_{nik} \right] + O\left(n^{-2} h^{-p}\right) \mathbb{E}\left[u_j^4 f_j^4 \mathbf{K}_{nij} \right]$$

$$= O\left(n^{-1}\right) + O\left(n^{-2} h^{-p}\right).$$

The desired result follows. $\square$

The following result, known as Bochner's Lemma (see Theorem 1.1.1. of Bochner (1955)) will be repeatedly use in the following. We recall it for the sake of completeness.

Lemma 5. *For any function $l(\cdot) \in \mathcal{U}^p$ and any integrable kernel $K(\cdot)$,*

$$\sup_{x \in \mathbb{R}^p} \left| \int l(y) \frac{1}{h^p} K\left(\frac{x-y}{h} \right) dy - l(x) \int K(u) \, du \right| \to 0.$$

In the following we provide the proofs for rates for the remaining terms in the decomposition of I_n, see Propositions 3 and 4. For this purpose, we use the following a decomposition for $U-$statistics that can be found in Lavergne and Vuong (2000): if $U_n = (1/n^{(m)}) \sum_a H_n(Z_{i_1}, \ldots, Z_{i_m})$, then

$$\mathbb{E}\left[U_n^2 \right] = \left(\frac{1}{n^{(m)}} \right)^2 \sum_{c=0}^{m} \frac{n^{(2m-c)}}{c!} \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1, \Delta_2)$$

$$= \sum_{c=0}^{m} O\left(n^{-c}\right) \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1, \Delta_2),$$

where $\sum^{(c)}$ denotes summation over sets Δ_1 and Δ_1 of ordered positions of length c,

$$I(\Delta_1, \Delta_2) = \mathbb{E}\left[H_n(Z_{i_1}, \ldots, Z_{i_m}) H_n(Z_{j_1}, \ldots, Z_{j_m}) \right]$$

and the i's position in Δ_1 coincide with the j's position in Δ_2 and are pairwise distinct otherwise. Now, we will bound $\mathbb{E}[U_n^2]$ using the $\xi_c = \sum^{(c)} I(\Delta_1, \Delta_2)$ and

the fact that by Cauchy's inequality,

$$I^2\left(\Delta_1, \Delta_2\right) = \mathbb{E}^2\left[\mathbb{E}\left[H_n\left(Z_{i_1}, \ldots, Z_{i_m}\right) \mid Z_c\right] \mathbb{E}\left[H_n\left(Z_{j_1}, \ldots, Z_{j_m}\right) \mid Z_c\right]\right]$$
$$\leq \mathbb{E}\left[\mathbb{E}^2\left[H_n\left(Z_{i_1}, \ldots, Z_{i_m}\right) \mid Z_c\right]\right] \mathbb{E}\left[\mathbb{E}^2\left[H_n\left(Z_{j_1}, \ldots, Z_{j_m}\right) \mid Z_c\right]\right]$$

where Z_c denotes the common Z_i's.

Proof of Proposition 3. After bounding the ψ_{ij}'s by $\|\psi\|_\infty$ the arguments are very similar to those used in Lavergne and Vuong (2000). We prove only the first statement.

(i) $I_{1,3}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_l) = u_i f_i u_l L_{njl} K_{nij} \psi_{ij}$. We need to bound the ξ_c, $c = 0, 1, 2, 3$.

1. $\mathbb{E}[H_n] = 0$, thus $\xi_0 = 0$.

2. $\xi_1 = O(\delta_n^2)$. Indeed, $\mathbb{E}[H_n \mid Z_l] = \delta_n u_l \mathbb{E}[d_i f_i L_{njl} K_{nij} \psi_{ij} \mid Z_l]$ and $\mathbb{E}[H_n \mid Z_i] = 0 = \mathbb{E}[H_n \mid Z_j]$. Then

$$\mathbb{E}\left[\mathbb{E}^2\left[H_n \mid Z_l\right]\right] \leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E}\left[u_l^2 \mathbb{E}^2\left[d_i f_i L_{njl} K_{nij} \mid Z_l\right]\right]$$
$$= O\left(\delta_n^2\right) \mathbb{E}\left[u_l^2 \mathbb{E}^2\left[L_{njl} d_j f_j^2 \mid Z_l\right]\right] = O\left(\delta_n^2\right).$$

3. $\xi_2 = O(g^{-p})$. Indeed, we have

$$\mathbb{E}\left[H_n \mid Z_i, Z_j\right] = u_i f_i K_{nij} \psi_{ij} \mathbb{E}\left[u_l L_{njl} \mid Z_j\right] = 0,$$
$$\mathbb{E}\left[H_n \mid Z_i, Z_l\right] = u_i f_i u_l \mathbb{E}\left[L_{njl} K_{nij} \psi_{ij} \mid Z_i, Z_l\right],$$
$$\mathbb{E}\left[H_n \mid Z_j, Z_l\right] = u_l L_{njl} \mathbb{E}\left[u_i f_i K_{nij} \psi_{ij} \mid Z_j\right]$$
$$= \delta_n u_l L_{njl} \mathbb{E}\left[d_i f_i K_{nij} \psi_{ij} \mid Z_j\right].$$

By successive applications of Lemma 5,

$$\mathbb{E}\left[\mathbb{E}^2\left(H_n \mid Z_i, Z_l\right)\right] \leq \|\psi\|_\infty^2 \mathbb{E}\left[u_i^2 f_i^2 u_l^2 \mathbb{E}\left[L_{njl} K_{nij} \mid Z_i, Z_l\right]\right.$$
$$\left. \times \mathbb{E}\left[L_{nj'l} K_{nij'} \mid Z_i, Z_l\right]\right]$$
$$= O\left(g^{-p}\right) \mathbb{E}\left[u_i^2 f_i^2 u_l^2 \mathbb{E}\left[\mathbf{L}_{njl} \mathbf{K}_{nij} \mid Z_i, Z_l\right]\right.$$
$$\left. \times \mathbb{E}\left[\mathbf{K}_{nij'} \mid Z_i, Z_l\right]\right]$$
$$= O\left(g^{-p}\right) \mathbb{E}\left[u_i^2 f_i^3 u_l^2 \mathbf{L}_{njl} \mathbf{K}_{nij}\right] = O\left(g^{-p}\right),$$
$$\mathbb{E}\left[\mathbb{E}^2\left[H_n \mid Z_j, Z_l\right]\right] \leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E}\left[u_l^2 L_{njl}^2 \mathbb{E}^2\left[d_i f_i K_{nij} \mid Z_j\right]\right]$$
$$\leq O\left(\delta_n^2\right) \mathbb{E}\left[u_l^2 L_{njl}^2 d_j^2 f_j^4\right]$$
$$= O\left(\delta_n^2\right) O\left(g^{-p}\right) \mathbb{E}\left[u_l^2 \mathbf{L}_{njl} d_j^2 f_j^4\right] = O\left(g^{-p}\right).$$

4. $\xi_3 = O(g^{-p} h^{-p})$, as $\mathbb{E}[H_n^2]$ equals

$$\mathbb{E}\left[u_i^2 u_l^2 f_i^2 L_{njl}^2 K_{nij}^2 \psi_{ij}^2\right] = O\left(g^{-p} h^{-p}\right) \mathbb{E}\left[u_i^2 u_l^2 f_i^2 \mathbf{L}_{njl} \mathbf{K}_{nij}\right]$$
$$= O\left(g^{-p} h^{-p}\right).$$

Collecting results, $\mathbb{E}[(n h^{p/2} I_{1,3})^2] = O(\delta_n^2 n h^p) + O(h^p/g^p) + O(n^{-1} g^{-p}) = o(1)$. $\qquad\square$

Proof of Proposition 4. As in Proposition 3, we only prove the first statement. We will use the following lemma, which is similar to Lemma 2 of Lavergne and Vuong (2000), and whose proof is then omitted.

Lemma 6. *Let* $\Delta f_i^j = \widehat{f_i^j} - f_i$. *If* $f(\cdot) \in \mathcal{U}^p$ *and* $ng^p \to \infty$,

$$\mathbb{E}\left[\Delta^2 f_i^j \mid Z_i, Z_j, Z_{i'}, Z_{j'}\right] = o\,(1)$$

and

$$E\left[\Delta^2 f_i^{j,l} \mid Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'}\right] = o\,(1)$$

uniformly in the indices.

(i) Let us denote $\Delta f_i^j = \widehat{f_i^j} - f_i$. We have $I_{1,1} = (1/n^{(2)})\sum_a u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij}$ so that

$$\mathbb{E}\left[I_{1,1}^2\right] = \left(\frac{1}{n^{(2)}}\right)^2 \left[\sum_a u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij}\right]$$

$$\times \left[\sum_a u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'} \psi_{i'j'}\right] \qquad (7)$$

where the first (respectively the second) sum is taken over all arrangements of different indices i and j (respectively different indices i' and j'). Let $\overline{W}$ denote the sample of W_i, $1 \le i \le n$, and let $\lambda_n = E[\Delta^2 f_i^j \mid Z_i, Z_j, Z_{i'}, Z_{j'}]$. By Lemma 6, $\lambda_n = o(1)$ uniformly in the indices. By Equation (7), $\mathbb{E}[I_{1,1}^2]$ is equal to a normalized sum over four indices. This sum could split in three sums of the following types.

1. All indices are different, that is a sum of $n^{(4)}$ terms. Each term in the sum can be bounded as follows:

$$\mathbb{E}\left[u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij} u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'} \psi_{i'j'}\right]$$

$$\le \quad \|\psi\|_\infty^2 \delta_n^4 \mathbb{E}\left[\Delta f_i^j f_j \Delta f_{i'}^{j'} f_{j'} \mathbb{E}\left[d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} \mid \overline{W}\right]\right]$$

$$\le \quad \|\psi\|_\infty^2 \delta_n^4 \mathbb{E}\left[f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'}\right.$$

$$\left. \times \mathbb{E}\left[\Delta f_i^j \Delta f_{i'}^{j'} \mid Z_i, Z_j, Z_{i'}, Z_{j'}\right]\right]$$

$$\le \quad O(\delta_n^4 \lambda_n)\mathbb{E}\left|f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'}\right| = O\left(\delta_n^4 \lambda_n\right).$$

2. One index is common to $\{i,j\}$ and $\{i',j'\}$, that is a sum of $4n^{(3)}$ terms. For each of such terms we can write

$$(i' = i) \qquad \mathbb{E}\left[u_i^2 \Delta f_i^j u_j f_j K_{nij} \psi_{ij} \Delta f_i^{j'} u_{j'} f_{j'} K_{nij'} \psi_{ij'}\right]$$

$$\le \quad \|\psi\|_\infty^2 \delta_n^2 \mathbb{E}\left[\Delta f_i^j f_j \Delta f_i^{j'} f_{j'} E\left[u_i^2 d_j d_{j'} K_{nij} K_{nij'} \mid \overline{W}\right]\right]$$

$$\le \quad O(\delta_n^2 \lambda_n)\mathbb{E}\left|f_j f_{j'} u_i^2 d_j d_{j'} K_{nij} K_{nij'}\right| = O\left(\delta_n^2 \lambda_n\right),$$

$$(j' = j) \qquad \mathbb{E}\left[u_i \Delta f_i^j u_j^2 f_j^2 K_{nij} \psi_{ij} u_{i'} \Delta f_{i'}^j K_{ni'j} \psi_{i'j}\right]$$
$$\leq \quad \|\psi\|_\infty^2 \, \delta_n^2 \mathbb{E}\left[\Delta f_i^j f_j^2 \Delta f_{i'}^j \mathbb{E}\left[d_i u_j^2 d_{i'} K_{nij} K_{ni'j} \mid \overline{W}\right]\right]$$
$$\leq \quad O(\delta_n^2 \lambda_n) \mathbb{E}\left|f_j^2 d_i u_j^2 d_{i'} K_{nij} K_{ni'j}\right| = O\left(\delta_n^2 \lambda_n\right),$$
$$(i' = j) \qquad \mathbb{E}\left[u_i \Delta f_i^j u_j^2 f_j K_{nij} \psi_{ij} \Delta f_j^{j'} u_{j'} f_{j'} K_{njj'} \psi_{jj'}\right]$$
$$\leq \quad \|\psi\|_\infty^2 \, \delta_n^2 \mathbb{E}\left[\Delta f_i^j f_j \Delta f_j^{j'} f_{j'} E\left[d_i u_j^2 d_{j'} K_{nij} K_{njj'} \mid \overline{W}\right]\right]$$
$$\leq \quad O(\delta_n^2 \lambda_n) \mathbb{E}\left|f_j f_{j'} d_i u_j^2 d_{j'} K_{nij} K_{njj'}\right| = O\left(\delta_n^2 \lambda_n\right).$$

The case $j' = i$ is similar to $i' = j$.

3. Two indices in common to $\{i, j\}$ and $\{i', j'\}$, that is a sum of $2n^{(2)}$ terms. For each term in the sum we can write

$$\mathbb{E}\left[u_i^2 u_j^2 \left(\Delta f_i^j\right)^2 f_j^2 K_{nij}^2 \psi_{ij}^2\right] = O\left(\lambda_n h^{-p}\right)$$

and

$$\mathbb{E}\left[u_i^2 u_j^2 \Delta f_i^j \Delta f_j^i f_i f_j K_{nij}^2 \psi_{ij}^2\right] = O\left(\lambda_n h^{-p}\right).$$

Therefore, $\mathbb{E}[(nh^{p/2} I_{1,1})^2] = \delta_n^4 n^2 h^p O(\lambda_n) + \delta_n^2 n h^p O\left(\lambda_n\right) + O\left(\lambda_n\right) = O(\lambda_n)$. The result then follows from Lemma 6. $\qquad\blacksquare$

Proof of Lemma 3. We only prove the result for $\Delta \hat{r}_i \hat{f}_i$, as the reasoning is similar for $\Delta \hat{f}_i$. We have

$$\Delta \hat{r}_i \hat{f}_i = \frac{1}{(n-1)g^p} \sum_{k \neq i} \left\{Y_k L\left((W_i - W_k)g^{-1}\right) - \mathbb{E}\left[YL\left((W_i - W)g^{-1}\right)\right]\right\}$$
$$+ \mathbb{E}\left[r(W)g^{-p}L\left((W_i - W)g^{-1}\right)\right] - r(W_i)f(W_i)$$
$$= \Delta_{1i} + \Delta_{2i}.$$

The uniform continuity of $r(\cdot)f(\cdot)$ implies $\sup_i |\Delta_{2i}| = o_p(1)$ by Lemma 5. For $\sup_i |\Delta_{1i}|$, we use empirical process tools. Let us introduce some notation. Let $\mathcal{G}$ be a class of functions of the observations with envelope function G and let

$$J(\delta, \mathcal{G}, L^2) = \sup_Q \int_0^\delta \sqrt{1 + \ln N(\varepsilon \|G\|_2, \mathcal{G}, L^2(Q))} d\varepsilon, \qquad 0 < \delta \leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions Q on the space of the observations, and $\|G\|_2$ denotes the norm of G in $L^2(Q)$. Let $Z_1, \ldots, Z_n$ be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma(Z_i), \qquad \gamma \in \mathcal{G}$$

be the empirical process indexed by $\mathcal{G}$. If the covering number $N(\varepsilon, \mathcal{G}, L^2(Q))$ is of polynomial order in $1/\varepsilon$, there exists a constant $c > 0$ such that $J(\delta, \mathcal{G}, L^2) \leq c\delta\sqrt{\ln(1/\delta)}$ for $0 < \delta < 1/2$. Now if $\mathbb{E}\gamma^2 < \delta^2 \mathbb{E}G^2$ for every γ and some

$0 < \delta < 1$, and $\mathbb{E}G^{(4v-2)/(v-1)} < \infty$ for some $v > 1$, under mild additional measurability conditions, Theorem 3.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n \gamma| = J(\delta, \mathcal{G}, L^2) \left(1 + \frac{J(\delta^{1/v}, \mathcal{G}, L^2)}{\delta^2 \sqrt{n}} \frac{\|G\|_{(4v-2)/(v-1)}^{2-1/v}}{\|G\|_2^{2-1/v}} \right)^{v/(2v-1)} \|G\|_2 O_p(1),$$

(8)

where $\|G\|_2^2 = \mathbb{E}G^2$ and the $O_p(1)$ term is independent of n. Note that the family $\mathcal{G}$ could change with n, as soon as the envelope is the same for all n. We apply this result to the family of functions $\mathcal{G} = \{YL((W-w)/g) : w \in \mathbb{R}^p\}$ for a sequence g that converges to zero and the envelope $G(Y, W) = Y \sup_{w \in \mathbb{R}^p} L(w)$. Its entropy number is of polynomial order in $1/\varepsilon$, independently of n, as $L(\cdot)$ is of bounded variation, see for instance van der Vaart and Wellner (1996). Now for any $\gamma \in \mathcal{G}$, $\mathbb{E}\gamma^2(Y, W) \leq Cg^p \mathbb{E}G^2(Y, W)$, for some constant C. Let $\delta = g^{p/2}$, so that $\mathbb{E}\gamma^2(Y, W) \leq C'\delta^2 \mathbb{E}G^2(Y, W)$, for some constant C' and $v = 3/2$, which corresponds to $\mathbb{E}G^8 < \infty$ that is guaranteed by our assumptions. The bound in (8) thus yields

$$\sup_{\mathcal{G}} \left| \frac{1}{g^p \sqrt{n}} \mathbb{G}_n \gamma \right| = \frac{\ln^{1/2}(n)}{\sqrt{ng^p}} \left[1 + n^{-1/2} g^{-2p/3} \ln^{1/2}(n) \right]^{3/4} O_p(1),$$

where the $O_p(1)$ term is independent of n. Since $ng^{4p/3}/\ln n \to \infty$, the expected result follows. $\square$

Proof of Lemma 4. We have

$$\hat{u}_i^* \hat{f}_i = \frac{1}{n-1} \sum_{k \neq i} (Y_i^* - Y_k^*) L_{nik}$$

$$= u_i^* \hat{f}_i - \frac{1}{n-1} \sum_{k \neq i} u_k^* L_{nik} + \frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik}$$

where

$$\frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik} = \frac{1}{n-1} \sum_{k \neq i} (r_i - r_k) L_{nik} + (\hat{r}_i - r_i) \hat{f}_i$$

$$- \frac{1}{(n-1)^2 \hat{f}_k} \sum_{k \neq i} \sum_{k' \neq k} (r_{k'} - r_k) L_{nkk'} L_{nik}$$

$$- \frac{1}{(n-1)^2 \hat{f}_k} \sum_{k \neq i} \sum_{k' \neq k} u_{k'} L_{nkk'} L_{nik}.$$

By Lemma 3 and the fact that $f(\cdot)$ is bounded away from zero, deduce that $\sup_i |\hat{r}_i - r_i| = o_p(1)$. From this and applying several times the arguments in the proof of Lemma 3 we obtain

$$\frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik} = o_p(1).$$

On the other hand,

$$\left| \frac{1}{n-1} \sum_{k \neq i} u_k^* L_{nik} \right| \leq \left| \frac{1}{n-1} \sum_{k \neq i} \eta_k u_k L_{nik} \right| + \frac{\sup_j |\hat{r}_j - r_j|}{n-1} \sum_{k \neq i} |\eta_k| \mathbf{L_{nik}}$$
$$= o_p(1),$$

where we used again the arguments for Δ_{1i} in the proof of Lemma 3 (here with $\eta_k u_k$ and $|\eta_k|$ in the place of Y_k) to derive the last rate. □

References

AIT-SAHALIA, Y., BICKEL, P. J. and STOKER, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics* **105** 363–412.

BOCHNER, S. (1955). *Harmonic Analysis and the Theory of Probability.* University of California Press, Berkeley and Los Angeles. MR0072370 (17,273d)

CHEN, X. and FAN, Y. (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* **91** 373–401.

DELGADO, M. A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.* **17** 199–204. MR1229937 (94g:62098)

DELGADO, M. A. and GONZÁLEZ MANTEIGA, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29** 1469–1507. MR1873339 (2002i:62074)

DETTE, H. and NEUMEYER, N. (2001). Nonparametric analysis of covariance. *Annals of Statistics* **29** 1361–1400.

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. MR1395735 (97a:62103)

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

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics. MR624435 (83a:60001)

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774 (94k:62057)

HURST, S. (1995). The characteristic function of the Student t distribution. Technical Report, Center for Financial Mathematics, Canberra.

JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions.* Wiley, New-York.

LAVERGNE, P. (2001). An equality test across nonparametric regressions. *J. Econometrics* **103** 307–344. Studies in estimation and testing. MR1838202

LAVERGNE, P. and VUONG, Q. (2000). Nonparametric significance testing. *Econometric Theory* **16** 576–601.

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 (94h:62075)

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *Ann. Statist.* **31** 880–920. MR1994734 (2004g:62075)

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *Annals of Statistics* **31** 880–920.

Racine, J. (1997). Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics* **15** 369–378.

Racine, J. S., Hart, J. and Li, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Rev.* **25** 523–544. MR2283340

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* **119** 99–130.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316 (99f:62066)

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics.* Springer-Verlag, New York. With applications to statistics. MR1385671 (97g:60035)

van der Vaart, A. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.* **5** 192–203. MR2792551 (2012k:60067)

Volgushev, S., Birke, M., Dette, H. and Neumeyer, N. (2013). Significance testing in quantile regression. *Electronic Journal of Statistics* **7** 105–145.