The Annals of Statistics
2001, Vol. 29, No. 5, 1469–1507

SIGNIFICANCE TESTING IN NONPARAMETRIC REGRESSION BASED ON THE BOOTSTRAP[1]

BY MIGUEL A. DELGADO AND WENCESLAO GONZÁLEZ MANTEIGA

Universidad Carlos III de Madrid and Universidad de Santiago de Compostela

This paper proposes a test for selecting explanatory variables in nonparametric regression. The test does not need to estimate the conditional expectation function given all the variables, but only those which are significant under the null hypothesis. This feature is computationally convenient and solves, in part, the problem of the "curse of dimensionality" when selecting regressors in a nonparametric context. The proposed test statistic is based on functionals of a U-process. Contiguous alternatives, converging to the null at a rate $n^{-1/2}$ can be detected. The asymptotic null distribution of the statistic depends on certain features of the data generating process, and asymptotic tests are difficult to implement except in rare circumstances. We justify the consistency of two easy to implement bootstrap tests which exhibit good level accuracy for fairly small samples, according to the reported Monte Carlo simulations. These results are also applicable to test other interesting restrictions on nonparametric curves, like partial linearity and conditional independence.

1. Introduction. This paper proposes a testing procedure for choosing significant variables in nonparametric regression. The test only needs a smooth nonparametric estimator of the regression function depending on the explanatory variables which are significant under the null hypothesis. In contrast to other alternative procedures, it is able to detect contiguous alternatives converging to the null at the parametric rate $n^{-1/2}$. The asymptotic null distribution of the test depends on certain features of the data generating process and, therefore, an asymptotic test is difficult to implement except in rare circumstances. In order to estimate the critical values, we propose resampling procedures based on wild bootstrapping of the nonparametric residuals. The method can also be applied to test other restrictions on the nonparametric regression curve, like partial linearity, monotonicity or additivity; and also restrictions on other nonparametric curves. For example, conditional distributions might be tested for conditional independence.

There is a large literature on specification testing, consistent in the direction of general alternatives ("lack-of-fit tests") based on two leading methodologies. On the one hand, tests have been proposed based on some distance between the fitted nonparametric regression, using some smoother, and the parametric fit under the null hypothesis; see, for example, Eubank and Spiegelman

Received October 1998; revised March 2001.

[1]Research funded DGES Grant PB98-0025 and PB98-0182-C02-02.

AMS 2000 *subject classifications*. 62G07, 62G09, 62G10.

Key words and phrases. Nonparametric regression, selection of variables, higher order kernels, U-processes, wild bootstrap, restrictions on nonparametric curves.

(1990), Härdle and Mammen (1993) and the recent monograph by Hart (1997). The test statistics have the form of degenerate U-statistics which, under the null, converge to a standard normal. However, the convergence to the limiting distribution is slow, and Härdle and Mammen (1993) have recommended implementations bootstrap tests in practice. On the other hand, other authors have proposed tests based on a comparison between the empirical integrated regression and the estimated parametric integrated regression function under the specification in the null; see, for example, Brunk (1970), Hong-zhy and Bin (1991), Sue and Wei (1991) and Stute (1997). These tests are based on a marked empirical process and, in general, their null asymptotic distribution depends on certain features of the data generating process. The limiting distribution can be tabulated when the distribution of the regressors is known. Also Stute, Thies and Zhu (1998) and Koul and Stute (1999) suggest a transformation of the underlying empirical process, when the regression depends only on one variable, which is asymptotically distribution free under the null. Transformations when the regression model depends on more than one variable are still unexplored. However, asymptotic critical values can be accurately estimated using bootstrap techniques, as suggested by Stute, González-Manteiga and Presedo-Quindimil (1998). Related to this method are Bierens' tests [see, e.g., Bierens and Ploberger (1997)]. The first testing methodology resembles the goodness-of-fit tests of distribution functions based on the distance between nonparametric and parametric estimates of the probability density curve [see, e.g., Rosenblatt (1975)]. Tests of the second type resemble the typical goodness-of-fit tests of distribution functions based on some distance between the empirical distribution function and the fitted distribution function under the specification on the null. The two methodologies are not comparable from a theoretical viewpoint [see Hart (1997), Chapters 5 and 6, for a discussion]. Tests based on marked empirical processes are able to detect contiguous alternatives converging to the null at a rate $n^{-1/2}$, such as H_{1n} in the next section, which are not detected by tests based on smoothers. However, these last tests detect high frequency alternatives, such as those considered by Rosenblatt [(1975), Section 3], which are not detected by the former tests.

The two methodologies discussed above, which have been developed for specification testing of parametric regression functions, are applicable to testing different restrictions on nonparametric regression curves. Significance testing is a relevant example of restrictions to be tested, since the "curse of dimensionality" may lead one to reduce the number of explanatory variables in the regression curve as much as possible. Let $(S, \mathscr{S}, P)$ be the probability space of the random vector $\chi = (Y, W)$, where Y is scalar and $W = (X, Z)$, $X$is $\mathbb{R}^q$-valued and Z is $\mathbb{R}^p$-valued. We want to test,

$$H_0 : E(Y \mid W) = m(X) \qquad \text{a.s.,}$$

where $m(\cdot) = E(Y|X = \cdot)$. The alternative hypothesis, H_1, is the complement of H_0. Fan and Li (1996) have proposed a significance test inspired by the first methodology discussed above. That is, with the null hypothesis expressed as,

$$H_0 : E\{[E(Y \mid W) - m(X)]^2 \eta(W)\} = 0,$$

where η is a suitable weight function which does not change its sign in the support of W, the test statistic is an estimator of the above expectation, which employes smoothers to estimate the nonparametric expectations and the weight function η, involving the density function of X and W, in order to avoid stochastic denominators. Hence, this testing procedure requires the estimation of two nonparametric regression curves with q and $p+q$ regressors respectively, and the choice of two different bandwidths for each regression, one converging to zero faster than the other. The resulting test statistic has the form of a degenerate U-statistic with a standard normal limiting distribution under the null. However, to the best of our knowledge, bootstrap tests have not been justified in this context.

In this paper, we propose to apply another methodology, which only requires one to estimate the regression function under the null using smoothers, assuming that the distribution of X admits a density, f say. Notice that

$$H_0 : E[Y - m(X) \mid W] = 0 \qquad \text{a.s.},$$

is equivalent to

$$f(X)\, E[Y - m(X) \mid W] = 0 \qquad \text{a.s.},$$

using the fact that $f(X) > 0$ a.s., or

$$T(W) = 0 \qquad \text{a.s.},$$

where, for $w = (x, z)$,

$$T(w) = E\{f(X)[Y - m(X)]\, 1_w(W)\}, \tag{1}$$

where $1_w(W) = 1_x(X) 1_z(Z)$, and $1_v(V) = 1(V \leq v)$, $1(A)$ is the indicator function of the event A, and for two vectors v and w of equal dimension, "$v \leq w$" means that each coordinate of v is less than or equal to the corresponding coordinate of w. Hence, (1) is the difference between the weighted integrated regression function of Y given W and of Y given X. The reason of writing H_0 in this form is mainly technical, in order to avoid the random denominator in the conditional expectation. Test statistics are suitable functionals of a T estimate.

In next section, we study the asymptotic properties of test statistics. Asymptotic tests are difficult to implement, since the asymptotic distribution of the statistic under the null depends on unknown features of the underlying distribution of χ. In Section 3, we propose consistent bootstrap tests, which are easy to implement. A Monte Carlo study, in Section 4, illustrates the properties of the proposed bootstrap tests in practice. In Section 5, we propose the extension of this testing methodology to other restrictions on nonparametric curves, discussing in detail a test for partial linearity and a test for conditional independence. Proofs of the main results are deferred to Section 6. They are based on some lemmas, which are listed in Section 7.

2. Significance testing. Let $\mathscr{Y}_n = \{\chi_i, i = 1, \ldots, n\}$, $\chi_i = (Y_i, W_i)$, $W_i = (X_i, Z_i)$, be independent copies of $\chi = (Y, W)$, which has probability space $(S, \mathscr{S}, P)$. The test is based on the $T(w)$ estimate

$$\begin{aligned} T_n(w) &= \frac{1}{n} \sum_i \hat{f}(X_i)(Y_i - \hat{m}(X_i)) 1_w(W_i) \\ &= \frac{1}{n^2} \sum_i \sum_j \frac{1}{a^q} K_{ij}(Y_i - Y_j) 1_w(W_i) \end{aligned} \tag{2}$$

where

$$\hat{m}(X_i) = \frac{1}{\hat{f}(X_i)} \frac{1}{na^q} \sum_j K_{ij} Y_j \quad \text{and} \quad \hat{f}(X_i) = \frac{1}{na^q} \sum_j K_{ij},$$

where $K_{ij} = K\left((X_i - X_j)/a\right)$, $K(u) = \prod_{j=1}^q k(u_j)$, k is an univariate kernel and $a = a(n) \in \mathbb{R}^+$ is a bandwidth. The test statistic is a functional of the random element $n^{1/2} T_n$, for instance, the Cramér-von Mises' statistic of the form

$$C_n = \int \left[n^{1/2} T_n(w)\right]^2 dF_{Wn}(w) = \sum_i T_n(W_i)^2,$$

where, henceforth, F_ζ is the distribution function of the real valued random variable ζ, and $F_{\zeta n}$ its corresponding empirical distribution function; or the Kolmogorov-Smirnov statistic of the form

$$K_n = \sup_w \left|n^{1/2} T_n(w)\right|.$$

T_n is a U-process of the type considered by Stute (1994). We can write

$$T_n(w) = \frac{(n-1)}{n} U_n(w, \infty; \infty, \infty),$$

where

$$U_n(s_1; s_2) = \frac{1}{n(n-1)} \sum_{i \neq j} \psi_a(\chi_i, \chi_j) 1_{s_1}(\chi_i) 1_{s_2}(\chi_j),$$

with $s_j = (w_j, y_j)$, $j = 1, 2$, and,

$$\psi_a(\chi_i, \chi_j) = (Y_i - Y_j) \frac{1}{a^q} K_{ij}.$$

Write

$$U_n(s_1; s_2) = \hat{U}_n(s_1; s_2) + R_n(s_1; s_2),$$

where,

$$\begin{aligned} &\hat{U}_n(s_1; s_2) \\ &\quad = \left\{ \frac{1}{n} \sum_i 1_{s_1}(\chi_i) \int 1_{s_2}(\bar{s}) \psi_a(\chi_i, \bar{s}) P(d\bar{s}) - E\left[\psi_a(\chi_1, \chi_2) 1_{s_1}(\chi_1) 1_{s_2}(\chi_2)\right] \right\} \end{aligned}$$

$$+\frac{1}{n}\sum_i 1_{s_2}(\chi_i)\int 1_{s_1}(\bar{s})\psi_a(\bar{s},\chi_i)P(d\bar{s})$$
$$=\hat{U}_n^{(1)}(s_1;s_2)+\hat{U}_n^{(2)}(s_1;s_2)$$

is the Hájek projection of U_n and the remainder $R_n = U_n - \hat{U}_n$ is a degenerate U-process. Assuming that ψ_a has second moments for each $a \in \mathbb{R}^+ \setminus \{0\}$, a generalization of Stute [(1994), Theorem 1.1] to the multivariate case (see Proposition 4 in Section 6) shows that, for each $a \in \mathbb{R}^+ \setminus \{0\}$,

$$E\left[\sup_{s_1,s_2}\left|n^{1/2}R_n(s_1;s_2)\right|^2\right] \le C\frac{1}{n}E\left[\left|\psi_a(\chi_1,\chi_2)\right|^2\right],$$

where C is a constant independent of ψ_a. The next Proposition shows that $n^{1/2}R_n$ vanishes in probability uniformly in (s_1, s_2) under fairly weak conditions.

PROPOSITION 1. *Let* $\sup_x f(x) < \infty$, $\sup_u |k(u)| + \int |k(u)|du < \infty$, $E(Y^2) < \infty$ *and* $(na^q)^{-1} \to 0$ *as* $n \to \infty$. *Then,*

$$\sup_{s_1,s_2}\left|R_n(s_1;s_2)\right| = o_p\left(n^{-1/2}\right).$$

Thus, uniformly in w,

$$T_n(w) = \frac{n-1}{n}\left[\bar{U}_n^{(1)}(w) + \bar{U}_n^{(2)}(w)\right] + o_p\left(n^{-1/2}\right),$$

where, henceforth, $\bar{U}_n^{(j)}(w) := \hat{U}_n^{(j)}(w,\infty;\infty,\infty)$, $j = 1,2$. The empirical processes $\bar{U}_n^{(1)}$ and $\bar{U}_n^{(2)}$ are of different natures. On one hand,

$$\bar{U}_n^{(1)}(w) = \left[\frac{1}{n}\sum_i 1_w(W_i)\int (Y_i - m(\bar{x}))f(\bar{x})\frac{1}{a^q}K\left(\frac{X_i - \bar{x}}{a}\right)d\bar{x}\right]$$
$$-E\left[1_w(W_1)(Y_1 - m(X_2))\frac{1}{a^q}K_{12}\right]$$

is a centered classical marked empirical process, with marks depending on a, which is expected, under H_0, to be asymptotically equivalent to,

$$\tilde{U}_n^{(1)}(w) = \frac{1}{n}\sum_i 1_w(W_i)(Y_i - m(X_i))f(X_i),$$

under suitable smoothness assumptions on f and m, which must be related to the kernel order and bandwidth rate of convergence. $\tilde{U}_n^{(1)}$ is very similar to the empirical process considered by Stute (1997) for testing the simple hypothesis $E\left(Y\middle| X = \cdot\right) = m(\cdot)$, with m known. He considers the process indexed by a single parameter, showing that, conveniently normalized, it converges in distribution in the càdlàg functional space $D(-\infty, \infty)$ endowed with the Skorohod's norm to a Gaussian element. On the other hand, under H_0,

$$\bar{U}_n^{(2)}(w) = \frac{1}{n}\sum_i \int \gamma_w(Y_i, \bar{x})\frac{1}{a^q}K\left(\frac{X_i - \bar{x}}{a}\right)d\bar{x},$$

where

$$\gamma_w(\bar{y}, \bar{x}) = (m(\bar{x}) - \bar{y})\,1_x(\bar{x})\,r_z(\bar{x})\,f(\bar{x})$$

and $r_z(\bar{x}) = \Pr\left(Z \le z \middle| X = \bar{x}\right)$, is a smoothed version of the empirical process,

$$\tilde{U}_n^{(2)}(w) = \frac{1}{n}\sum_i \gamma_w(Y_i, X_i).$$

It is also expected that, under H_0, $n^{1/2}\bar{U}_n^{(2)}$ and $n^{1/2}\tilde{U}_n^{(2)}$ are asymptotically equivalent, assuming similar conditions than those needed to show the asymptotic equivalence between $n^{1/2}\bar{U}_n^{(1)}$ and $n^{1/2}\tilde{U}_n^{(1)}$. The following definitions, introduced by Robinson (1988), are helpful in the presentation of these regularity conditions. The first one defines a class of higher order kernels and the second one defines a class of smooth functions.

DEFINITION 1. $\mathscr{K}_\ell$, $\ell \ge 1$ is the class of even functions of uniformly bounded variation, $k: \mathbb{R} \to \mathbb{R}$, satisfying

$$k(u) = O\left(\left(1 + |u|^{\ell+1+\varepsilon}\right)^{-1}\right) \qquad \text{some } \varepsilon > 0,$$

$$\int_{\mathbb{R}} u^i k(u)\,du = \delta_{i0}, \qquad i = 0, \dots, \ell - 1,$$

where δ_{ij} is Kroneker's delta.

It is interesting that whereas the classes $\mathscr{K}_\ell$ play useful roles in bias-reduction and widening the spectrum of admissible bandwidths in nonparametric estimation, they are crucial in our problem, which requires dealing with a greater ($n^{1/2}$) norming than in the central limit theorem for *q−variate* nonparametric estimators ($(na^q)^{1/2}$).

DEFINITION 2. $\mathscr{G}_\beta^\alpha$, $\alpha > 0$, $\beta > 0$, is the class of functions $g: \mathbb{R}^q \to \mathbb{R}$ satisfying: g is uniformly $(b-1)$-times continuously differentiable, for $b - 1 \le \beta \le b$; for some $\rho > 0$, there exists a function d such that

$$\sup_{v \in \mathscr{S}_{u\rho}} \left|g(v) - g(u) - Q(v, u)\right| / \|v - u\|^\beta \le d(u),$$

for all u, where $\mathscr{S}_{u\rho} = \{v : \|v - u\| < \rho\}$; $Q = 0$ when $b = 1$; Q is a $(b-1)$th degree homogeneous polynomial in $v - u$ with coefficients being the partial derivatives of g at u of orders 1 through $b - 1$ when $b > 1$; and $g(u)$, its partial derivatives of orders $b-1$ and less, and $d(u)$, have finite αth moments.

The functions $\mathscr{J}_\beta^\alpha$ are thus expanded in a Taylor series with a local Lipschitz condition on the remainder; (α, β) depending simultaneously on smoothness and moment properties. Bounded functions in $Lip(\beta)$ (the Lipschitz class of degree β) for $0 < \beta \leq 1$ are in $\mathscr{J}_\beta^\infty$; for $\beta > 1$, $\mathscr{J}_\beta^\infty$ contains the bounded and $(b-1)$−times boundedly differentiable functions whose $(b-1)$th partial derivatives are in $Lip(\beta - b + 1)$. In applying $\mathscr{J}_\beta^\alpha$ to f, we take $\alpha = \infty$, but we allow for $\alpha < \infty$ in Definition 2, because we have no wish that m is a.s. bounded. The next set of regularity conditions are usually assumed for showing the $\sqrt{n}$*−consistency* of semiparametric estimators.

A1. $f \in \mathscr{J}_\lambda^\infty$, for some $\lambda > 0$.
A2. $m \in \mathscr{J}_\tau^2$, for some $\tau > 0$.
A3. $k \in \mathscr{K}_{\ell+t-1}$, where $\ell - 1 < \lambda \leq \ell$ and $t - 1 < \tau \leq t$.
A4. $(na^q)^{-1} + na^{2\min(\tau,\lambda+1)} \to 0$ as $n \to \infty$.
A5. $E[|Y - m(X)|^{2+\delta}] < \infty$ for some $\delta > 0$.

Assumptions A3 and A4 have to be satisfied simultaneously for λ and τ satisfying the stated inequalities, so that, for example, when $k \in \mathscr{K}_2$ only, the lower bounds on a's rate of decay are not better than $na^4 \to 0$, no matter the degree of smoothness of m and f. A necessary condition for A4 is $\tau > q/2$ and $\lambda > q/2 - 1$. Thus, a necessary condition for A3 is $k \in \mathscr{K}_{q-1}$. Fan and Li (1996) assume four moments for the errors, while we only need more than two. Additionally, they require smoothness conditions on the density of W and conditional moments of the regression errors, as well as, four moments for m.

PROPOSITION 2. *Under* H_0, *if* A1 *to* A5 *hold*,

$$\sup_w \left| \bar{U}_n^{(1)}(w) - \tilde{U}_n^{(1)}(w) \right| = o_p\left(n^{-1/2}\right).$$

In order to prove a similar result for $\bar{U}_n^{(2)} - \tilde{U}_n^{(2)}$, we need some smoothness assumption on the family of functions $\mathscr{R} = \{r_z : z \in \mathbb{R}^p\}$.

A6. $\mathscr{R} \subset \mathscr{J}_v^\infty$, some $v > 0$.

PROPOSITION 3. *Under* H_0, *if* A1 *to* A6 *hold*,

$$\sup_w \left| \bar{U}_n^{(2)}(w) - \tilde{U}_n^{(2)}(w) \right| = o_p\left(n^{-1/2}\right).$$

Write

$$\tilde{U}_n(w) := \tilde{U}_n^{(1)}(w) + \tilde{U}_n^{(2)}(w) = \frac{1}{n} \sum_i \xi_w(\chi_i),$$

where

$$\xi_w(\chi_i) = (Y_i - m(X_i)) f(X_i) 1_x(X_i) [1_z(X_i) - r_z(X_i)].$$

We interpret $\tilde{U}_n$ as a process indexed by a class of functions, rather than indexed by two sets of parameters, x and z. Consider the family of functions $\mathscr{H} := \{\xi_w : w \in \mathbb{R}^{p+q}\}$. Under A1 and A5, $\mathscr{H} \subset \mathscr{L}_2(S, \mathscr{S}, P)$ such that,

$$H(s) := \sup_{h\in\mathscr{H}} \left|h(s)\right| < \infty \qquad \text{for all } s \in S.$$

$$\sup_{h\in\mathscr{H}} \left|Ph\right| < \infty,$$

where, henceforth, we use the abbreviation $Qg = \int g dQ$ for any generic measure Q. Thus, the maps

$$\delta_s : \mathscr{H} \mapsto \mathbb{R} \text{ given by } \delta_s(h) = h(s),$$

$$P : \mathscr{H} \mapsto \mathbb{R} \text{ given by } Ph = \int h(s)\, dP(s),$$

are bounded. So, δ_s and P belong to $\ell^\infty(\mathscr{H})$, the Banach space of real bounded functions on $\mathscr{H}$, equipped with the supremum norm, $\|t\|_{\mathscr{H}} = \sup_{h\in\mathscr{H}} \left|t(h)\right|$. The empirical measure, defined as,

$$P_n = \frac{1}{n}\sum_i \delta_{\chi_i},$$

induces a map from $\mathscr{H}$ to $\mathbb{R}$ given by, $h \mapsto P_n h$. That is, the empirical process indexed by the class of functions $\mathscr{H}$, $\{\tilde{U}_n(w),\ w \in \mathbb{R}^{p+q}\} := \{P_n h : h \in \mathscr{H}\}$ can be viewed as a random element with values in $\ell^\infty(\mathscr{H})$. Let $\{G_P h : h \in \mathscr{H}\}$, the P-Brownian bridge indexed by $\mathscr{H}$, be the centered Gaussian process with covariance function,

$$E[G_P g G_P h] = Pgh - PgPh,\ \ g, h \in \mathscr{H}.$$

For any finite subset $J \subset \mathscr{H}$,

$$\{P_n g,\ g \in J\} \text{ converges in distribution to } \{G_P g :\ g \in J\}$$

by the multivariate central limit theorem. This convergence is made "uniform" over all $\mathscr{H}$ using the Hoffmann-Jørgensen (1984) definition: If $\{S_n\}_{n=0}^\infty$ are $\ell^\infty(\mathscr{H})$ random valued elements, and S_∞ is measurable and has separable support, then,

$$S_n \text{ converges in distribution to } S_\infty \text{ in } \ell^\infty(\mathscr{H}),$$

if and only if

$$E^*(H(S_n)) \to E(H(S_\infty))$$

for all $H : \ell^\infty(\mathscr{H}) \to \mathbb{R}$ bounded and continuous. E^* stands for outer expectation. We say that the process G_P is sample continuous, or a P-tight Borel measurable element of $\ell^\infty(\mathscr{H})$, when it has a version with bounded ρ_P-continuous trajectories, where

$$\rho_P^2(g, h) = E\left[(G_P g - G_P h)^2\right], \qquad g, h \in \mathscr{H},$$

which means that G_P is measurable and has support $\mathscr{C}_u(\mathscr{H}, \rho_P)$ ($\mathscr{C}_u$ =bounded uniformly continuous functions), which is separable in $(\ell^\infty(\mathscr{H}), \|\cdot\|_{\mathscr{H}})$. Henceforth, we use Dudley's definition [Dudley (1999), page 94] of "CLT for the empirical process uniform over $\mathscr{H}$": $\mathscr{H}$ is a P-Donsker class if and only if:

(i) G_P is sample continuous, and
(ii) $n^{1/2}(P_n - P)$ converges in distribution to G_P in $\ell^\infty(\mathscr{H})$.

In Section 6 we show that $\mathscr{H}$ is a P-Donsker class. Under H_0, $\sup_{h\in\mathscr{H}}|Ph| = 0$, and $G_P \overset{d}{=} B_P$ is a *P-Brownian motion* in $\ell^\infty(\mathscr{H})$, a Gaussian process with zero mean and $E[B_P g B_P h] = Pgh$ for $g, h \in \ell^\infty(\mathscr{H})$. The following Theorem is a direct consequence of Propositions 1, 2, 3 and the fact that $\mathscr{H}$ is P-Donsker.

THEOREM 1. *Under H_0, if* A1 *to* A6 *hold,*

$$n^{1/2}T_n \textit{ converges in distribution to } B_P \textit{ in } \ell^\infty(\mathscr{H}),$$

where B_P is sample continuous.

We can obtain the asymptotic distribution of any continuous functional $\varphi : \ell^\infty(\mathscr{H}) \mapsto \mathbb{R}$, under H_0, applying the Continuous Mapping Theorem [e.g., Dudley (1999), Theorem 3.6.7, page 116]. That is, if $\mathscr{H}$ is P-Donsker and the functional $\varphi : \ell^\infty(\mathscr{H}) \mapsto \mathbb{R}$ is continuous, under H_0, for any continuous and bounded function $H : \mathbb{R} \mapsto \mathbb{R}$,

$$E\left(H\left(\varphi\left(n^{1/2}T_n\right)\right)\right) \to E\left(H\left(\varphi\left(B_P\right)\right)\right),$$

since $H \circ \varphi$ is a continuous and bounded functional, and, therefore,

$$\varphi\left(n^{1/2}T_n\right) \text{ converges in distribution to } \varphi\left(B_P\right),$$

meaning standard convergence in distribution on the real line. Next Corollary establishes the asymptotic distribution of K_n and C_n under the null.

COROLLARY 1. *Under H_0, if* A1 *to* A6 *hold,*

$$K_n \textit{ converges in distribution to } \|B_P h\|_{\mathscr{H}}$$

and

$$C_n \textit{ converges in distribution to } \int \left|B_P \xi_w\right|^2 F_W(dw).$$

From Propositions 2 and 3, it is immediate that, uniformly in w, under H_1,

$$\bar{U}_n^{(1)}(w) = \tilde{U}_n^{(1)}(w) - E\left[1_w(W)(Y - m(X)) f(X)\right] + o_p\left(n^{-1/2}\right),$$

$$\bar{U}_n^{(2)}(w) = \tilde{U}_n^{(2)}(w) + \frac{1}{n}\sum_i \int b_w(\bar{x}) f(\bar{x}) \frac{1}{a^q} K\left(\frac{X_i - \bar{x}}{a}\right) d\bar{x} + o_p\left(n^{-1/2}\right),$$

where $b_w(\bar{x}) = E\left[1_w(W)(Y - m(X)) \middle| X = \bar{x}\right]$. Thus, applying Proposition 1, uniformly in w,

$$T_n(w) = \tilde{U}_n(w) + \left\{ \frac{1}{n} \sum_i \int b_w(\bar{x}) f(\bar{x}) \frac{1}{a^q} K\left(\frac{X_i - \bar{x}}{a}\right) d\bar{x} - E\left[b_w(X) f(X)\right] \right\}$$
$$+ o_p\left(n^{-1/2}\right).$$

Since $\{b_w \cdot f, w \in \mathbb{R}^{p+q}\}$ is P-Donsker (see the proof of Proposition 3 in Section 6), and applying Proposition 5 in Section 6, the second term on the right hand side of the last expression is $O_p\left(n^{-1/2}\right)$, uniformly in w, under suitable smoothness assumptions on b_w (e.g., $b_w \in \mathscr{J}_v^\infty$ for some $v > 0$ and all w.) Thus, using the fact that $\mathscr{H}$ is P-Donsker, uniformly in w,

$$T_n(w) = E\left[\xi_w(\chi)\right] + O_p\left(n^{-1/2}\right),$$

which guarantees the consistency of tests based on continuous functionals of $n^{1/2} T_n$, in particular those based on C_n and K_n. That is, given asymptotic critical values,

$$c_\alpha^K = \inf\left\{t : F_{\|B_P\|_{\mathscr{H}}}(t) \geq 1 - \alpha\right\}$$

and

$$c_\alpha^C = \inf\left\{t : F_{\int |B_P \xi_w|^2 F_W(dw)}(t) \geq 1 - \alpha\right\}$$

under H_1, $F_{K_n}\left(c_\alpha^K\right) = o(1)$ and $F_{C_n}\left(c_\alpha^C\right) = o(1)$. Under contiguous alternatives of the form

$$H_{1n} : E(Y \mid W) = m(X) + \frac{\Psi(W)}{n^{1/2}} \qquad \text{a.s.}$$

applying a similar argument,

$$n^{1/2} T_n(w) = n^{1/2} \tilde{U}_n(w)$$
$$+ \left\{ \frac{1}{n} \sum_i \int e_w(\bar{x}) f(\bar{x}) \frac{1}{a^q} K\left(\frac{X_i - \bar{x}}{a}\right) d\bar{x} - E\left[e_w(X) f(X)\right] \right\} + o_p(1).$$

where $e_w(\bar{x}) = E\left[1_w(W)\Psi(W) \middle| X = \bar{x}\right]$. Under suitable smoothness assumptions on e_w, uniformly in w,

$$n^{1/2} T_n(w) = n^{1/2} \tilde{U}_n(w) + o_p(1).$$

Noticing that

$$n^{1/2} \tilde{U}_n(w) = n^{1/2} P_n \xi_w^0 + P_n \xi_w^1,$$

where $\xi_w^0(\chi) = \left(Y - E\left(Y \middle| W\right)\right) f(X) 1_x(X)\left[1_z(Z) - r_z(X)\right]$ and $\xi_w^1(\chi) = \Psi(W) f(X) 1_x(X)\left[1_z(Z) - r_z(X)\right]$, and $\mathscr{H}^0 = \left\{\xi_w^0 : w \in \mathbb{R}^{p+q}\right\}$ is, like $\mathscr{H}$,

P-Donsker, it follows that

$$K_n \text{ converges in distribution to } \sup_w \left| B_P \xi_w^0 + P\xi_w^1 \right|,$$

$$C_n \text{ converges in distribution to } \int \left| B_P \xi_w^0 + P\xi_w^1 \right|^2 dF_W(w).$$

Therefore, under H_{1n}, $F_{K_n}\left(c_\alpha^K\right) \leq 1 - \alpha + o(1)$ and $F_{K_n}\left(c_\alpha^C\right) \leq 1 - \alpha + o(1)$.

These test statistics are not asymptotically pivotal, and asymptotic tests are difficult to implement in practice. This is why we propose bootstrap tests in the next section.

3. Bootstrap tests. A bootstrap version of $\tilde{U}_n$ is

$$\tilde{U}_n^*(w) = \frac{1}{n} \sum_i V_i (Y_i - \hat{m}(X_i)) \hat{f}(X_i) \left[1_w(W_i) - \hat{\phi}_w(X_i)\right],$$

where

$$\hat{\phi}_w(X_i) = \frac{1}{na^q \hat{f}(X_i)} \sum_j 1_w(W_j) K_{ij}$$

is an estimate of

$$\phi_w(X_i) = 1_x(X_i) r_z(X_i),$$

and $\{V_i, i = 1, \ldots, n\}$ are random variables such that:

A8. $\{V_i, i = 1, \ldots, n\}$ are bounded, iid independent of $\mathscr{Y}_n = \{(Y_i, W_i), i = 1, \ldots, n\}$, such that $E(V_1) = 0$ and $E(V_1^2) = 1$.

From a computational view-point, it is worth noticing that $\tilde{U}_n^*$ can also be written as

$$\tilde{U}_n^*(w) = \frac{1}{n} \sum_i (\hat{\varepsilon}_i^* - \bar{\varepsilon}_i^*) \hat{f}(X_i) 1_w(W_i),$$

where $\{\hat{\varepsilon}_i^* = V_i \hat{\varepsilon}_i, i = 1, \ldots, n\}$ is the bootstrap resample of the nonparametric residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, and $\bar{\varepsilon}_i^* = (na^q \hat{f}(X_i))^{-1} \sum_j \hat{\varepsilon}_j^* K_{ij}$. Thus, if we have a program for computing T_n, with input $\{(Y_i, W_i), i = 1, \ldots, n\}$, we can use the same program with input $\{(\hat{\varepsilon}_i^*, W_i), i = 1, \ldots, n\}$ to compute $\tilde{U}_n^*$. The test statistic is the bootstrap version of $\varphi(n^{1/2}\tilde{U}_n)$, $\varphi(n^{1/2}\tilde{U}_n^*)$, for some given continuous functional φ. Notice that

$$T_n(w) = \frac{1}{n} \sum_i (Y_i - m(X_i)) \hat{f}(X_i) \left[1_w(W_i) - \hat{\phi}_w(X_i)\right]$$
$$+ \frac{1}{n} \sum_i m(X_i) \hat{f}(X_i) \left[1_w(W_i) - \hat{\phi}_w(X_i)\right].$$

Hence, $\tilde{U}_n^*$ can be interpreted as a bootstrap version of the first term on the right hand side of this last expression, neglecting the effect of the second term, which can be interpreted as a bias term. The bootstrap analog of T_n is

$$T_n^*(w) = \tilde{U}_n^* + \frac{1}{n}\sum_i \hat{m}(X_i)\hat{f}(X_i)\left[1_w(W_i) - \hat{\phi}_w(X_i)\right]$$

$$= \frac{1}{n}\sum_i (Y_i^* - \hat{m}^*(X_i))\hat{f}(X_i)1_w(W_i),$$

where $Y_i^* = \hat{m}(X_i) + \hat{\varepsilon}_i^*$ and $\hat{m}^*(X_i) = (na^q\hat{f}(X_i))^{-1}\sum_i Y_i^* K_{ij}$, which takes into account this bias term. That is, T_n^* is the bootstrap version of T_n computed with the "wild resample" $\{(Y_i^*, X_i), i = 1, \ldots, n\}$. This resampling procedure was introduced by Wu (1986) in the context of estimation in heteroskedastic linear models. In specification testing of a parametric regression functions, the "wild bootstrap" has been applied by Härdle and Mammen (1993) in tests statistics based on smoothers, and by Stute, González-Manteiga and Presedo-Quindimil (1998) in test statistics based on estimates of the integrated regression function.

Bootstrap versions of K_n are

$$K_n^* = \sup_w \left|n^{1/2}\tilde{U}_n^*(w)\right| \quad \text{and} \quad K_n^{**} = \sup_w \left|n^{1/2}T_n^*(w)\right|,$$

and bootstrap versions of C_n are

$$C_n^* = \sum_i \tilde{U}_n^*(w)^2 \quad \text{and} \quad C_n^{**} = \sum_i T_n^*(w)^2.$$

Because $\tilde{U}_n^*$ and T_n^* have a random denominator, we need the following assumption.

A9. $\Pr(f(X) > \vartheta) = 1$ for some $\vartheta > 0$.

Assumption A9 rules out important distributions, like the Beta and Normal. However, from a practical view point, this assumption does not seem so damaging, since truncated distributions can always be considered. Another way of dealing with the random denominator problem, avoiding assumption A9, consists of introducing some trimming as suggested by Robinson (1988). It will imply the choice of a trimming parameter, whose rate of convergence must be related to the bandwidth a. We also need stronger conditions on the rate of convergence of a.

A4′. $\left(na^{2q}\right)^{-1} + na^{2\min(\tau,\lambda+1,2\lambda)} \to 0$ as $n \to \infty$.

Consider the infeasible version of $\tilde{U}_n^*(w)$,

$$\tilde{U}_n^{0*}(w) = \frac{1}{n}\sum_i V_i \xi_w(\chi_i).$$

The next two theorems establish that $n^{1/2}\tilde{U}_n^*$ and $n^{1/2}T_n^*$ are asymptotically equivalent to $n^{1/2}\tilde{U}_n^{0*}$.

THEOREM 2. *Under* A1–A4′, A5–A9,

$$\sup_w \left|\tilde{U}_n^*(w) - \tilde{U}_n^{0*}(w)\right| = o_p\left(n^{-1/2}\right).$$

THEOREM 3. *Under* A1–A4′, A5–A9,

$$\sup_w \left|T_n^*(w) - \tilde{U}_n^{0*}(w)\right| = o_p\left(n^{-1/2}\right).$$

The bootstrap empirical measure,

$$P_n^* = \frac{1}{n}\sum_i V_i \delta_{\chi_i},$$

also induces a map from $\mathscr{H}$ to $\mathbb{R}$, and $\{\tilde{U}_n^{0*}(w) : w \in \mathbb{R}^{p+q}\} := \{P_n^* h,\ h \in \mathscr{H}\}$ can be viewed as a random element of $\ell^\infty(\mathscr{H})$. Since $\mathscr{H}$ is P-Donsker with square integrable envelope, the almost sure conditional multiplier central limit theorem [Ledoux and Talagrand (1988)] establishes that

$$n^{1/2}\tilde{U}_n^{0*} \text{ converges in distribution to } B_P \text{ in } \ell^\infty(\mathscr{H}) \qquad \text{a.s,}$$

where B_P is sample continuous [an excelent exposition of convergence of bootstrap distributions is in Giné (1997)]. Thus, for a continuous functional $\varphi : \ell^\infty(\mathscr{H}) \mapsto \mathbb{R}$,

$$d\left(F^*_{\varphi\left(n^{1/2}\tilde{U}_n^{0*}\right)}, F_{\varphi(B_P)}\right) = o(1) \qquad \text{a.s.,}$$

where d is a distance metrizing weak convergence on the real line, and F^* denotes the conditional distribution given the sample $\mathscr{Y}_n$. Therefore, applying Theorems 2 and 3,

$$d\left(F^*_{\varphi\left(n^{1/2}\tilde{U}_n^{*}\right)}, F_{\varphi(B_P)}\right) = o_p(1) \quad \text{and} \quad d\left(F^*_{\varphi(n^{1/2}T_n^*)}, F_{\varphi(B_P)}\right) = o_p(1),$$

and we say that,

$$\varphi\left(n^{1/2}\tilde{U}_n^*\right) \text{ and } \varphi\left(n^{1/2}T_n^*\right) \text{ converge in distribution to } \varphi(B_P) \text{ in probability.}$$

The next Corollary justifies the implementation of these tests in practice, using as critical values the quantiles of the conditional distribution of the bootstrap statistics given the sample $\mathscr{Y}_n$.

COROLLARY 2. *Under* A1–A4′, A5–A9,

K_n^* *and* K_n^{**} *converge in distribution to* $\|B_P\|_{\mathscr{H}}$ *in probability*,

C_n^* *and* C_n^{**} *converge in distribution to* $\int \left|B_P \xi_w\right|^2 dF_w(w)$ *in probability.*

Let η_n be the statistic used for testing H_0 (e.g., C_n or K_n) and η_n^* the corresponding bootstrap statistic (e.g., C_n^*, C_n^{**}, K_n^* or K_n^{**}). Define the bootstrap critical value $c_{n\alpha} = \inf\{t : F^*_{\eta_n^*}(t) \geq 1-\alpha\}$, and let η_∞ be the asymptotic distribution of η_n under H_0. Corollary 1 and 2 implies that, if F_{η_∞} is continuous at $c_\alpha = \inf\{t : F_{\eta_\infty}(t) \geq 1-\alpha\}$, under H_0, $\Pr(\eta_n \geq c_{n\alpha}|\mathscr{Y}_n) = \alpha + o_p(1)$ and also $\Pr(\eta_n \geq c_{n\alpha}) = \alpha + o(1)$. Corollary 2 holds also under H_1 and H_{1n}, which guarantees consistency, and power in the direction of contiguous alternatives H_{1n}, of the test based on critical values $c_{n\alpha}$. In practice, bootstrap critical values $c_{n\alpha}$ can be approximated, as accurately as desired, by Monte Carlo. Beran, Le Cam and Millar (1987) justify this statement, showing that, whenever a bootstrapped limit theorem holds in probability, then, the empirical distributions of the bootstrapped laws also converge in probability. The implementation of the tests is as follows. We generate B bootstrap residual samples, $\{(\hat{\varepsilon}_i^{*b}, i = 1, \dots, n), b = 1, \dots, B\}$, according to our resampling procedure, and the corresponding bootstrap statistics $\{\eta_b^*, b = 1, \dots, B\}$ are computed. Then, $c_{n\alpha}$ is approximated by $c^B_{n\alpha} = \inf\{t : B^{-1}\sum_{b=1}^B 1(\eta_b^* > t) \geq 1-\alpha\}$; the larger B is the better is the approximation of $c_{n\alpha}$.

4. Monte Carlo. In this Monte Carlo study we provide evidence on the sensitivity of the test to the bandwidth choice and the dimension of Z and X in small samples. We choose $a = Cn^{-1/3q}$ for $C = 0.25$, 0.5, 1 and 2, which is compatible with A4 and A4$'$. The bootstrap tests are compared with the parametric asymptotic Wald's test of significance of regressors Z in a linear regression model. We consider the case $q = 1, 2$ and $p = 1, 2$ under different designs. We choose the Epanechnikov's kernel, $k(u) = 1(|u| \leq 1)(1-u^2)3/4$, of different orders, depending on q, as suggested by our sufficient conditions. The V_i variables are the same as in Härdle and Mammen (1993) and Stute et al. (1998). That is, we consider a two point distribution attaching masses $(\sqrt{5}+1)/2\sqrt{5}$ and $(\sqrt{5}-1)/2\sqrt{5}$ to the points $-(\sqrt{5}-1)/2$ and $(\sqrt{5}+1)/2$, respectively. The tables report the proportion of rejections in 2000 Monte Carlo samples using 2000 bootstrap samples for approximating the critical values by Monte Carlo. We only present simulation results for C_n. Simulation results for K_n, which are not reported here, are very similar. Samples are generated according to the model

$$Y_i = m(X_i) + \beta \sin\left(\gamma Z_i^{(1)}\right) + U_i,\ i = 1, \dots, n,$$

where $U_i \sim N(0,1)$ and the regressors are iid $U(0,1)$, independent of U_i. As γ increases, in the sine model, the regression curve has more oscillations and the correlation between Y_i and $Z_i^{(1)}$ decreases.

Table 1 examines the behavior of the test under the null hypothesis ($\beta = 0$) when $q = 1$ and $p = 1, 2$. We consider a linear model $m(x) = 1 + x$ and a sinusoidal model $m(x) = 1 + \sin(10x)$. As could be expected, the empirical size of Wald's test is very close to the theoretical one in all cases. The bootstrap tests exhibit good level accuracy in the linear model for all the bandwidth choices. However, for the sine model, which is more difficult to estimate due to the

TABLE 1

Proportion of rejections in 2000 Monte Carlo samples, under H_0: $E(Y \mid W) = E(Y \mid X)$ a.s., $p = 1, 2$ for the bootstrap test and an asymptotic t-ratio based on a linear regression model

$m(x) = 1 + x, \quad p = 1$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.098				0.103			
	0.05	0.054				0.052			
	0.01	0.016				0.013			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.147	0.124	0.100	0.100	0.128	0.120	0.109	0.114
	0.05	0.080	0.062	0.049	0.059	0.073	0.061	0.054	0.056
	0.01	0.011	0.009	0.007	0.007	0.016	0.011	0.011	0.010
C_n^{*}	0.1	0.153	0.128	0.105	0.129	0.132	0.123	0.114	0.143
	0.05	0.084	0.061	0.055	0.059	0.074	0.063	0.056	0.077
	0.01	0.013	0.008	0.010	0.013	0.017	0.011	0.011	0.019
$m(x) = 1 + x, \quad p = 2$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.126				0.105			
	0.05	0.070				0.053			
	0.01	0.021				0.016			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.152	0.120	0.100	0.095	0.142	0.113	0.097	0.093
	0.05	0.067	0.049	0.038	0.034	0.066	0.054	0.046	0.044
	0.01	0.011	0.004	0.004	0.003	0.011	0.011	0.010	0.008
C_n^{*}	0.1	0.159	0.121	0.104	0.110	0.147	0.116	0.099	0.103
	0.05	0.073	0.053	0.043	0.049	0.070	0.057	0.047	0.058
	0.01	0.011	0.005	0.004	0.007	0.012	0.010	0.009	0.011
$m(x) = 1 + \sin(10x), \quad p = 1$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.099				0.106			
	0.05	0.052				0.054			
	0.01	0.012				0.011			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.153	0.130	0.142	0.310	0.129	0.120	0.127	0.509
	0.05	0.075	0.062	0.059	0.156	0.072	0.061	0.064	0.272
	0.01	0.011	0.009	0.009	0.021	0.016	0.011	0.014	0.048
C_n^{*}	0.1	0.157	0.142	0.195	0.335	0.132	0.123	0.169	0.595
	0.05	0.082	0.070	0.093	0.181	0.074	0.063	0.080	0.348
	0.01	0.012	0.011	0.013	0.031	0.017	0.011	0.021	0.075

TABLE 1 (*Continued*)

$m(x) = 1 + \sin(10x)$, $p = 2$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.132				0.111			
	0.05	0.076				0.055			
	0.01	0.019				0.015			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.155	0.121	0.110	0.177	0.138	0.114	0.110	0.238
	0.05	0.065	0.051	0.043	0.083	0.063	0.055	0.050	0.109
	0.01	0.010	0.004	0.005	0.010	0.012	0.010	0.009	0.015
C_n^{*}	0.1	0.165	0.134	0.134	0.190	0.143	0.119	0.126	0.280
	0.05	0.073	0.061	0.062	0.095	0.067	0.059	0.062	0.109
	0.01	0.011	0.005	0.007	0.012	0.012	0.010	0.013	0.015

Bootstrap tests are based on 2000 bootstrap samples, $h = Cn^{-1/3}$ for $C = 0.25, 0.5, 1, 2$. Model: $Y_i = m(X_i) + \varepsilon_i$, $i = 1, \dots, n$, $X_i \sim U(0,1)$, $Z_i^{(1)} \sim U(0,1)$, $Z_i^{(2)} \sim U(0,1)$, $\varepsilon_i \sim N(0,1)$ independent.

number of oscillations in the interval (0,1), higher bandwidth values produce serious size distortions. As in other simulation studies for specification tests of parametric functions based on smoothers, it seems advisable, in practice, to undersmooth, rather than oversmooth, in order to obtain good level accuracy. The size properties of the test are not very affected by the dimension of the vector Z. We present simulation results for the two bootstrap procedures (i.e., C_n^{**} and C_n^{*}.) The bootstrap test based on C_n^{**} performs slightly better than the test based on C_n^{*}.

Table 2 examines the power properties of the test under the alternative ($\beta = 1$), for $q = 1$ and $p = 1, 2$. We consider $\gamma = 5, 10$. The correlation between Y_i and $Z_i^{(1)}$ is close to 1 when $\gamma = 5$ and to 0 when $\gamma = 10$. Therefore, the power of the Wald's test decreases as γ increases. When $\gamma = 5$, all the tests are very powerful. When $\gamma = 10$, the power of the Wald's test is very close to the theoretical size. However, the bootstrap tests are still powerful, though bigger sample sizes than in the previous case are needed. The results are quite insensitive to the choice of smoothing parameter and the dimension of the vector Z.

In Table 3, we examine the level accuracy of the bootstrap test under the null ($\beta = 0$) when $q = 2$ and we only report results for $p = 1$. In this case, a kernel of order higher than two is needed, according to our assumptions. In order to illustrate the sensitivity of the test to the order of the kernel chosen, we report simulations for Epanechnikov's kernels of order 2 and 4. We consider the linear model $m(x) = 1 + x^{(1)} + x^{(2)}$. We observe that the kernel of order 4 is less affected by extreme bandwidth choices. Here, we also report results for $n = 200$. As it could be expected, greater sample sizes must be used when $q = 2$ than when $q = 1$.

TABLE 2

Proportion of rejections in 2000 *Monte Carlo samples, under* $H_1 : E(Y|W) \neq E(Y|X)$ *a.s.*, $p = 1, 2$ *for the bootstrap test and an asymptotic t-ratio based on a linear regression model*

$\gamma = 5, \quad p = 1$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.980				1.000			
	0.05	0.963				1.000			
	0.01	0.878				0.997			
	$\alpha \backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.950	0.962	0.960	0.927	1.000	1.000	1.000	0.999
	0.05	0.901	0.921	0.922	0.866	0.998	1.000	1.000	0.998
	0.01	0.661	0.724	0.734	0.643	0.988	0.991	0.990	0.982
C_n^{*}	0.1	0.954	0.964	0.962	0.933	1.000	1.000	1.000	0.999
	0.05	0.914	0.924	0.919	0.877	0.998	1.000	1.000	0.999
	0.01	0.710	0.740	0.731	0.629	0.990	0.990	0.991	0.982
$\gamma = 5, \quad p = 2$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.971				1.000			
	0.05	0.944				1.000			
	0.01	0.824				0.992			
	$\alpha \backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.623	0.790	0.837	0.609	0.974	0.982	1.000	0.937
	0.05	0.437	0.672	0.728	0.476	0.942	0.969	1.000	0.880
	0.01	0.138	0.316	0.409	0.222	0.812	0.903	0.997	0.688
C_n^{*}	0.1	0.766	0.826	0.844	0.656	0.976	0.985	1.000	0.968
	0.05	0.585	0.716	0.742	0.518	0.944	0.974	0.986	0.928
	0.01	0.279	0.387	0.408	0.219	0.823	0.916	0.939	0.743
$\gamma = 10, \quad p = 1$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.096				0.096			
	0.05	0.054				0.052			
	0.01	0.013				0.012			
	$\alpha \backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.445	0.434	0.404	0.387	0.788	0.810	0.806	0.783
	0.05	0.252	0.247	0.226	0.218	0.603	0.610	0.607	0.590
	0.01	0.058	0.057	0.047	0.046	0.238	0.236	0.230	0.227
C_n^{*}	0.1	0.464	0.444	0.421	0.430	0.798	0.812	0.814	0.823
	0.05	0.269	0.252	0.242	0.262	0.613	0.614	0.619	0.634
	0.01	0.066	0.060	0.053	0.071	0.247	0.241	0.241	0.263

TABLE 2 (*Continued*)

$\gamma = 10, \quad p = 2$									
	α	$n = 50$				$n = 100$			
t-ratio	0.1	0.130				0.105			
	0.05	0.073				0.057			
	0.01	0.016				0.019			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.366	0.374	0.313	0.202	0.487	0.631	0.622	0.618
	0.05	0.167	0.205	0.167	0.090	0.315	0.421	0.419	0.418
	0.01	0.020	0.026	0.020	0.013	0.088	0.123	0.111	0.116
C_n^{*}	0.1	0.585	0.427	0.339	0.219	0.616	0.656	0.638	0.742
	0.05	0.355	0.258	0.193	0.112	0.448	0.456	0.442	0.580
	0.01	0.105	0.042	0.030	0.020	0.160	0.141	0.131	0.279

Bootstrap tests are based on 2000 bootstrap samples, $h = Cn^{-1/3}$ for $C = 0.25, 0.5, 1, 2$. Model $Y_i = 1 + X_i + \sin(\gamma Z_i) + \varepsilon_i,\ i = 1, \dots, n,\ X_i \sim U(0,1),\ Z_i^{(1)} \sim U(0,1),\ Z_i^{(2)} \sim U(0,1)$, $\varepsilon_i \sim N(0,1)$ independent.

In Table 4, we report the proportion of rejections under the alternative ($\beta = 1$) with $q = 2$. We consider m as in Table 3, and $\gamma = 5$ and 10, as in Table 2. The results are similar to the case where $q = 1$, though this comparison is not fair, since the test over-rejects when $q = 2$.

5. Testing other restrictions on regression curves. Different restrictions on nonparametric regression curves can be tested applying the methodology developed in preceding sections. Suppose we want to test

$$H_0 : E\left(Y \middle| W\right) = m_0(W) \qquad \text{a.s.},$$

where m_0 is the regression function when certain restrictions have been imposed; for example, mean independence is the case considered before. Other restrictions could be partial linearity, monotonicity, additivity, etc. The null hypothesis can be alternatively be written as

$$H_0 : T(W) = 0 \qquad \text{a.s.},$$

where $T(w) = E\left[(Y - m_0(W))\,\eta(W)\,1(W \le w)\right]$, and η is a weight function which does not change sign in the support of W. Let $\hat{m}_0$ and $\hat{\eta}$ be suitable estimates of m_0 and η respectively. A test can be based on the U-process,

$$Q_n(w) = \frac{1}{n} \sum_i (Y_i - \hat{m}_0(W_i))\,\hat{\eta}(W_i)\,1_w(W_i).$$

The choice of η, the limiting distribution of Q_n, and the construction of bootstrap tests will depend on the particular testing problem. Here, we only discuss the implementation of this methodology for testing partial linearity and conditional independence. However, application to tests of other restrictions seems also possible.

TABLE 3

Proportion of rejections in 2000 *Monte Carlo samples, under* $H_0 : E(Y|W) = E(Y|X)$ *a.s.,* $p = 1$, *for the bootstrap test and an asymptotic t-ratio based on a linear regression model*

$m(x) = 1 + x^{(1)} + x^{(2)}$ Kernel of order 2.

	α	$n = 50$				$n = 100$				$n = 200$			
$t-$ ratio	0.1	0.115				0.101				0.098			
	0.05	0.057				0.045				0.050			
	0.01	0.012				0.009				0.010			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.262	0.158	0.114	0.338	0.248	0.118	0.127	0.629	0.211	0.123	0.128	0.790
	0.05	0.113	0.067	0.045	0.155	0.135	0.054	0.059	0.446	0.108	0.053	0.063	0.619
	0.01	0.011	0.006	0.003	0.015	0.020	0.010	0.004	0.132	0.025	0.011	0.013	0.273
C_n^{*}	0.1	0.344	0.190	0.190	0.421	0.286	0.143	0.209	0.716	0.227	0.140	0.262	0.917
	0.05	0.185	0.095	0.100	0.219	0.159	0.065	0.117	0.601	0.120	0.065	0.146	0.844
	0.01	0.037	0.010	0.019	0.047	0.028	0.014	0.004	0.324	0.030	0.014	0.043	0.616

$m(x) = 1 + x^{(1)} + x^{(2)}$ Kernel of order 4.

	α	$n = 50$				$n = 100$				$n = 200$			
$t-$ ratio	0.1	0.115				0.101				0.098			
	0.05	0.057				0.045				0.050			
	0.01	0.012				0.009				0.010			
	$\alpha\backslash C$	0.25	0.5	1	2	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.1	0.280	0.193	0.116	0.203	0.226	0.145	0.108	0.215	0.175	0.150	0.113	0.200
	0.05	0.118	0.085	0.048	0.092	0.119	0.074	0.045	0.103	0.098	0.074	0.050	0.102
	0.01	0.011	0.009	0.004	0.009	0.022	0.013	0.006	0.013	0.030	0.017	0.010	0.018
C_n^{*}	0.1	0.504	0.235	0.134	0.312	0.331	0.163	0.165	0.380	0.229	0.157	0.123	0.425
	0.05	0.261	0.117	0.058	0.194	0.188	0.009	0.091	0.254	0.131	0.079	0.057	0.294
	0.01	0.069	0.017	0.006	0.006	0.043	0.017	0.019	0.008	0.039	0.016	0.012	0.111

Bootstrap tests are based on 2000 bootstrap samples, $h = Cn^{-1/6}$ for $C = 0.25, 0.5, 1, 2$. Model: $Y_i = 1 + m(X_i) + \varepsilon_i$, $i = 1, \ldots, n$, $X_i^{(k)} \sim U(0,1)$, $Z_i^{(k)} \sim U(0,1)$ $k = 1, 2$, $\varepsilon_i \sim N(0,1)$ independent.

5.1. *Specification testing of partially linear models.* The partially linear model is a compromise between the linear and the nonparametric regression model. It permits one to reduce the curse of dimensionality in the estimation of a nonparametric curve. Estimators of this model have been proposed by Heckman (1986), Robinson (1988) and Speckman (1988) among others. Consider the null hypothesis

$$H_0 : E(Y \mid W) = Z'\theta_0 + \gamma(X) \qquad \text{a.s. for some } \theta_0 \in \Theta \subset \mathbb{R}^m,$$

where θ_0 is an unknown parameter vector belonging to the parameter space Θ, and γ is an unknown function. Henceforth, a' means a transpose. Noticing that $\gamma(\cdot) = m(\cdot) - m_Z(\cdot)'\theta_0$, where $m_Z(\cdot) = E(Z| X = \cdot)$, the null hypothesis

TABLE 4

Proportion of rejections in 2000 *Monte Carlo samples, under* $H_1 : E(Y|W) \neq E(Y|X)$ *a.s.,* $p = 1$, *for the bootstrap test and an asymptotic t-ratio based on a linear regression model*

		$\gamma = 5$							
	α	$n = 50$				$n = 100$			
t-ratio	0.10	0.969				1.000			
	0.05	0.944				1.000			
	0.01	0.847				0.997			
	$\alpha \backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.10	0.623	0.790	0.837	0.609	0.837	0.982	0.995	0.937
	0.05	0.437	0.672	0.728	0.476	0.793	0.969	0.985	0.880
	0.01	0.138	0.316	0.409	0.222	0.622	0.903	0.938	0.688
C_n^{*}	0.10	0.766	0.826	0.844	0.656	0.883	0.985	0.995	0.968
	0.05	0.585	0.716	0.742	0.518	0.842	0.974	0.986	0.928
	0.01	0.279	0.387	0.408	0.219	0.684	0.916	0.939	0.743
		$\gamma = 10$							
	α	$n = 50$				$n = 100$			
t-ratio	0.10	0.103				0.093			
	0.05	0.045				0.048			
	0.01	0.008				0.008			
	$\alpha \backslash C$	0.25	0.5	1	2	0.25	0.5	1	2
C_n^{**}	0.10	0.366	0.374	0.313	0.356	0.487	0.631	0.622	0.618
	0.05	0.167	0.205	0.167	0.191	0.315	0.421	0.419	0.418
	0.01	0.020	0.026	0.020	0.027	0.088	0.123	0.111	0.116
C_n^{*}	0.10	0.585	0.427	0.339	0.447	0.616	0.656	0.638	0.742
	0.05	0.355	0.258	0.193	0.303	0.448	0.456	0.442	0.580
	0.01	0.105	0.042	0.030	0.105	0.160	0.141	0.131	0.279

Bootstrap tests are based on 2000 bootstrap samples, $h = Cn^{-1/6}$ for $C = 0.25, 0.5, 1, 2$. Model $Y_i = 1 + X_i^{(1)} + X_i^{(2)} + \sin(\gamma Z_i) + \varepsilon_i,\ i = 1, \ldots, n,\ X_i^{(k)} \sim U(0,1),\ Z_i^{(k)} \sim U(0,1),\ k = 1, 2,$ $\varepsilon_i \sim N(0,1)$ independent.

can be also written as

$$H_0 : E\left(Y - m(X) - (Z - m_Z(X))' \theta_0 \mid W\right) = 0$$
$$\text{a.s. for some } \theta_0 \in \Theta \subset \mathbb{R}^m,$$

Fan and Li (1996) have considered a test of H_0 based on a distance between the semiparametric model fit and the nonparametric fit using the whole set of regressors W. As in Section 2, we propose a test which only requires estimates of conditional expectations given X, $m(\cdot)$ and $m_Z(\cdot)$. Given a $\sqrt{n}$*–consistent* estimator of θ_0, $\hat{\theta}_n$ say, as proposed by Robinson (1988), the test statistic is based on the U-process,

$$G_n(w) = \frac{1}{n}\sum_i \hat{\varepsilon}_i^s \hat{f}(X_i) 1_w(W_i),$$

where $\hat{\varepsilon}_i^s = [Y_i - \hat{m}(X_i) - \hat{\theta}_n'[Z_i - \hat{m}_Z(X_i)]]$ and

$$\hat{m}_Z(X_i) = (na^q)^{-1}\sum_j Z_j K_{ij}/\hat{f}(X_i)$$

estimates $m_Z(X_i)$. It seems fairly straightforward to obtain, under regularity conditions in Robinson (1988) and the results in Section 2, that $\sup_w |G_n(w) - G_n^o(w)| = o_p(n^{-1/2})$, where

$$G_n^o(w) = \frac{1}{n}\sum_i \varepsilon_i^s f(X_i)[1_w(W_i) - \phi_w(X_i)],$$

with $\varepsilon_i^s = \left[Y_i - m(X_i) - \theta_0'(Z_i - m_Z(X_i))\right]$. A bootstrap version of $\tilde{U}_n^s$ is

$$G_n^{o*}(w) = \frac{1}{n}\sum_i V_i \hat{\varepsilon}_i^s \hat{f}(X_i)\left[1_w(W_i) - \hat{\phi}_w(X_i)\right].$$

Using similar conditions and arguments as in Theorem 2, it can be shown that the resulting test is consistent. The bootstrap analog of the process can be obtained from the resample $\mathscr{Y}_n^* = \{(Y_i^*, X_i), i = 1, \ldots, n\}$, where $Y_i^* = Z_i'\hat{\theta}_n + \hat{m}(X_i) - \hat{m}_Z(X_i)'\hat{\theta}_n + V_i\hat{\varepsilon}_i^s$. The consistency of the corresponding bootstrap test can be proved using similar arguments as in Theorem 3, exploiting its relationship with the previous bootstrap.

5.2. *Testing conditional independence.* Suppose we want to test that the conditional distribution of Y given W does not depend on Z. That is, the null hypothesis is

$$H_0 : E\left[1_y(Y)\middle| W\right] = E\left[1_y(Y)\middle| X\right] \qquad \text{a.s.} \ \forall y \in S_Y,$$

where S_Y is the support of Y. In fact, we are testing the significance of Z, for all y, in a nonparametric regression curve where the dependent variable is $1_y(Y)$. The null hypothesis can be alternatively written as

$$H_0 : L(Y, W) = 0 \qquad \text{a.s.,} \tag{3}$$

where $L(y, w) = E\left[f(X)\left(1_y(Y) - F(y|X)\right)1_w(W)\right]$, and $F(\cdot \mid \cdot)$ is the distribution function of Y given X. Let $\hat{\varepsilon}_i(y) = 1_y(Y) - \hat{F}_n(y|X)$ be the estimator of $\varepsilon_i(y) = 1_y(Y_i) - F(y|X_i)$, where

$$\hat{F}_n(y|X_i) = \frac{1}{\hat{f}(X_i)}\frac{1}{na^q}\sum_i 1_y(Y_i) K_{ij}.$$

The curve $L(y, w)$ is estimated by

$$L_n(y, w) = \frac{1}{n}\sum_i \hat{\varepsilon}_i(y)\hat{f}(X_i)1_w(W_i).$$

Thus, L_n is identical to T_n, but with $\hat{\varepsilon}_i$ substituted by $\hat{\varepsilon}_i(y)$. Thus, reasoning as in Section 2, $\sup_{w,y}\left|L_n(y, w) - L_n^0(y, w)\right| = o_p\left(n^{-1/2}\right)$, where

$$n^{1/2}L_n^0(y, w) = \frac{1}{n^{1/2}}\sum_i \varepsilon_i(y) f(X_i)\left[1_w(W_i) - \phi_w(X_i)\right].$$

A bootstrap version of $L_n^0(y, w)$ is

$$L_n^{0*}(y, w) = \frac{1}{n}\sum_i V_i\hat{\varepsilon}_i(y)\hat{f}(X_i)\left[1_w(W_i) - \hat{\phi}_w(X_i)\right].$$

The consistency of the resulting bootstrap test can be proved using similar arguments than in the proof of Theorem 2. In order to take into account the bias, we could also use

$$L_n^*(y, w) = L_n^{0*}(y, w) + \frac{1}{n}\sum_i \hat{f}(X_i)\hat{F}_n(y|X_i)\left[1_w(W_i) - \hat{\phi}_w(X_i)\right],$$

which takes into account the bias term, and its consistency is proved using similar arguments to those in the proof of Theorem 3.

6. Proofs. The next two propositions and the lemmas in Section 7 provide the basic tools for proving the results in this paper. The first one is a moment inequality for degenerated U-processes of any degree indexed by a general class of functions, which is useful for showing that the remainder term in the Hoeffding decomposition vanishes uniformly. The second proposition is a general result for the asymptotic equivalence between perturbed empirical processes indexed by general classes of functions, which appear in the Hayék projections of the U-processes considered in this paper, and their non-perturbed versions.

Let $\mathscr{G} = \{g(\zeta_1, \ldots, \zeta_m)\}$ be a class of real functions of m variables where the $\zeta_i, i = 1, \ldots, n$ are iid with common probability space $(B, \mathscr{B}, Q)$, such that its envelope, $G = \sup_{g\in\mathscr{G}}|g|$, is measurable – otherwise take as G the least measurable majorant of $\sup_{g\in\mathscr{G}}|g|$. With the notation $Q_1 \times \cdots \times Q_m g = \int g d(Q_1 \times \cdots \times Q_m)$, the Hoeffding projections of $g : B^m \mapsto \mathbb{R}$ are defined as

$$\pi_k g = \left(\delta_{\zeta_1} - Q\right) \times \cdots \times \left(\delta_{\zeta_k} - Q\right) \times Q^{m-k} g, \ \ k = 1, \ldots, m.$$

If g is symmetric in its entries, these projections induce the Hoeffding decomposition

$$U_n^{(m)}(g) - Q^m g = \sum_{k=1}^m \binom{m}{k} U_n^{(k)}(\pi_k g)$$

where

$$U_n^{(r)}(d) = \frac{1}{n(n-1)\cdots(n-r+1)} \sum_{1\le i_1\neq\cdots\neq i_r\le n} d\left(\zeta_{i_1},\ldots,\zeta_{i_r}\right).$$

A Vapnik-Červonenkins type (VC-type) class of functions is a class of functions such that for every Q for which $QG^2 < \infty$, we have

$$\mathscr{N}(\varepsilon, L_2(Q), \mathscr{G}) \le \left(\frac{A\,\|G\|_{L_2(Q)}}{\varepsilon}\right)^v, \tag{4}$$

where $\mathscr{N}(\varepsilon, L_2(Q), \mathscr{G})$ is the smallest number of $L_2(Q) - balls$ of radius less or equal to ε and centers in $\mathscr{G}$ needed to cover $\mathscr{G}$. We call A and v the VC-characteristics of the class. $\mathscr{G}$ might not be a VC-type for its envelope, but for some other function $J \ge G$ [meaning (4) holds with J replacing G]. We thank Evarist Giné for bringing the next proposition to our attention.

PROPOSITION 4. *Let $\mathscr{G}$ be a class of kernels of m variables with envelope G, symmetric in their entries, which is VC-type, and let Q be the common law of the iid ζ_i, $i = 1,\ldots,n$ variables (which take values in a measurable space B, the g's in $\mathscr{G}$ are $B^m \to \mathbb{R}$.) Suppose $P^m G^2 < \infty$. Then, there is a constant, C say, that depends on the VC characteristics of $\mathscr{G}$ such that, for all $k = 1,\ldots,m$,*

$$E\left\{\sup_{g\in\mathscr{G}}\left|\frac{1}{n^{k/2}}\sum_{1\le i_1\neq\ldots\neq i_k\le n}(\pi_k g)\left(\zeta_{i_1},\ldots,\zeta_{i_k}\right)\right|^2\right\} \le CQ^m G^2.$$

PROOF. We use decoupling, Rademacher randomization and the chaos inequalities in de la Peña and Giné [(1999), Corollary 5.1.8, page 221, together with Lemma 5.3.5 and the results on page 246, and taking Remark 5.3.9 into account]. A similar result can be found, for $m = 2$, in Nolan and Pollard [(1987) Theorem 6]; see also Sherman [(1994), main corollary on page 447-448] and Ghosal, Sen and Van der Vaart [(2000), Theorem A1 and A2].

PROPOSITION 5. *Let $\mathscr{G}$ be a Donsker class of measurable functions that is closed under translation, and let ζ_i, $i = 1,\ldots,n$ be independent copies of ζ (taking values in a measurable space B, the g's in $\mathscr{G}$ are $B \to \mathbb{R}$). Let μ_n be non-random, signed measures of uniformly bounded variation, that converge to the Dirac measure at zero. Then*

$$\sup_{g\in\mathscr{G}}\left|\frac{1}{n}\sum_i \int \left[g(\zeta_i + y) - g(\zeta_i)\right]\mu_n(dy)\right| = o_p\left(n^{-1/2}\right),$$

if and only if

$$\sup_{g\in\mathscr{G}} E\left[\left(\int \left[g(\zeta + y) - g(\zeta)\right] d\mu_n(y)\right)^2\right] = o(1)$$

and

$$\sup_{g\in\mathscr{G}}\left|E\left[\int\left[g\left(\zeta+y\right)-g\left(\zeta\right)\right]d\mu_n\left(y\right)\right]\right|=o\left(n^{-1/2}\right).$$

PROOF. See the theorem in Van der Vaart (1994).

PROOF OF PROPOSITION 1. Write

$$R_n\left(s_1,s_2\right)=U_n^{(2)}\left(\pi_2\beta_{sa}\right)$$

where $s=\left(s_1,s_2\right)$, and

$$\beta_{sa}\left(\chi_i,\chi_j\right)=\tfrac{1}{2}\psi_a\left(\chi_i,\chi_j\right)\left[1_{s_1}\left(\chi_i\right)1_{s_2}\left(\chi_j\right)-1_{s_1}\left(\chi_j\right)1_{s_2}\left(\chi_i\right)\right]$$

are symmetric functions depending on a. The set of functions $\{1_s : s\in\mathbb{R}^{p+q+1}\}$ is VC-type [e.g., Example 2.6.1 in van der Vaart and Wellner (1996)] and therefore, so is the class $\{\psi_a 1_s : s\in\mathbb{R}^{p+q+1}\times\mathbb{R}^{p+q+1}\}$ for each $a\in\mathbb{R}^+\setminus\{0\}$ [see, e.g., Lemma 2.6.18 (vi) in van der Vaart and Wellner (1996)], which has envelope $|\psi_a|$ and CV-characteristics independent of ψ_a. Now, if $\mathscr{G}=\{g_1+g_2 : g_1\in\mathscr{G}_1 \text{ and } g_2\in\mathscr{G}_2\}$ and $\mathscr{G}_1$ and $\mathscr{G}_2$ are VC with envelopes G_1 and G_2 respectively, then, $\mathscr{G}_1+\mathscr{G}_2$ is also VC with envelope G_1+G_2 [this statement follows easily from the fact that if $g_1,\dots,g_N$ are the centers of an $\varepsilon/2-cover$ of $\mathscr{G}_1$ for $L_2(Q)$ and $\bar{g}_1,\dots,\bar{g}_{\bar{N}}$ are those for an $\varepsilon/2-cover$ of $\mathscr{G}_2$ for $L_2(Q)$, then $g_i+\bar{g}_j,\ i=1,\dots,N,\ j=1,\dots,\bar{N}$, are centers of a cover of $\mathscr{G}$ by a $L_2(Q)$-ball of radius less or equal to ε]. Therefore, the class of functions indexed by s, $\{\beta_{sa} : s\in\mathbb{R}^{p+q+1}\times\mathbb{R}^{p+q+1}\}$ for fixed $a\in\mathbb{R}^+\setminus\{0\}$, which is the sum of two VC-type classes, is also VC-type, with envelope $|\psi_a|$ and VC characteristics independent of ψ_a. Thus, applying Proposition 4, there exists a constant C, which does not depend on a, such that

$$\begin{aligned}E\left[\sup_{s_1,s_2}\left|n^{1/2}R_n\left(s_1,s_2\right)\right|^2\right]&\le C\frac{1}{n}E\left[\psi_a\left(\chi_1,\chi_2\right)^2\right]\\&=C\frac{1}{na^{2q}}E\left[\left(Y_1-Y_2\right)^2K_{12}^2\right]\\&=O\left(\frac{1}{na^q}\right),\end{aligned}$$

by Lemma 3. □

PROOF OF PROPOSITION 2. Write

$$\eta_a\left(\bar{y},\bar{x}\right)=\int\left(\bar{y}-m\left(\tilde{x}\right)\right)f\left(\tilde{x}\right)\frac{1}{a^q}K\left(\frac{\bar{x}-\tilde{x}}{a}\right)d\tilde{x}-\left(\bar{y}-m\left(\bar{x}\right)\right)f\left(\bar{x}\right)$$

and

$$\left\{\bar{U}_n^{(1)}\left(w\right)-\tilde{U}_n^{(1)}\left(w\right),\ w\in\mathbb{R}^{p+q}\right\}:=\left\{\left(P_n-P\right)g,\ g\in\left\{1_w\eta_a : w\in\mathbb{R}^{p+q}\right\}\right\}.$$

That is, reasoning as above, under H_0, $\bar{U}_n^{(1)} - \tilde{U}_n^{(1)}$ is a centered empirical process indexed by a VC-type class of functions with VC characteristics independent of η_a and envelope $|\eta_a|$. Therefore, applying Proposition 4 with $m = 1$, there exists a constant C independent of a, such that

$$\begin{aligned}
&E\left[\sup_w \left|n^{1/2}\left(\bar{U}_n^{(1)}(w) - \tilde{U}_n^{(1)}(w)\right)\right|^2\right] \\
&\quad \le CE\left[\eta_a(\chi_1)^2\right] \\
&\quad \le 2CE\left[\left|E\left((m(X_1) - m(X_2))\frac{1}{a^q}K_{12}\middle| X_1\right)\right|^2\right] \\
&\qquad + 2CE\left[\left|Y_1 - m(X_1)\right|^2\left|E\left(f(X_1) - \frac{1}{a^q}K_{12}\middle| X_1\right)\right|^2\right].
\end{aligned}$$

The first term in this last expression is an $O(a^{2\min(\lambda+1,\tau)})$ by Lemma 5, and, for $\delta > 0$, the second term is bounded, applying Hölder's inequality by a constant times

$$\left|E\left(\left|Y_1 - m(X_1)\right|^{2+\delta}\right)\right|^{\frac{2}{2+\delta}}\left|E\left(\left|E\left(f(X_1) - \frac{1}{a^q}K_{12}\middle| X_1\right)\right|^{2+\frac{4}{\delta}}\right)\right|^{\frac{\delta}{2+\delta}} = O\left(a^{2\lambda}\right)$$

by Lemma 4. □

PROOF OF PROPOSITION 3. For any $z_1, z_2 \in \mathbb{R}^p$, $E[|r_{z_1}(X) - r_{z_2}(X)|^2] \le E[|1_{z_1}(Z) - 1_{z_2}(Z)|^2]$ by Jensen's inequality, and hence,

$$\sup_Q \mathcal{N}(\varepsilon, L_2(Q), \mathcal{R}) \le \mathcal{N}(\varepsilon, L_2(R), \{1_z : z \in \mathbb{R}^p\}),$$

where R is a fixed probability measure on the support of Z. That is, the covering numbers of $\mathcal{R} = \{r_z : z \in \mathbb{R}^p\}$ are bounded by the covering numbers of the family of functions $\{1_z : z \in \mathbb{R}^p\}$ [see also Ghosal, Sen and Van der Vaart (2000), Lemma A2, for a related result]. Therefore, $\mathcal{R}$ is VC-type with absolute bounded envelope, and $\{r_z \wedge 1_x : z \in \mathbb{R}^p, x \in \mathbb{R}^q\}$ is also VC-type [see Van der Vaart and Wellner (1996), Lemma 2.6.18 (i)]. If $g : S \to \mathbb{R}$ is a fixed function, such that $Pg^2 < \infty$, $\{[1_x \wedge r_z] \cdot g : z \in \mathbb{R}^p, x \in \mathbb{R}^q\}$ is P-Donsker [see, e.g., Van der Vaart and Wellner (1996), Example 2.10.23]. Therefore, taking into account that $\gamma_w(\bar{y}, \bar{x}) = [1_x(\bar{x}) \wedge r_z(\bar{x})] \cdot (m(\bar{x}) - \bar{y}) f(\bar{x})$, and $\sup_x f(x) < \infty$, $E\left[(Y - m(X))^2\right] < \infty$, $\{\gamma_w : w \in \mathbb{R}^{p+q}\}$ is P-Donsker. Now, notice that

$$\begin{aligned}
\bar{U}_n^{(2)}(w) - \tilde{U}_n^{(2)}(w) &= \frac{1}{n}\sum_i \int [\gamma_w(Y_i, X_i + \bar{x}) - \gamma_w(Y_i, X_i)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x} \\
&= \frac{1}{n}\sum_i \int [\gamma_w(Y_i + \bar{y}, X_i + \bar{x}) - \gamma_w(Y_i, X_i)] \mu_n(d\bar{y}, d\bar{x}),
\end{aligned}$$

where

$$\mu_n(B_1 \times B_2) = 1(0 \in B_1) \int_{y \in B_2} \frac{1}{a^q} K\left(\frac{y}{a}\right) dy,$$

is a non-random, signed measure of uniformly bounded variation, which is degenerate at the first coordinate, and converges to the Dirac measure at zero. Thus, applying Proposition 5, it suffices to show that

$$\sup_w E\left[\left(\int [\gamma_w(Y, X+\bar{x}) - \gamma_w(Y, X)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right)^2\right] = o(1) \tag{5}$$

and

$$\sup_w \left|\int E[\gamma_w(Y, X+\bar{x}) - \gamma_w(Y, X)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right| = o\left(n^{-1/2}\right). \tag{6}$$

The left-hand side of (5) is bounded by twice

$$\sup_w E\left\{\left[\int (m(X+\bar{x}) - m(X))(f \cdot 1_x \cdot r_z)(X+\bar{x}) \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right]^2\right\} \tag{7}$$

$$+\sup_w E\left\{(Y - m(X))^2 \right. \tag{8}$$

$$\left. \times \left[\int [(f \cdot 1_x \cdot r_z)(X+\bar{x}) - (f \cdot 1_x \cdot r_z)(X)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right]^2\right\}.$$

Equation (7) is bounded by

$$\begin{aligned} &E\left\{\left[\int \left|m(X+\bar{x}) - m(X)\right| f(X+\bar{x}) \frac{1}{a^q} \left|K\left(\frac{\bar{x}}{a}\right)\right| d\bar{x}\right]^2\right\} \\ &\quad = E\left[\left|E\left(\left|m(X_2) - m(X_1)\right| \frac{1}{a^q} \left|K\left(\frac{X_1 - X_2}{a}\right)\right| \middle| X_1\right)\right|^2\right] \\ &\quad = O\left(a^2\right) \end{aligned}$$

by Lemma 6; and (8) is bounded, applying Hölder's inequality, by

$$\left[E\left((Y - m(X))^{2+\delta}\right)\right]^{\frac{2}{2+\delta}}$$

$$\times \left\{\sup_{x,z} E\left[\left|\int [(f \cdot 1_x \cdot r_z)(X+\bar{x}) - (f \cdot 1_x \cdot r_z)(X)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right|^{2+\frac{4}{\delta}}\right]\right\}^{\frac{\delta}{2+\delta}},$$

with $\delta > 0$. Now, applying Jensen's inequality, for $\alpha = 2 + 4/\delta$,

$$\begin{aligned} &\sup_{x,z} E\left[\left|\int [(f \cdot 1_x \cdot r_z)(X+\bar{x}) - (f \cdot 1_x \cdot r_z)(X)] \frac{1}{a^q} K\left(\frac{\bar{x}}{a}\right) d\bar{x}\right|^{\alpha}\right] \\ &\quad \le C \sup_{x,z} \int E\left[\left|(f \cdot 1_x \cdot r_z)(X+\bar{x}) - (f \cdot 1_x \cdot r_z)(X)\right|^{\alpha}\right] \frac{1}{a^q} \left|K\left(\frac{\bar{x}}{a}\right)\right| d\bar{x}, \end{aligned}$$

and this last expression is bounded by a constant times

$$\sup_{x,z}\int E\left[\left|1_x(X+\bar{x})\,r_z(X+\bar{x})\,[f(X+\bar{x})-f(X)]\right|^{\alpha}\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$+\sup_{x,z}\int E\left[\left|f(X)\,1_x(X+\bar{x})\,[r_z(X+\bar{x})-r_z(X)]\right|^{\alpha}\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$+\sup_{x,z}\int E\left[\left|f(X)\,r_z(X)\,[1_x(X+\bar{x})-1_x(X)]\right|^{\alpha}\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$\le C\int E\left[\left|f(X+\bar{x})-f(X)\right|\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x} \tag{9}$$
$$+C\sup_{z}\int E\left[\left|r_z(X+\bar{x})-r_z(X)\right|\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$+C\sup_{x}\int E\left[\left|1_x(X+\bar{x})-1_x(X)\right|\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$=O(a),$$

applying Lemma 6, and noticing that

$$\int E\left[\left|1_x(X+\bar{x})-1_x(X)\right|\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$\le\sum_{j=1}^{q}\int E\left[\left|1_{x^{(j)}}\left(X^{(j)}+\bar{x}^{(j)}\right)-1_{x^{(j)}}\left(X^{(j)}\right)\right|\right]\frac{1}{a^q}\left|K\left(\frac{\bar{x}}{a}\right)\right|d\bar{x}$$
$$\le C\sum_{j=1}^{q}\int\left|F_{X^{(j)}}\left(x^{(j)}-\bar{x}^{(j)}\right)-F_{X^{(j)}}\left(x^{(j)}\right)\right|\frac{1}{a}\left|k\left(\frac{\bar{x}^{(j)}}{a}\right)\right|d\bar{x}^{(j)},$$

where $X=(X^{(1)},\dots,X^{(q)})'$, $x=(x^{(1)},\dots,x^{(q)})'$, $\bar{x}=(\bar{x}^{(1)},\dots,\bar{x}^{(q)})'$, and that $F_{X^{(j)}}$, f and r_z are Lipschitz. Finally, the left hand side of (6) is equal to

$$\sup_{x,z}\left|\int E[m(X)-m(X+\bar{x})]\,f(X+\bar{x})\,1_x(X+\bar{x})\,r_z(X+\bar{x})\frac{1}{a^q}K\left(\frac{\bar{x}}{a}\right)d\bar{x}\right|$$
$$=\sup_{x,z}\left|E\left[\left[m(X_1)-m(X_2)\frac{1}{a^q}\right]K\left(\frac{X_1-X_2}{a}\right)1_x(X_2)\,r_z(X_2)\right]\right|$$
$$\le E\left[\left|E\left[[m(X_1)-m(X_2)]\frac{1}{a^q}K\left(\frac{X_1-X_2}{a}\right)\right|X_2\right]\right|\right]$$
$$=O\left(a^{\min(\lambda+1,\tau)}\right),$$

by Lemma 5. □

PROOF OF THEOREM 1.

$$\sup_w \left|T_n(w) - \tilde{U}_n(w)\right|$$
$$= \sup_w \left|\frac{n-1}{n}\left(\bar{U}_n(w) - \tilde{U}_n(w)\right) - \frac{1}{n}\tilde{U}_n(w) + \frac{n-1}{n}R_n(w,\infty;\infty,\infty)\right|$$
$$= o_p\left(n^{-1/2}\right)$$

since $\sup_w |\tilde{U}_n(w)| = O_p(1)$, $\sup_w |R_n(w,\infty;\infty,\infty)| = o_p(n^{-1/2})$ by Proposition 1, and $\sup_w |\bar{U}_n(w) - \tilde{U}_n(w)| = o_p(n^{-1/2})$ by Propositions 2 and 3. The family of functions $\mathscr{H} = \mathscr{H}^1 + \mathscr{H}^2$, where $\mathscr{H}^1 = \{\gamma_w : w \in \mathbb{R}^{p+q}\}$ and $\mathscr{H}^2 = \{\omega_w : w \in \mathbb{R}^{p+q}\}$, with $\omega_w(\chi) = 1_w(W)(Y - m(X))f(X)$. In the proofs of Proposition 1 and 3, we have seen that $\mathscr{H}^1$ and $\mathscr{H}^2$ are VC-type with common square P-integrable envelope proportional to $|Y - m(X)|$. Since the sum of two VC-type classes of functions is also VC-type with envelope the sum of the envelopes (see the proof of Proposition 1), $\mathscr{H}$ is a VC-type class with square P-integrable envelope and, therefore, is P-Donsker (see the proof of Proposition 3). □

PROOF OF COROLLARY 1. Notice that the maps $T \mapsto \sup_w \left|T(w)\right|$ and $T \mapsto \int T^2(w)\,dF_W(w)$ are continuous, dF_W being a bounded measure, and the T's being bounded functionals. Thus, convergence in distribution of K_n under H_0 is a consequence of Theorem 1 and the Continuous Mapping Theorem (CMT). By Propositions 1, 2 and 3,

$$C_n = \int \left|n^{1/2}T_n(w)\right|^2 dF_W(w) + \int \left[n^{1/2}\tilde{U}_n(w)\right]^2 (dF_{Wn}(w) - dF_W(w)) + o_p(1).$$

The first term on the right hand side of the last expression converges in distribution to $\int \left|B_P\xi_w\right|^2 dF_W(w)$ by the Theorem 1 and the CMT, and the second term is equal to

$$\frac{1}{n^2}\sum_{i\neq j\neq k}\{\xi_{W_k}(\chi_i)\xi_{W_k}(\chi_j) - E[\xi_{W_k}(\chi_i)\xi_{W_k}(\chi_j)|\chi_i,\chi_j, j\neq i\neq k]\}$$
$$+\frac{1}{n^2}\sum_{i\neq k}\left\{\xi_{W_k}(\chi_i)^2 - E\left[\xi_{W_k}(\chi_i)^2\middle|\chi_i, i\neq k\right]\right\} \tag{10}$$
$$= o_p(1).$$

Since the first term of (10) is a completely degenerate U-statistic, noticing that, under H_0,

$$E\left[\xi_{W_k}(\chi_j)\middle|\chi_k, j\neq k\right]$$
$$= E\left[E\left(\varepsilon_j\middle|W_j\right)f\left(X_j\right)1_{X_k}\left(X_j\right)\left(1_{Z_k}\left(Z_j\right) - r_{Z_k}\left(X_j\right)\right)\middle|\chi_k, j\neq k\right] = 0,$$

and applying Proposition 4 version for U-statistics it is $O_p\left(n^{-1/2}\right)$. The second term of (10) is centered at zero and, thus, it is an $o_p(1)$ applying the LLN for $U-$statistics [see, e.g., de la Peña and Giné (1999), Theorem 4.1.4]. □

In order to simplify notation, henceforth, for any generic function γ, $\gamma_i = \gamma(\chi_i)$ and $\hat{\gamma}_i = \hat{\gamma}(\chi_i)$; also $\Pi_{iw} = 1_w(W_i)$, $r_{zi} = r_z(X_i)$, $\phi_{wi} = \phi_w(X_i) := 1_x(X_i) r_z(X_i)$, $\varepsilon_i = Y_i - m_i$ and $\hat{\varepsilon}_i = Y_i - \hat{m}_i$.

PROOF OF THEOREM 2.

$$\begin{aligned} n^{1/2}\left[\tilde{U}_n^*(w) - \tilde{U}_n^{0*}(w)\right] &= \left\{\frac{1}{n^{1/2}}\sum_i V_i \hat{\varepsilon}_i \hat{f}_i \Pi_{wi} - \frac{1}{n^{1/2}}\sum_i V_i \varepsilon_i f_i \Pi_{wi}\right\} \\ &\quad - \left\{\frac{1}{n^{1/2}}\sum_i V_i \hat{\varepsilon}_i \hat{f}_i \hat{\phi}_{wi} - \frac{1}{n^{1/2}}\sum_i V_i \varepsilon_i f_i \phi_{wi}\right\} \\ &= A_{1n}(w) - A_{2n}(w). \end{aligned}$$

It suffices to show that $\sup_w \left|A_{jn}(w)\right| = o_p\left(n^{-1/2}\right)$, $j = 1, 2$. Write

$$A_{1n}(w) = \frac{1}{n^{3/2}}\sum_{i \neq j}\left[\psi_a\left(\chi_i, \chi_j\right) - \varepsilon_i f_i\right] V_i \Pi_{wi},$$

where ψ_a was defined in Proposition 1. Thus, reasoning as in the proof of Proposition 1 and taking into account that V_i's are iid and independent of the sample $\mathscr{Y}_n$,

$$\sup_w \left|A_{1n}(w)\right| = \sup_w \left|\frac{1}{n^{1/2}}\sum_i V_i \Pi_{wi} \eta_a(Y_i, X_i)\right| + O_p\left(\frac{1}{n^{1/2} a^{q/2}}\right),$$

where η_a was defined in Proposition 2. The first term on the right hand side of the last expression is $O_p(a^{\lambda}) + O_p(a^{\min(\lambda+1,\tau)})$, reasoning as in the proof of Proposition 2, taking into account that V_i's are iid and bounded with zero mean and independent of the sample $\mathscr{Y}_n$. Thus, $\sup_w |A_{1n}(w)| = o_p(1)$. Now write

$$\begin{aligned} A_{2n}(w) &= \frac{1}{n^{1/2}}\sum_i V_i \varepsilon_i \left[\hat{f}_i \hat{\phi}_{wi} - f_i \phi_{wi}\right] + \frac{1}{n^{1/2}}\sum_i V_i (m_i - \hat{m}_i) \hat{f}_i \hat{\phi}_{wi} \\ &= A_{2n}^1(w) + A_{2n}^2(w). \end{aligned}$$

Now

$$A_{2n}^1(w) = \frac{1}{n^{3/2}}\sum_i V_i \varepsilon_i \left[\frac{1}{a^q} K(0) \Pi_{wi} - f_i \phi_{wi}\right] + \frac{1}{n^{3/2}}\sum_{i \neq j} V_i \varepsilon_i \left[\frac{1}{a^q} K_{ij} \Pi_{wj} - f_i \phi_{wi}\right].$$

The first term on the right hand side of the last expression is bounded , uniformly in w, by a constant times $n^{-3/2}\sum_i \left|\varepsilon_i\right| = O_p\left(n^{-1/2}\right)$ by the LLN. Reasoning as in Proposition 1, the second term is, uniformly in w, equal to

$$\begin{aligned}&\frac{1}{n^{1/2}}\sum_i\int\left[V_i\varepsilon_i\phi_w\left(\bar{x}+X_i\right)f\left(\bar{x}+X_i\right)-V_i\varepsilon_i\phi_w\left(X_i\right)f\left(X_i\right)\right]\frac{1}{a^q}K\left(\frac{\bar{x}}{a}\right)d\bar{x}\\&\qquad+o_p(1)\\&\quad=\frac{1}{n^{1/2}}\sum_i\int\left[g_w\left(V_i\varepsilon_i,\bar{x}+X_i\right)-g_w\left(V_i\varepsilon_i,X_i\right)\right]\frac{1}{a^q}K\left(\frac{\bar{x}}{a}\right)d\bar{x}+o_p(1)\\&\quad=o_p(1),\end{aligned}$$

applying Proposition 5, as in the proof of Proposition 3, taking into account that $\sup_{w,\bar{x}}\left|E\left(g_w\left(V_1\varepsilon_1,\bar{x}\right)\right)\right|=0$, and

$$\sup_w E\left[\left(\int\left[g_w\left(V_1\varepsilon_1,\bar{x}+X_1\right)-g_w\left(V_1\varepsilon_1,X_1\right)\right]\frac{1}{a^q}K\left(\frac{\bar{x}}{a}\right)\right)^2\right]=o(1),$$

reasoning as in (9). Thus, $\sup_w\left|A_{2n}^1(w)\right|=o_p(1)$. Write

$$\begin{aligned}A_{2n}^2(w)&=\frac{1}{n^{1/2}}\sum_i\frac{1}{f_i}V_i(m_i-\hat{m}_i)\left(f_i-\hat{f}_i\right)\hat{f}_i\hat{\phi}_{wi}+\frac{1}{n^{1/2}}\sum_i\frac{1}{f_i}V_i(m_i-\hat{m}_i)\hat{f}_i^2\hat{\phi}_{wi}\\&=A_{2n}^{21}(w)+A_{2n}^{22}(w).\end{aligned}$$

Applying the Cauchy-Schwarz and Markov inequalities,

$$\begin{aligned}\sup_w\left|A_{2n}^{21}(w)\right|&=O_p\left(\left\{nE\left[(\hat{m}_1-m_1)^2\hat{f}_1^2\right]E\left[\left(f_1-\hat{f}_1\right)^2\right]\right\}^{1/2}\right)\\&=O_p\left(n^{1/2}\left(\frac{1}{n^{1/2}a^{q/2}}+a^{\min(\lambda+1,\tau)}\right)\left(\frac{1}{n^{1/2}a^{q/2}}+a^{\lambda}\right)\right),\end{aligned}\tag{11}$$

since

$$E\left[(\hat{m}_1-m_1)^2\hat{f}_1^2\right]=O\left(\frac{1}{na^q}+a^{2\min(\lambda+1,\tau)}\right)$$

by Lemmas 3 and 5, and

$$E\left[\left(f_1-\hat{f}_1\right)^2\right]=O\left(\frac{1}{na^q}+a^{2\lambda}\right),$$

by Lemmas 2 and 4. Therefore, $\sup_w\left|A_{2n}^{21}(w)\right|=o_p(1)$ using A4′. Applying Lemma 3 and the LLN, we obtain that, uniformly in w,

$$A_{2n}^{22}(w)=\frac{1}{n^{5/2}}\sum_{i\neq j\neq k}V_i\frac{1}{f_i}(m_i-Y_j)K_{ij}K_{ik}\Pi_{wk}+O_p\left(\frac{1}{a^qn^{1/2}}\right)+O_p\left(\frac{1}{a^{2q}n^{3/2}}\right).$$

The first term on the right hand side of the last expression can be written, according to the notation at the beginning of this section, as

$$\frac{(n-2)(n-1)}{n^2}n^{1/2}U_n^{(3)}(\varphi_{wa}),$$

where

$$\varphi_{wa}\left(\chi_i^\dagger, \chi_j^\dagger, \chi_k^\dagger\right) = V_i\kappa_{ijk}\Pi_{wk} + V_k\kappa_{kji}\Pi_{wi} + V_j\kappa_{jik}\Pi_{wk} + V_k\kappa_{kij}\Pi_{wj} + V_i\kappa_{ikj}\Pi_{wj} + V_j\kappa_{jki}\Pi_{wi},$$

with $\chi_i^\dagger = (\chi_i, V_i)$, and

$$\kappa_{ijk} = \kappa_a\left(\chi_i, \chi_j, \chi_k\right) = \frac{1}{6f_i a^{2q}} K_{ji}K_{ik}\left(m_i - Y_j\right). \tag{12}$$

The class of functions $\{\varphi_{wa} : w \in \mathbb{R}^{q+p}\}$ is CV-type for each $a \in \mathbb{R}^+ \setminus \{0\}$, with CV-characteristics independent of a, and envelope proportional to

$$G_a\left(\chi_i, \chi_j, \chi_k\right) = \left|\kappa_{ijk}\right| + \left|\kappa_{kji}\right| + \left|\kappa_{jik}\right| + \left|\kappa_{kij}\right| + \left|\kappa_{ikj}\right| + \left|\kappa_{jki}\right|. \tag{13}$$

Thus,

$$\begin{aligned} P^3G_a^2 &\leq \frac{C}{a^{2q}}E\left\{(m_1 - Y_2)^2\frac{1}{a^q}K_{12}^2E\left[\frac{1}{a^q}K_{13}^2\middle| X_1\right]\right\} \\ &= O\left(\frac{1}{a^{2q}}\right), \end{aligned} \tag{14}$$

by Lemmas 2 and 3. Hence, applying Proposition 4,

$$E\left[\sup_w\left|n^{1/2}U_n^{(3)}\left(\pi_3\varphi_{wa}\right)\right|^2\right] \leq O\left(\frac{1}{n^2a^{2q}}\right),$$

$$E\left[\sup_w\left|n^{1/2}U_n^{(2)}\left(\pi_2\varphi_{wa}\right)\right|^2\right] \leq O\left(\frac{1}{na^{2q}}\right).$$

Thus, by Markov's inequality,

$$\begin{aligned} \sup_w\left|n^{1/2}U_n^{(3)}\left(\varphi_{wa}\right)\right| &= \sup_w\left|\sum_{k=1}^3\binom{3}{k}n^{1/2}U_n^{(k)}\left(\pi_k\varphi_{wa}\right)\right| \\ &= 3\sup_w\left|n^{1/2}U_n^{(1)}\left(\pi_1\varphi_{wa}\right)\right| + O_p\left(\frac{1}{na^q} + \frac{1}{n^{1/2}a^q}\right). \end{aligned}$$

Finally,

$$\begin{aligned} &3\sup_w\left|n^{1/2}U_n^{(1)}(\pi_1\varphi_{wa})\right| \\ &\quad= \sup_w\left|\frac{1}{n^{1/2}}\sum_i V_i\frac{1}{f_i}E\left\{\frac{1}{a^q}K_{ij}\left[(m_i - m_j) - \varepsilon_j\right]\frac{1}{a^q}K_{ik}\Pi_{wk}\middle|\chi_i, V_i, i\neq j\neq k\right\}\right| \\ &\quad= \sup_w\left|\frac{1}{n^{1/2}}\sum_i V_i\frac{1}{f_i}E\left\{E\left[\frac{1}{a^q}K_{ij}(m_i - m_j)\middle|X_i\right]\frac{1}{a^q}K_{ik}\phi_{wk}\middle|X_i, i\neq k\right\}\right| \end{aligned}$$

$$\le C\frac{1}{n^{1/2}}\sum_i\left|E\left\{\left|E\left[\frac{1}{a^q}K_{ij}(m_i-m_j)\middle|X_i\right]\right|\left|\frac{1}{a^q}K_{ik}\right|\middle|X_i,i\neq k\right\}\right|$$
$$=O\left(n^{1/2}a^{\min(\lambda+1,\tau)}\right)$$

by Lemmas 2 and 5 after applying Cauchy-Schwarz and Markov inequalities. Hence, we have shown that $\sup_w\left|A_{2n}^{22}(w)\right|=o_p(1)$. Therefore, we have proved that, $\sup_w\left|A_{2n}(w)\right|=o_p\left(n^{-1/2}\right)$. □

PROOF OF THEOREM 3. It suffices to show that $\sup_w|T_n^*(w)-\tilde{U}_n^*(w)|=o_p(n^{-1/2})$ and then apply Theorem 2. Define $\tilde{m}_i=(na^q)^{-1}\sum_j m_jK_{ij}/\hat{f}_i$ and $\tilde{\varepsilon}_i=(na^q)^{-1}\sum_j\varepsilon_jK_{ij}/\hat{f}_i$ and write, using that $\sum_i m_i\hat{f}_i\hat{\phi}_{wi}=\sum_i\tilde{m}_i\hat{f}_i\Pi_{wi}$,

$$\begin{aligned}
&n^{1/2}\left[T_n^*(w)-\tilde{U}_n^*(w)\right]\\
&\quad=\frac{1}{n^{1/2}}\sum_i(\hat{m}_i-m_i)\hat{f}_i\left[\Pi_{wi}-\hat{\phi}_{wi}\right]+\frac{1}{n^{1/2}}\sum_i m_i\hat{f}_i\left[\Pi_{wi}-\hat{\phi}_{wi}\right]\\
&\quad=\frac{1}{n^{1/2}}\sum_i(m_i-\tilde{m}_i)\hat{f}_i\phi_{wi}+\frac{1}{n^{1/2}}\sum_i\tilde{\varepsilon}_i\hat{f}_i[\Pi_{wi}-\phi_{wi}]\\
&\qquad+\frac{1}{n^{1/2}}\sum_i\frac{(\hat{m}_i-m_i)\hat{f}_i\left(f_i-\hat{f}_i\right)\left[\phi_{wi}-\hat{\phi}_{wi}\right]}{f_i}+\frac{1}{n^{1/2}}\sum_i\frac{\tilde{\varepsilon}_i\hat{f}_i^2\left[\phi_{wi}-\hat{\phi}_{wi}\right]}{f_i}\\
&\qquad+\frac{1}{n^{1/2}}\sum_i\frac{(\tilde{m}_i-m_i)\hat{f}_i^2\left[\phi_{wi}-\hat{\phi}_{wi}\right]}{f_i}\\
&\quad=B_{1n}(w)+B_{2n}(w)+B_{3n}(w)+B_{4n}(w)+B_{5n}(w).
\end{aligned}$$

So, it suffices to prove that, $\sup_w\left|B_{kn}(w)\right|=o_p(1)$, for $k=1,2,3,4,5$. First, $\sup_w\left|B_{1n}(w)\right|=o_p(1)$ mimicking the arguments in the proofs of Propositions 1, 2 and 3. Applying Lemma 3,

$$\sup_w\left|B_{2n}(w)\right|=\sup_w\left|\frac{1}{n^{3/2}}\sum_{i\neq j}\varepsilon_j\frac{1}{a^q}K_{ij}[\Pi_{wi}-\phi_{wi}]\right|+O_p\left(\frac{1}{na^q}\right).$$

The first term on the right hand side of the last expression is a completely degenerate U-process, which, applying Proposition 4 and noticing that $\{\Pi_w-\phi_w:w\in\mathbb{R}^{p+q}\}$ is VC-type, has second moments bounded by a constant independent of a times $(na^{2q})^{-1}E(\varepsilon_1^2K_{12}^2)=O((na^q)^{-1})$. Third, $\sup_w|B_{3n}(w)|=o_p(1)$

reasoning as in (11). Fourth, applying Lemma 3, uniformly in w,

$$\begin{aligned}B_{4n}(w) &= \frac{1}{n^{5/2}} \sum_{i\neq j\neq k} \frac{1}{f_i} \varepsilon_j \frac{1}{a^q} K_{ij} \frac{1}{a^q} K_{ik} (\phi_{wi} - \Pi_{wk}) \\ &\quad + O_p\left(\frac{1}{n^{1/2}a^q}\right) + O_p\left(\frac{1}{n^{3/2}a^{2q}}\right) \\ &= \frac{(n-1)(n-2)}{n^2} n^{1/2} U_n^{(3)}(\rho_{wa}) + O_p\left(\frac{1}{n^{1/2}a^q}\right) + O_p\left(\frac{1}{n^{3/2}a^{2q}}\right)\end{aligned}$$

where

$$\begin{aligned}\rho_{wa}(\chi_i, \chi_j, \chi_k) = \{&\psi_{ijk}(\phi_{wi} - \Pi_{wk}) + \psi_{kji}(\phi_{wk} - \Pi_{wi}) + \psi_{jik}(\phi_{wj} - \Pi_{wk}) \\ &+ \psi_{kij}(\phi_{wk} - \Pi_{wj}) + \psi_{ikj}(\phi_{wi} - \Pi_{wj}) + \psi_{jki}(\phi_{wj} - \Pi_{wi})\}\end{aligned}$$

and

$$\psi_{ijk} = \psi_a(\chi_i, \chi_j, \chi_k) = \frac{1}{6} \frac{1}{f_i} \varepsilon_j \frac{1}{a^q} K_{ij} \frac{1}{a^q} K_{ik}.$$

The class of functions $\{\rho_{wa} : w \in \mathbb{R}^{q+p}\}$ is CV-type, for each , $a \in \mathbb{R}^+ \setminus \{0\}$, with CV-characteristics independent of a, and envelope

$$L_a(\chi_i, \chi_j, \chi_k) = \sum_{(6)} |\psi_a(\chi_i, \chi_j, \chi_k)|,$$

where $\sum_{(6)}$ runs over all possible permutations of the integers (i, j, k). By Lemmas 2 and 3,

$$P^3 L_a^2 \leq CE\left[\left|\varepsilon_2 \frac{1}{a^q} K_{12}\right|^2 E\left(\left|\frac{1}{a^{2q}} K_{13}^2\right| \Big| X_1\right)\right] = O\left(\frac{1}{a^{2q}}\right).$$

Thus, taking into account that $\sup_w |P^3 \rho_{wa}| = 0$ and applying Proposition 5,

$$\begin{aligned}\sup_w \left|n^{1/2} U_n^{(3)}(\rho_{wa})\right| &= \sup_w \left|\sum_{k=1}^{3} \binom{3}{k} n^{1/2} U_n^{(k)}(\pi_k \rho_{wa})\right| \\ &\leq 3 \sup_w \left|n^{1/2} U_n^{(1)}(\pi_1 \rho_{wa})\right| + O_p\left(\frac{1}{na^q} + \frac{1}{n^{1/2}a^q}\right).\end{aligned} \tag{15}$$

Because of notational convenience, let us define

$$E_i(\gamma(\chi_i, \chi_j)) = E(\gamma(\chi_i, \chi_j) | \chi_i, i \neq j),$$

where γ is a generic function. Now,

$$\begin{aligned}3n^{1/2} U_n^{(1)}(\pi_1 \rho_{wa}) &= \frac{1}{n^{1/2}} \sum_i \varepsilon_i E_i\left(\frac{1}{f_j} \frac{1}{a^q} K_{ij} \frac{1}{a^q} K_{jk} (\phi_{wj} - \phi_{wk})\right) \\ &= \frac{1}{n^{1/2}} \sum_i \varepsilon_i \left(\phi_{wi} f_i - E_i\left(\frac{1}{f_j} \frac{1}{a^q} K_{ij} \frac{1}{a^q} K_{jk} \phi_{wk}\right)\right)\end{aligned}$$

$$+\frac{1}{n^{1/2}}\sum_i \varepsilon_i\left[E_i\left(\frac{1}{a^q}K_{ij}\phi_{wj}\right)-\phi_{wi}f_i\right]$$
$$+\frac{1}{n^{1/2}}\sum_i \varepsilon_i E_i\left(\frac{1}{f_j}\frac{1}{a^q}K_{ij}\phi_{wj}E_j\left(\frac{1}{a^q}K_{jk}-f_j\right)\right)$$
$$=D_{1n}(w)+D_{2n}(w)+D_{3n}(w).$$

Next we show that $\sup_w \left|D_{jn}(w)\right| = o_p(1)$, $j=1,2,3$. First,

$$\begin{aligned}
&\sup_w\left|D_{1n}(w)\right|\\
&=\sup_w\left|\frac{1}{n^{1/2}}\sum_i\varepsilon_i\left\{\phi_w(X_i)f(X_i)\right.\right.\\
&\qquad\left.\left.-\int\int\frac{1}{a^{2q}}K\left(\frac{\bar{x}-\tilde{x}}{a}\right)K\left(\frac{X_i-\bar{x}}{a}\right)\phi_w(\tilde{x})f(\tilde{x})d\tilde{x}d\bar{x}\right\}\right|\\
&=\sup_w\left|\frac{1}{n^{1/2}}\sum_i\left\{\int[\varepsilon_i\phi_w(X_i)f(X_i)-\varepsilon_i\phi_w(X_i+y)f(X_i+y)]\frac{1}{a^q}\mathbb{K}\left(\frac{y}{a}\right)dy\right\}\right|\\
&=\sup_w\frac{1}{n^{1/2}}\left|\sum_i\int[g_w(\varepsilon_i,X_i)-g_w(\varepsilon_i,X_i+\bar{x})]\frac{1}{a^q}\mathbb{K}\left(\frac{y}{a}\right)dy\right|\\
&=\sup_w\left|\frac{1}{n^{1/2}}\sum_i\int[g_w(\varepsilon_i,X_i)-g_w(\varepsilon_i+\bar{u},X_i+\bar{x})]d\mu_n(\bar{u},\bar{x})\right|,
\end{aligned}$$

where $g_w(\bar{\varepsilon},\bar{x})=\bar{\varepsilon}\phi_w(\bar{x})f(\bar{x})$, $\mathbb{K}(y)=\int K(u)K(u-y)\,du$, and

$$\mu_n(B_1\times B_2)=1(0\in B_1)\int_{y\in B_2}\frac{1}{a^q}\mathbb{K}\left(\frac{y}{a}\right)dy,$$

is a signed measure, degenerate in the first coordinate, which converges to the Dirac measure at zero. Hence, applying Proposition 5, it suffices to show that

$$\sup_w E\left[\left|\int[g_w(\varepsilon_1,X_1)-g_w(\varepsilon_1,X_1+\bar{x})]\frac{1}{a^q}\mathbb{K}\left(\frac{y}{a}\right)dy\right|^2\right]=o(1),$$
$$\sup_w\left[E\left|\int[g_w(\varepsilon_1,X_1)-g_w(\varepsilon_1,X_1+\bar{x})]\frac{1}{a^q}\mathbb{K}\left(\frac{y}{a}\right)dy\right|\right]=o\left(n^{-1/2}\right),$$

which follows reasoning as in the proofs of (5) and (6), noticing that $\mathbb{K}$, like K, is a multiplicative kernel, which satisfies A3. Second, $\sup_w\left|D_{2n}(w)\right|=o_p(1)$ applying Proposition 5 in the same way. Third, notice that

$$D_{3n}(w)=\frac{1}{n^{1/2}}\sum_i\bar{t}_{wa}(Y_i,X_i),$$

where $\bar{t}_{wa}(y,x) := \int\int t_{wa}(y,x,\bar{x},\tilde{x})\,d\bar{x}d\tilde{x}$, and

$$t_{wa}(y,x,\bar{x},\tilde{x}) = (y - m(x))\frac{1}{a^q}K\left(\frac{x-\bar{x}}{a}\right)\phi_w(\bar{x})\left[\frac{1}{a^q}K\left(\frac{\bar{x}-\tilde{x}}{a}\right) - f(\bar{x})\right]f(\tilde{x}).$$

Since $\{t_{wa} : w \in \mathbb{R}^{p+q}\}$ is VC-type with VC-characteristics independent of a and envelope

$$M_a(y,x,\bar{x},\tilde{x}) = |y - m(x)|\frac{1}{a^q}\left|K\left(\frac{x-\bar{x}}{a}\right)\right|\left|\frac{1}{a^q}K\left(\frac{\bar{x}-\tilde{x}}{a}\right) - f(\bar{x})\right|f(\tilde{x}),$$

$\{\bar{t}_{wa} : w \in \mathbb{R}^{p+q}\}$ is also VC with VC-characteristics independent of a, applying Lemma A2 in Ghosal, Sen and van der Vaart (2000) [see also Lemma 5 and 6 in Sherman (1994)] and it has envelope $\bar{M}_a(y,x) = \int\int M_a(y,x,\bar{x},\tilde{x})\,d\bar{x}d\tilde{x}$. Therefore, by Proposition 4 applied with $m = 1$, there exists a constant C independent of a such that,

$$E\left[\sup_w\left|\frac{1}{n^{1/2}}\sum_i \bar{t}_{wa}(Y_i, X_i)\right|^2\right] \le CP\bar{M}_a^2 = O\left(a^{2\lambda}\right),$$

by Lemmas 2 and 4, after applying Hölder's inequality. Therefore $\sup_w |D_{3n}(w)| = O\left(a^{\lambda}\right)$ by Markov's inequality. Finally, applying Lemma 3,

$$\begin{aligned} B_{5n}(w) &= \frac{1}{n^{5/2}}\sum_{i\neq j\neq k}\frac{1}{f_i}(m_i - m_j)\frac{1}{a^q}K_{ij}\frac{1}{a^q}K_{jk}\Pi_{wi}\Pi_{wk} \\ &\quad + O_p\left(\frac{1}{n^{1/2}a^q}\right) + O_p\left(\frac{1}{n^{3/2}a^{2q}}\right) \\ &= \frac{(n-1)(n-2)}{n^2}n^{1/2}U_n^{(3)}(\alpha_{wa}) + O_p\left(\frac{1}{n^{1/2}a^q}\right) + O_p\left(\frac{1}{n^{3/2}a^{2q}}\right), \end{aligned}$$

where

$$\begin{aligned} \alpha_{wa}(\chi_i,\chi_j,\chi_k) = \{&\varkappa_{ijk}(\phi_{wi} - \Pi_{wk}) + \varkappa_{kji}(\phi_{wk} - \Pi_{wi}) + \varkappa_{jik}(\phi_{wj} - \Pi_{wk}) \\ &+ \varkappa_{kij}(\phi_{wk} - \Pi_{wj}) + \varkappa_{ikj}(\phi_{wi} - \Pi_{wj}) + \varkappa_{jki}(\phi_{wj} - \Pi_{wi})\} \end{aligned}$$

and

$$\varkappa_{ijk} = \varkappa_a(\chi_i,\chi_j,\chi_k) = \frac{1}{f_i}(m_i - m_j)\frac{1}{a^q}K_{ij}\frac{1}{a^q}K_{ik}.$$

Notice that α_{wa} is identical to ρ_{wa}, after substituting ψ_a by $\varkappa_a$. Therefore, $\{\alpha_{wa} : w \in \mathbb{R}^{q+p}\}$ is CV-type, for each $a \in \mathbb{R}^+ \setminus \{0\}$, with CV-characteristics independent of a and envelope with second moments $O\left(n^{-2q}\right)$. Therefore, reasoning as in (15),

$$\sup_w\left|n^{1/2}U_n^{(3)}(\alpha_{wa})\right| = \sup_w\left|3n^{1/2}U_n^{(1)}(\pi_1\alpha_{wa}) + n^{1/2}P^3\alpha_{wa}\right| + O_p\left(\frac{1}{na^q} + \frac{1}{n^{1/2}a^q}\right).$$

Now,

$$
\begin{aligned}
n^{1/2}\sup_w\left|P^3\alpha_{wa}\right| &= n^{1/2}\sup_w\left|E\left[\frac{1}{f_1}(m_1-m_2)\frac{1}{a^q}K_{12}\frac{1}{a^q}K_{13}\right]\right|\\
&= n^{1/2}\sup_w\left|E\left[\frac{1}{f_1}E_1\left((m_1-m_2)\frac{1}{a^q}K_{12}\right)\frac{1}{a^q}K_{13}\right]\right|\\
&= O\left(n^{1/2}a^{\min(\lambda+1,\tau)}\right),
\end{aligned}
$$

by Lemma 5. Applying Lemma A2 in Ghosal, Sen and van der Vaart (2000), $\left\{P^2\alpha_{wa} : w\in\mathbb{R}^{p+q}\right\}$ is also VC-type for each $a\in\mathbb{R}^+\setminus\{0\}$, with VC-characteristics independent of a, and envelope

$$
\begin{aligned}
N_a(X_i) = E_i\Bigg[&\left|m_j-m_i\right|\frac{1}{a^q}\left|K_{ji}\right|\frac{1}{a^q}E_j\left(\left|K_{jk}\right|\right)+\left|m_k-m_i\right|\frac{1}{a^q}\left|K_{ki}\right|\frac{1}{a^q}E_k\left(\left|K_{kj}\right|\right)\\
&+E_j\left(\left|m_j-m_k\right|\frac{1}{a^q}\left|K_{jk}\right|\right)\frac{1}{a^q}\left|K_{ji}\right|\Bigg].
\end{aligned}
$$

Therefore, by Proposition 4, applied with $m=1$,

$$
\begin{aligned}
E\left[\sup_w\left|n^{1/2}U_n^{(1)}(\pi_1\alpha_{wa})\right|\right] &\le CPN_a^2\\
&= O\left(a^2\right)
\end{aligned}
$$

by Lemmas 2 and 6, which concludes the proof. □

PROOF OF COROLLARY 2. Write $\{\tilde{U}_n^{0*}(w) : w\in\mathbb{R}^{q+p}\} = \{P_n^* g : g\in\mathscr{H}\}$ as a process indexed by functions in $\mathscr{H}$. Since $\mathscr{H}$ is P-Donsker with squared integrable envelope, and V_i are bounded, the conditional multiplier uniform central limit theorem [see Ledoux and Talagrand (1988) and Theorem 2.9.7 in Van der Vaart and Wellner (1996); see also Problem 5, page 186] establishes that, $n^{1/2}P_n^*$ converges in distribution to B_P in $\ell^\infty(\mathscr{H})$ almost surely, where B_P is sample continuous, and for a continuous funtional $\varphi : \ell^\infty(\mathscr{H})\mapsto\mathbb{R}$,

$$
d\left(F^*_{\varphi\left(n^{1/2}\tilde{U}_n^{0*}\right)}, F_{\varphi(B_P)}\right) = o(1) \qquad \text{a.s.},
$$

where d is a distance metrizing weak convergence on the real line, and F^* is the conditional distribution given the sample $\mathscr{Y}_n$. Therefore, applying Theorems 2 and 3,

$$
d\left(F^*_{\varphi\left(n^{1/2}\tilde{U}_n^{*}\right)}, F_{\varphi(B_P)}\right) = o_p(1) \quad \text{and} \quad d\left(F^*_{\varphi\left(n^{1/2}T_n^{*}\right)}, F_{\varphi(B_P)}\right) = o_p(1).
$$

Then, the Corollary follows reasoning as in the proof of Corollary 1. □

7. Lemmas. Proofs of Lemmas 1 to 5 can be found in Robinson (1988).

LEMMA 1. *Let* $\sup_u |k(u)| + \int |u^\gamma k(u)| du < \infty$, *for some* $\gamma \geq 0$. *Then, uniformly in* x,

$$\int_{\mathbb{R}^q} \|x - s\|^\gamma \left| K\left(\frac{x - s}{a}\right)\right| ds \leq Ca^{q+\gamma}. \tag{16}$$

LEMMA 2. *Let* $\sup_x f(x) < \infty$, $\sup_u |k(u)| + \int |k(u)| du < \infty$. *Then uniformly in* x,

$$E\left| K\left(\frac{X - x}{a}\right)\right| \leq Ca^q.$$

LEMMA 3. *Let* $\sup_x f(x) < \infty$, $E\left(|s(X)|\right) < \infty$, $\sup_u |k(u)| + \int |k(u)| du < \infty$. *Then,*

$$E\left[s(X_1)\left| K\left(\frac{X_1 - X_2}{a}\right)\right|\right] \leq Ca^q.$$

LEMMA 4. *Suppose* λ *satisfies* $l - 1 < \lambda \leq l$, *where* $l \geq 1$ *is an integer, and let* $f \in \mathscr{J}_\lambda^\infty$, $k \in \mathscr{K}_l$. *Then, for* $\alpha > 0$,

$$E\left[\left| E\left[\frac{1}{a^q} K\left(\frac{X_1 - X_2}{a}\right)\right| X_1\right] - f(X_1)\right|^\alpha\right] = O\left(a^{\alpha\lambda}\right).$$

LEMMA 5. *Suppose* λ, τ *satisfying* $l - 1 < \lambda \leq l$, $t - 1 < \tau \leq t$, *where* $l \geq 1$, $t \geq 1$ *are integers, let* $f \in \mathscr{J}_\lambda^\infty$ *and* $m \in \mathscr{J}_\tau^\alpha$ *for some* $\alpha > 0$, $k \in \mathscr{K}_{l+t-1}$. *Then*

$$E\left\{\left| E\left([m(X_1) - m(X_2)] K\left(\frac{X_1 - X_2}{a}\right)\right| X_1\right)\right|^\alpha\right\} = O\left(a^{\alpha[q + \min(\lambda+1, \tau)]}\right).$$

LEMMA 6. *For* $t - 1 < \tau \leq t$, $t \geq 1$, $\sup_x f(x) < \infty$, *let* $s \in \mathscr{J}_\tau^\alpha$, $\alpha > 0$, $\sup_u |k(u)| + \int |k(u)| du < \infty$. *Then*

$$E\left\{\left| E\left(|s(X_1) - s(X_2)| \left| K\left(\frac{X_1 - X_2}{a}\right)\right|\right| X_1\right)\right|^\alpha\right\} = O\left(a^{\alpha[q + \min(1, \tau)]}\right). \tag{17}$$

PROOF. Let $Q(v, x)$ be a $(t-1)-th$ homogeneous polynomial in $v - x$ with coefficients the partial derivatives of s at x of orders 1 through $t - 1$, when $t > 1$; and $Q = 0$ when $t = 1$. Let d_1 be a function like d in Definition 1, and let d_2 depend on the derivatives of s. Let $L(\gamma)$ be the left hand side of (16). As in Robinson's (1988) proof of Lemma 5, the left hand side of (17) is bounded by

$$\int_{S_{x\rho}} |s(v) - s(x) - Q(v, x)| \left| K\left(\frac{v - x}{a}\right)\right| f(v) dv$$
$$+ \int_{S_{x\rho}} |Q(v, x)| \left| K\left(\frac{v - x}{a}\right)\right| f(v) dv$$

$$+\int_{\bar{S}_{x\rho}} \left|s(v) - s(x)\right| \left|K\left(\frac{v-x}{a}\right)\right| f(v)\,dv$$

$$\leq C\left\{d_1(x)\,L(\tau) + d_2(x)\sum_{i=1}^{t-1} L(i)\right.$$

$$\left. + \left[\left|s(x)\right| + E\left(\left|s(X)\right|\right)\right] a^{q+1} \sup_u \left\{|u|^{q+1}\left|k(u)\right|^q\right\}\right\},$$

where $E\left[\left|d_1(X)\right|^\alpha + \left|d_2(X)\right|^\alpha\right] < \infty$. Then, the lemma follows applying Lemma 1 and dominated convergence. □

Acknowledgments. We thank the associate editor and two anonymous referees for suggestions that have greatly improved the paper. We are most grateful to Evarist Giné and Winfried Stute for valuable and insightful comments on the second version of this manuscript.

REFERENCES

BERAN, R., LE CAM, L. and MILLAR, P.W. (1987). Convergence of stochastic empirical measures. *J. Multivariate Anal.* **23** 159–168.

BIERENS, H. and PLOBERGER, W. (1997). Asymptotic theory of integrated conditional moment test. *Econometrica* **65** 1153-1174.

BRUNK, H.D. (1970). Estimation by isotonic regression. In *Nonparametric Techniques in Statistical Inference* (M.L. Puri ed.) 177–197. Cambridge Univ. Press.

DUDDLEY, R.M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press.

DE LA PEÑA, V.H. and GINÉ, E. (1999). *Decoupling: From Dependence to Independence.* Springer, Berlin.

EUBANK, R. and SPIEGELMAN, S. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques, *J. Amer. Statist. Assoc.* **85** 387–392.

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

GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082.

GINÉ, E. (1997). *Lectures on Some Aspects of the Bootstrap. Ecole de Éte de Calcul de Probabilités de Saint-Flour. Lecture Notes in Math.* **1665.** Springer, Berlin. (See also www.math.uconn.edu/~gine/Corrections.)

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

HART, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, Berlin.

HECKMAN, N.E. (1986). Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B* **48** 244–248.

HOFFMAN-JØRGENSEN, J. (1984). Stochastic processes on Polish spaces. [Published (1991). Various Publication Series No. 39. Matematisk Institute, Aarhus Univ.]

HONG-ZHY, A. and BING, C. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.* **59** 287–307.

KOUL, H.L. and STUTE, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27** 204–236.

LEDOUX, M. and TALAGRAND, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probab. Theory Related Fields* **77** 29–47.

NOLAN, D. and POLLARD, D. (1987). *U*-processes: rates of convergence. *Ann. Statist.* **15** 780–799.

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

ROSENBLATT, M. (1975). A quadratic measure of deviations of two-dimensional density estimates an a test of independence. *Ann. Statist.* **3** 1–14.

SHERMAN, R.P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22** 439–459.

SPECKMAN, P. (1988). Kernel smoothing in partially linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–446.

STUTE, W. (1994). U-Statistic processes: a martingale approach. *Ann. Probab.* **22** 1725–1744.

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.

STUTE, W., GONZÁLEZ-MANTEIGA, W. and PRESEDO-QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149.

STUTE, W., THIES, S. and ZHU, L.X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26** 1916–1934.

SUE, J.Q. and WEI, L.J. (1991). A lack of fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86** 420–426.

VAN DER VAART, A.W. (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **21** 501–504.

VAN DER VAART, A.W. and WELLNER, J.A. (1996). *Weak Convergence of Empirical Processes.* Springer, New York.

WU, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350.

DEPARTAMENTO DE ESTADÍSTICA
Y ECONOMETRÍA
UNIVERSIDAD CARLOS III DE MADRID
C/ MADRID 126-128
28903 GETAFE
SPAIN
E-MAIL: delgado@est-econ.uc3m.es

DEPARTAMENTO DE ESTADÍSTICA
E INVESTIGACIÓN OPERATIVA
FACULTAD DE MATEMÁTICAS
UNIVERSIDAD DE SANTIAGO DE COMPOSTELA
15782 SANTIAGO DE COMPOSTELA
SPAIN
E-MAIL: wences@zmat.usc.es