Working Paper 98-68
Statistics and Econometrics Series 31
October 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34)91 624-9849

SIGNIFICANCE TESTING IN NONPARAMETRIC REGRESSION BASED ON THE BOOTSTRAP*

Miguel A. Delgado, and Wenceslao González Manteiga.*

Abstract

We propose 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 an empirical process marked by nonparametric residuals. 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 bootstrap tests easy to implement, which exhibit good level accuracy for fairly small samples, according to the Monte Carlo simulations reported. These results are also applicable to test other interesting restrictions on nonparametric regression curves, like partial linearity and conditional independence.

Keywords:
Nonparametric regression; selection of variables; higher order kernels; marked empirical processes; Wild bootstrap; restrictions on nonparametric curves.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Ph: 34-91-624.98.04; Fax: 34-91-624.98.49, e-mail: delgado@est-econ.uc3m.es; González Manteiga, Universidad de Santiago de Compostela.
Research funded by Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB95-0292 and PB95-0826.

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. We propose resampling procedures, in order to estimate the critical values of the test, 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, like conditional independence.

There is a large literature on consistent specification testing, consistent in the direction of general alternatives, based on two leading methodologies. On one hand, there have been proposed tests based on some distance between the fitted nonparametric regression, using some smoother, and the parametric fit under the null hypothesis, see e.g. Eubank and Spiegelman (1990), Härdle and Mammen (1993), Horowitz and Härdle (1994), Gozalo (1993), Hong and White (1995) and Zheng (1996) among others. 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 e.g. Brunk (1970), Hong-zhy and Bin (1991), Sue and Wei (1991), Stute (1997), Andrews (1997) and Delgado et al (1998) among others. 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 and, unlike the tests based on smoothers, asymptotic critical values are difficult to compute. Related to this method are Bierens' tests (see e.g. Bierens (1982, 1990) and 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)). The second type of tests resembles the tipical 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 (see e.g. Kolmogorov (1933), Cràmer (1928), Smirnov (1936) and v. Mises (1931)).

The two methodologies discussed above, which have been developed for specification testing of parametric regresion functions, are applicable to test different restrictions on nonparametric regression curves. Significance testing is a relevant example of restrictions to be tested, since the "curse of dimensionality" motivates to reduce the number of explanatory variables in the regression curve as much as possible. Given a random vector (Y, W), where Y is scalar and $W = (X', Z')'$, $X = \left(X^{(1)}, X^{(2)}, ..., X^{(q)}\right)'$ and $Z = \left(Z^{(1)}, Z^{(2)}, ..., Z^{(p)}\right)'$ are $\mathbb{R}^q - valued$ and $\mathbb{R}^p - valued$ respectively, we want to test

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

where $m(\cdot) = E(Y|X = \cdot)$. The alternative hypothesis, H_1, is the negation of H_0. Fan and Li (1996) have proposed a significance test inspired in the first methodology discussed above. That is, the null hypothesis can alternatively be written as,

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

where ς is a suitable weight function which does not change 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 ς involves the density function of X and W, in order to avoid the stochastic denominators in the resulting statistics. So, this testing procedure requires to estimate two nonparametric regression curves with q and $p+q$ regressors respectively, and to choose two different bandwidth numbers for each regression, one converging to zero faster than the other. The resulting test statistic has the form of a degenerate $U - statistic$ converging to a standard normal under H_0. In this paper, we propose to apply the second methodology, which only requires to estimate $E(Y \mid X)$ using smoothers. Hereforth, for two vectors v and w of equal dimension, "$v \leq w$" means that each coordinate of v is less or equal to the corresponding coordinate of w and $1(A)$ is the indicator function of the event A. Notice that,

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

$$\Leftrightarrow H_0 : E\{[Y - m(X)] 1(W \leq w)\} = 0, \ \forall w = (x', z')' \in \mathcal{W}, \tag{1}$$

where $\mathcal{W}$ is the support of W. The expectation in (1) can also be written as,

$$\int_{-\infty}^{w} E(Y \mid X = x, Z = z) \, dF_W(x, z) - \int_{-\infty}^{w} m(x) \, dF_W(x, z), \tag{2}$$

where F_W is the distribution function of W and, hereforth, for a vector v and some function g, $\int_{-\infty}^{v} g(u) \, du = \int_{-\infty}^{\infty} 1(u \leq v) g(u) \, du$. Hence, (2) is the difference between the integrated regression funcion of Y given W and the integrated regression function of Y given X. Only under H_0, this difference will be equal to zero for all $w \in \mathcal{W}$. Let $f(\cdot)$ be the marginal density function of X. Since $f(x) \geq 0 \; \forall x$, the hypothesis H_0 in (1) can also be written as

$$H_0 : E\{f(X)[Y - m(X)] 1(W \leq w)\} = 0, \; \forall w \in \mathcal{W}. \tag{3}$$

The reason of writing H_0 in this form is purely technical, in order to avoid the random denominator in the conditional expectation. The same feature has been used by Powell et al (1989), Robinson (1989), Zheng (1996) and Fan and Li (1996) among others.

We propose tests statistics based on functions of estimates of the expectation in (3). In next section we present the test statistics, and we show that the resulting asymptotic tests has non trivial power under contiguous alternatives converging to the null at the parametric rate $n^{-1/2}$. However, asymptotic tests cannot be implemented except in exceptional circumstances, since the asymptotic distribution of the statistic under the null depends on unknown features of the underlying distribution function of (Y, W). In section 3, we propose consistent bootstrap tests, easy to implement. A Monte Carlo study, in section 4, illustrates the properties of the proposed bootstrap tests in practice. Finally, in section 5, we propose the extension of this testing methodology to test other restrictions on nonparametric curves, discussing in detail a test for partial linearity and a test for conditional independence.

2. NONPARAMETRIC SIGNIFICANCE TESTING.

Given independent observations $\{(Y_i, W_i), i = 1, ..., n\}$ of (Y, W), where $W_i = (X_i, Z_i)$, the expectation in (3) can be consistently estimated, when $U_i = Y_i - m(X_i)$ and $f(X_i)$ are known, by

$$T_n(w) = \frac{1}{n} \sum_{i=1}^{n} f(X_i) U_i 1(W_i \leq w), \tag{4}$$

which is a marked empirical process, with marks $f(X_i)(Y_i - m(X_i))$.

Applying a Central Limit Theorem argument, under H_0, $\sqrt{n}T_n$ has a normal limiting finite dimensional distribution, with covariance structure,

$$\begin{aligned} \Omega(w_1, w_2) &= Cov(T_n(w_1), T_n(w_2)) \\ &= E\left\{f(X)^2 \sigma^2(W) 1(W \leq \min(w_1, w_2))\right\}, \end{aligned}$$

where $w_j \in \mathbb{R}^{p+q}$, $j = 1, 2$ and $\sigma^2(\cdot) = Var(Y \mid W = \cdot)$. The tighness of the process follows using similar arguments as Stute (1997). Then,

$$\sqrt{n}T_n(w) \text{ converges weakly to } T_\infty^0(w) \text{ on } D\left(\mathbb{R}^{q+p}\right),$$

where T_{∞}^{0} is a Gaussian process centered at zero and with covariance structure Ω.

Since $m(X_i), i = 1, ..., n$ are unknown, they are estimated by the kernel regression estimator,

$$\hat{m}(X_i) = \frac{1}{\hat{f}(X_i)} \frac{1}{nh^q} \sum_{\substack{j=1 \\ j \neq i}}^{n} K_{ij} Y_j,$$

where

$$\hat{f}(X_i) = \frac{1}{nh^q} \sum_{\substack{j=1 \\ j \neq i}}^{n} K_{ij}$$

is the estimator of the density function of X evaluated at X_i, $f(X_i)$, and

$$K_{ij} = K\left(\frac{X_i - X_j}{h}\right),$$

where $K(u) = \prod_{j=1}^{q} k(u_j)$, k is an univariate kernel and h is a bandwidth number. So a feasible version of T_n is given by,

$$\begin{aligned} \hat{T}_n(w) &= \frac{1}{n} \sum_{i=1}^{n} \hat{f}(X_i) \hat{U}_i 1(W_i \leq w) \qquad (5) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{h^q} K_{ij} (Y_i - Y_j) 1(W_i \leq w). \end{aligned}$$

The test statistic is based on some continuous functional of $\sqrt{n}\hat{T}_n$. For instance, we can use a Cràmer-v.Mises statistic of the form

$$\hat{C}_n = n \int_{\mathbb{R}^{p+q}} \hat{T}_n(w)^2 \, d\hat{F}_{Wn}(w) = \sum_{i=1}^{n} \hat{T}_n(W_i)^2,$$

where $\hat{F}_{Wn}$ is the empirical distribution function of W. Kolmogorov-Smirnov statistics can also be constructed in a similar way.

Next, we provide first order asymptotic expansions of $\hat{T}_n$, which are very useful both, for deriving the asymptotic distribution of the tests statistics under the null and for motivating the bootstrap tests in next section. We need the following definitions introduced by Robinson (1988).

Definition 1 *$\mathcal{K}_\ell$, $\ell \geq 2$ is the class of even functions, $k : \mathbb{R}^q \rightarrow \mathbb{R}$ satisfying*

$$\int_{\mathbb{R}} u^i k(u) \, du = \begin{cases} 1 \text{ if } i = 0 \\ 0 \text{ if } i = 1, ..., \ell - 1, \end{cases}$$

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

Definition 2 *$\mathcal{G}_{\beta}^{\alpha}$, $\alpha > 0$, $\beta > 0$, is the class of functions $g : \mathbb{R} \to \mathbb{R}$ satisfying: g is $(b-1)$-times continously differentiable, for $b-1 \leq \beta \leq b$ and all u; $\sup_{v \in \mathcal{S}_{u\rho}} |g(v) - g(u) - Q(v,u)| / \|v-u\|^{\beta} \leq h(u)$, for some $\rho > 0$, where $\mathcal{S}_{u\rho} = \{v : \|v-u\| < \rho\}$; $Q = 0$ when $b = 1$; Q is a $(b-1)$th-degree homogeneous polynomial in $v-u$ with coefficients 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 $h(u)$ have finite αth moments.*

We also need the following assumptions.

A1.- $f \in \mathcal{G}_{\lambda}^{\infty}$, for some $\lambda > 0$.

A2.- $m \in \mathcal{G}_{\mu}^{2}$, for some $\mu > 0$.

A3.- Uniformly in z, $r(\cdot, z) \equiv E[1(Z \leq z) \mid X = \cdot] \in \mathcal{G}_{\upsilon}^{\infty}$, for some $\upsilon > 0$.

A4.- $E[U_1^2] < \infty$.

A5.- $k \in \mathcal{K}_{\ell+m-1}$, where $\ell - 1 < \lambda \leq \ell$ and $m - 1 < \mu \leq m$.

A6.- $(nh^q)^{-1} + nh^{2\eta} \to 0$ as $n \to \infty$, where $\eta = \min(\mu, \lambda + 1)$.

Assumptions A1, A2, A5 and A6 are needed for bias reduction using "higher order kernels" in Definition 1, as suggested by Robinson (1988). A neccesary condition, reconciling the components of A6, is $\mu > q/2$, $\lambda > q/2 - 1$. Condition A3 is not very restrictive, since υ is not related to the conditions A5 and A6. Let us define $\phi(X_i, w) = 1(X_i \leq x)\, r(X_i, z)$ and $\hat{\phi}(X_i, w) = (nh^q)^{-1} \sum_{j=1, j \neq i}^{n} 1(W_j \leq w) K_{ij} / \hat{f}(X_i)$. The next Theorem provides first order expansions for $\hat{T}_n$.

Theorem 1 *Under A1 to A6, uniformly in w,*

$$\hat{T}_n(w) = \frac{1}{n}\sum_{i=1}^{n} U_i \hat{f}(X_i)\left[1(W_i \leq w) - \hat{\phi}(X_i, w)\right] + o_p\left(n^{-1/2}\right) \tag{6}$$

$$= \frac{1}{n}\sum_{i=1}^{n} U_i f(X_i)\left[1(W_i \leq w) - \phi(X_i, w)\right] + o_p\left(n^{-1/2}\right). \tag{7}$$

Notice that, according to Theorem 1, uniformly in w

$$\hat{T}_n(w) = T_n(w) - \frac{1}{n}\sum_{i=1}^{n} U_i f(X_i)\,\phi(X_i, w) + o_p\left(n^{-1/2}\right). \tag{8}$$

Then, applying a central theorem argument, the finite dimensional distribution of $\sqrt{n}\hat{T}_n$ is Gaussian with covariance structure,

$$\Theta(w_1, w_2) = E\left\{\sigma^2(W)\, f(X)^2 \left[1(W \leq w_1) - \phi(X, w_1)\right]\left[1(W \leq w_2) - \phi(X, w_2)\right]\right\}.$$

Thus, $\sqrt{n}\hat{T}_n$ and the first term in (7), multiplied by $\sqrt{n}$, converges weakly to the same limiting process under H_0. The following Theorem provides an invariance principle for $\sqrt{n}\hat{T}_n$.

Theorem 2 *Under H_0, A1 to A6,*

$$\sqrt{n}\hat{T}_n(w) \text{ converges weakly to } T_{\infty}^{1}(w) \text{ in } D\left(\mathbb{R}^{q+p}\right),$$

where T_{∞}^{1} is a Gaussian process centered at cero and with covariance structure Θ.

The following Corollary stablishes the asymptotic distribution of the statistic under contiguous alternatives of the form,

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

where $\Pr(|g(W)| > 0) = 1$. Define the process,

$$T_{\infty}^{2}(w) = T_{\infty}^{1}(w) + S(w),$$

where

$$S(w) = E\{g(W)[1(W \leq w) - \phi(X, w)]\}.$$

Corollary 1 *Assume A1 to A6. and* $\Pr(|g(W)| > 0) = 1$, *with* $E|g(W)| < \infty$. *Under* H_0,

$$\hat{C}_n \xrightarrow{d} C_{\infty} \equiv \int_{\Re^{q+p}} T_{\infty}^{1}(w)^2 \, dF_W(w), \tag{9}$$

under H_{1n},

$$\hat{C}_n \xrightarrow{d} \int_{\Re^{q+p}} T_{\infty}^{2}(w)^2 \, dF_W(w), \tag{10}$$

and under H_1,

$$\hat{C}_n \xrightarrow{p} \infty. \tag{11}$$

The process T_{∞}^{1} depends on certain features of the distribution of (Y, W) and an asymptotic test cannot be implemented except in exceptional circunstances. This is why we propose a bootstrap test in next section, in order to estimate the critical values of the statistic.

3. BOOTSTRAP TESTS

We propose to estimate the exact critical values of the test statistic, $\hat{C}_n$, by the quantiles of the conditional distribution, given the sample $\mathcal{Y}_n = \{(Y_i, W_i), i = 1, ..., n\}$, of bootstrap statistics. We suggest two alternative bootstrap tests, both based on resamples $\left\{\hat{U}_i^*, i = 1, ..., n\right\}$ from the non-parametric residuals $\left\{\hat{U}_i, i = 1, ..., n\right\}$, where $\hat{U}_i^* = \hat{U}_i V_i$ and $\{V_i, i = 1, ..., n\}$ are random variables such that

A7 V_i are bounded, *iid*, independent of $\mathcal{Y}_n$ and such that $E(V_1) = 0$ and $E(V_1^2) = 1$.

This resample procedure, known as "wild bootstrap," was introduced by Wu (1986) in the context of estimation in heteroskedastic linear models.

The first type of bootstrap statistic is inspired in the first order asymptotic expansion (6), provided in Theorem 1, where the unobserved errors $\{U_i, i = 1, ..., n\}$ are substituted by the resample $\left\{\hat{U}_i^*, i = 1, ..., n\right\}$. That is, the bootstrap statistic is

$$\bar{C}_n^* = \sum_{i=1}^{n} \bar{T}_n^* (W_i)^2,$$

where

$$\bar{T}_n^* (w) = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^* \hat{f}(X_i) \left[1(W_i \leq w) - \hat{\phi}(X_i, w)\right]. \tag{12}$$

Rather than obtaining the bootstrap analogs of the original statistics from a resample of the original sample $\mathcal{Y}_n = \{(Y_i, W_i), i = 1, ..., n\}$, we are approximating the unknown asymptotic distribution of the statistics by the conditional distribution of the bootstrap statistics computed from $\bar{T}_n^*$, which is the bootstrap estimator of the asymptotic expansion provided in (6). This way of approximating the asymptotic null distribution of non-pivotal statistics has been used before, in other contexts, by Sue and Wei (1991), Lewbel (1995), De Jong (1996) and Hansen (1996) among others.

The second method consists of using the bootstrap analogs of the statistics from a bootstrap sample $\mathcal{Y}_n^* = \{(Y_i^*, X_i), i = 1, ..., n\}$, where $Y_i^* = \hat{m}(X_i) + \hat{U}_i^*$. This way of obtaining "wild bootstrap" samples in a nonparametric regresion context has been proposed by Härdle and Marron (1991) in order to compute bootstrap confidence intervals in nonparametric regression. However, the bootstrap method proposed here can be implemented with different badwidths, if desired. In specification testing of parametric regression fuctions, the "wild bootstrap" has been applied by Härdle and Mammen (1993) in tests statistics based on smoothers and by Stute et al (1998) in test statistics based on estimates of the integral regression function. So the bootstrap statistic is

$$\hat{C}_n^* = \sum_{i=1}^{n} \hat{T}_n^* (W_i)^2,$$

where

$$\hat{T}_n^* (w) = \frac{1}{n} \sum_{i=1}^{n} (Y_i^* - \hat{m}^* (X_i)) \hat{f}(X_i) 1(W_i \leq w), \tag{13}$$

and $\hat{f}(X_i) \hat{m}^*(X_i) = (nh^q)^{-1} \sum_{j=1, i \neq j}^{n} Y_i^* K_{ij}$.

The first bootstrap method is easier to compute than the second one and needs weaker regularity conditions in order to prove its consistency. However, unlike the second bootstrap method, the first one is not mimicing the behaviour of the sample under the null hypothesis. So, it is expected that, the first bootstrap test behave, in small samples, like the asymptotic tests and the second bootstrap tests, based on the bootstrap analogs computed by a resample of the original sample under H_0, is expected to enjoy better level properties.

Let $\hat{\eta}_n$ be the statistic used for testing H_0 (i.e. $\hat{C}_n$) and $\hat{\eta}_n^*$ the corresponding bootstrap statistic used for testing H_0 (i.e. $\hat{C}_n^*$ or $\bar{C}_n^*$). At the $\alpha-level$ of significance, H_0 is rejected when $\hat{\eta}_n \geq c^*_{n(1-\alpha)}$, where $c^*_{n(1-\alpha)}$ is the bootstrap critical value, such that $\Pr\left\{\hat{\eta}_n^* \geq c^*_{n(1-\alpha)} \middle| \mathcal{Y}_n\right\} = \alpha$. If under H_0, $\hat{\eta}_n \rightarrow_d \eta_\infty$, the bootstrap test is consistent if under H_0, H_1, or H_{1n}, $\hat{\eta}_n^* \rightarrow_{d^*} \eta_\infty$ (ie. η_∞ can be C_∞ or K_∞), where "$\rightarrow_{d^*}$" means convergence in bootstrap distribution; that is, $\Pr\{\hat{\eta}_n^* \geq \zeta | \mathcal{Y}_n\} \rightarrow_p G(\zeta)$, for each continuity point ζ of G, where G is the distribution of η_∞ under H_0. So, the distribution of $\hat{\eta}_n^*$, conditional on the sample $\mathcal{Y}_n$, consistently estimates the asymptotic distribution of $\hat{\eta}_n$. Hence, the bootstrap critical values consistently estimate the asymptotic critical values, both under the null and under the alternative, and the resulting test is consistent. In practice the critical values $c^*_{n(1-\alpha)}$ can be approximated, as accurately as desired, by Monte Carlo. That is, we generate B bootstrap samples, $\{\mathcal{Y}_n^{*b}, b = 1, ..., B\}$ according to our resample squeme and the corresponding bootstap statistics are computed. Then, $c^*_{n(1-\alpha)}$ is approximated by $c^{*B}_{n(1-\alpha)}$, where $B^{-1}\sum_{i=1}^n 1\left(\hat{\eta}_i^* > c^{*B}_{n(1-\alpha)}\right) = \alpha$. The larger B, the better the approximation of $c^*_{n(1-\alpha)}$.

Under the same assumptions than in Theorem 1, we provide an asymptotic expansion of $\sqrt{n}\bar{T}_n^*$, which is very useful in order to prove the consistency of the bootstrap tests. Let us define $U_i^* = U_i V_i$. Hereforth, for a sequence of random variables D_n^*, we say that $D_n^* = D_n + o_{p^*}(1)$ if $\Pr\{|D_n^* - D_n| > \varepsilon | \mathcal{Y}_n\} \rightarrow_p 0$, for all $\varepsilon > 0$.

Theorem 3 *Under H_0, H_1 or H_{1n} and if A1 to A.7, uniformly in w,*

$$\sqrt{n}\bar{T}_n^*(w) = \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i^* f(X_i)\left[1(W_i \leq w) - \phi(X_i, w)\right] + o_{p^*}(1).$$

Given the above theorem, the consistency of the bootstrap test follows straightforwadly from Stute et al (1998) results, as stated in the following Corollary.

Corollary 2 *Under H_0, H_1 or H_{1n} and if A1 to A.7 hold,*

$$\bar{C}_n^* \xrightarrow{d^*} C_\infty.$$

In order to show the consistency of the boostrap tests based on $\hat{T}_n^*$, we show first that $\sqrt{n}\hat{T}_n^*$ and $\sqrt{n}\bar{T}_n^*$ have the same asymptotic distribution. Since

$$\sqrt{n}\hat{T}_n^*(w) = \sqrt{n}\bar{T}_n^*(w) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{m}(X_i)\hat{f}(X_i)\left[1(W_i \leq w) - \hat{\phi}(X_i, w)\right], \qquad (14)$$

we must prove that the second term on the right side of (14) is $o_p(1)$ uniformly in w. Such term, has a random denominator. This is why we need the following assumption.

A8.- $\Pr(f(X) > \vartheta) = 1$ for some $\vartheta > 0$.

Assumption A8 does not allow important distributions, like the Beta or the Normal. However, from a practical view point, this assumption is not so damaging. Another way of dealing with the random denominator problem, avoiding assumption A8, consists of introducing some trimming as suggested in Robinson (1988). It will imply the choice of a trimming parameter whose rate of convergence will be related to h. We also need stronger conditions than A6.

A6'.- $(nh^{2q})^{-1} + nh^{2\eta} \to 0$ as $n \to \infty$, where $\eta = \min(\mu, \lambda + 1)$.

Notice that A6' implies A6. Now, necessary conditions to conciliate the components of A6$'$ are $\mu > q$, $\lambda > q - 1$. Under this condition we can prove the asymptotic equivalence, up to the first order, between $\sqrt{n}\hat{T}_n^*$ and $\sqrt{n}\bar{T}_n^*$, as stated in the following Theorem.

Theorem 4 *Under H_0, H_1 or H_{1n} and if A1 to A5, A6', A7 and A8 hold, uniformly in w,*

$$\sqrt{n}\hat{T}_n^*(w) = \sqrt{n}\bar{T}_n^*(w) + o_{p^*}(1).$$

From the above Theorem, and applying Theorem 2.1 in Stute et al (1998), the consistency of tests based on $\hat{T}_n^*$ is inmediate.

Corollary 3 *Under H_0, H_1 or H_{1n} and if A1 to A5, A6', A7 and A8 hold,*

$$\hat{C}_n^* \xrightarrow{d^*} C_\infty.$$

The performance of the bootstrap test in small samples is studied by means of a Monte Carlo experiment in next section.

4.- MONTE CARLO

We have carried out a small Monte Carlo experiment in order to study the small sample performance of the tests. 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 $q = 1$ and $p = 1, 2$ under different designs. We choose a Gaussian kernel and $h = Cn^{-1/2}$, $C = 0.25, 0.5, 1, 2$. In all designs, a bandwidth choice $h = Cn^{-\alpha}$ is consistent with A6 when $1 < \alpha < 1/4$, and is consistent with A6' when $1/2 < \alpha < 1/4$. So, our bandwidth choice is consistent with A6 and is in the limit in order to satisfy A6'. Tables 1 and 2 report the proportion of rejections in 2000 Monte Carlo samples using 2000 bootstrap samples for approximating the critical values.

In Table 1 we examine the level accuracy of the bootstrap test. Samples are generated according to the design

$$Y_i = m(X_i) + U_i,\ i = 1, ..., n,$$

where $U_i \sim N(0, 1)$ independent of X_i, $Z_i^{(1)}$, $Z_i^{(2)}$ which are independently generated as $U(0, 1)$. We consider a linear model $m(x) = 1 + x$ and a sin model $m(x) = 1 + \sin(\gamma x)$ with $\gamma = 8, 10$. As γ increases, in the sin model, the regression curve has more frequencies and, hence, is more difficult to estimate. As it could be expected, the empirical size of the Wald's test is 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 sin model, higher bandwidth values produce serious size distorsions. As in other simulation studies for specification tests of parametric functions based on smoothers (see e.g. Delgado et al (1998)), it seems advisable 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, p. The bootstrap test based on $\hat{C}_n^*$ performs slightly better than the test based on $\tilde{C}_n^*$.

Table 2 examines the power properties of the test under the design

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

with X_i, $Z_i^{(1)}$, $Z_i^{(2)}$, U_i generated as before. We consider $\gamma = 5, 8, 10$, the correlation between Y_i and $Z_i^{(1)}$ decreases as γ increases (such correlation is close to 1 when $\gamma = 5$, to 0.3 when $\gamma = 8$ 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 $\gamma = 8$, the power of the Wald's test dramatically decreases while the bootstrap tests are still powerful. Finally, 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 cases are needed. The results are quite unsensitive to the choice of smoothing parameter and the dimension of the vector Z.

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

$$H_0 : E(Y|W) = m_0(W) \ a.s.,$$

where m_0 is the regression function when certain restrictions have been imposed, e.g. mean independence is the case considered in preceding sections. Other restrictions could be partial linearity, monotonicity, additivity, etc. The null hypothesis can be alternatively be written as

$$H_0 : E\left[(Y - m_0(W))\,\xi(W)\,1(W \leq w)\right] = 0, \ \forall w \in \mathcal{W},$$

where ξ is a weight function which does not change sign in the support of W, $\mathcal{W}$. Let $\hat{m}_0$ be an estimator of m_0. A test can be based on the empirical process

$$\hat{Q}_n(w) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{m}_0(W_i))\,\xi(W_i)\,1(W_i \leq w).$$

The choice of ξ, the limiting distribution of the empirical process and the construction of bootstrap tests will depend on the particular testing problem. Here, we only discuss the implementation of this methodoloy to tests partial linearity and conditional independence. However, application to tests of other restrictions seem also possible.

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 to reduce the curse of dimensionality in the estimation of a nonparametric curve. Estimators of this model has been proposed by Heckman (1986), Robinson (1988) and Speckman (1988) among others. Consider the null hypothesis

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

where θ_0 is an unknown parameter vector belonging to the parameter space Θ, and γ is an unknown function. The null hypothesis can be also written as,

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

where $m_Z(\cdot) = E(Z \mid X = \cdot)$. 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 regresors W.

As in section 2, we propose a test which only requires estimates of conditional expectations given X, $m(\cdot)$ and $m_Z(.)$. 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 empirical process,

$$\hat{Q}_n(w) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(X_i)\hat{U}_i^s 1(W_i \leq w),$$

where $\hat{U}_i^s = Y_i - \hat{m}(X_i) - (Z_i - \hat{m}_Z(X_i))' \hat{\theta}_n$ estimates the semiparametric errors $U_i^s = Y_i - m(X_i) - (Z_i - m_Z(X_i))' \theta_0$, and $\hat{m}_Z(X_i) = (nh^q)^{-1} \sum_{j=1, j\neq i}^{n} Z_j K_{ij}$ estimates $m_Z(X_i)$. Notice that

$$\hat{Q}_n(w) = \hat{T}_n(w) - \hat{\theta}_n' \frac{1}{n}\sum_{i=1}^{n} \hat{f}(X_i) Z_i \left[1(W_i \leq w) - \hat{\phi}(X_i, w)\right],$$

where $\hat{T}_n$ was defined in (5). From the above expression, it seems relatively straightforward to obtain, under regularity conditions in Robinson (1988), the limiting process of $\hat{Q}_n$. Using similar arguments, as in Theorem 1, we can obtain the same asymptotic expansions in (6) and (7), substituting U_i by U_i^s. The limiting process of $\sqrt{n}\hat{Q}_n$ is straightforwadly obtained from this asymptotic expansion, as well as the limiting distribution of a test statistic based on some functional of $\sqrt{n}\hat{Q}_n$. Also, a bootstrap test, like in section 3, can be implemented using the asymptotic expansion. Given a bootstrap sample of the nonparametric residuals $\left\{\hat{U}_i^{s*}, i = 1, ..., n\right\}$, where $\hat{U}_i^{s*} = \hat{U}_i^s V_i$ and V_i are random variables holding A7, the bootstrap process is identical to $\hat{T}_n^*$ in (12), substituting $\hat{U}_i^*$ by $\hat{U}_i^{s*}$. Using similar conditions and arguments in Theorem 3 and Corollary 2, it can be showed than the resulting test is consistent. The bootstrap analog of the process can be obtained from the resample $\mathcal{Y}_n^* = \{(Y_i^*, X_i), i = 1, ..., n\}$, where $Y_i^* = Z_i'\hat{\theta}_n + \hat{\gamma}(X_i) + \hat{U}_i^*$. The consistency of the resulting bootstrap test can be proved using similar arguments as in Theorem 4 and Corollary 3.

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

$$H_0 : E[1(Y \leq y)| W] = E[1(Y \leq y)| X] \;\; a.s. \; \forall y \in \mathcal{Y},$$

where $\mathcal{Y}$ 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 $1(Y \leq y)$. The null hypothesis can be alternatively written as

$$H_0 : E[f(X)(1(Y \leq y) - F(y| X)) 1(W \leq w)] = 0, \;\; \forall y \in \mathcal{Y} \text{ and } w \in \mathcal{W}, \tag{15}$$

where $F(\cdot \mid \cdot)$ is the distribution function of Y given X. The expectation in (15) can be estimated by

$$\hat{Q}_n(y, w) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(X_i) \left[1(Y_i \leq y) - \hat{F}(y | X_i)\right] 1(W_i \leq w),$$

where

$$\hat{F}(y | X_i) = \frac{1}{\hat{f}(X_i)} \frac{1}{nh^q} \sum_{i=1}^{n} 1(Y_i \leq y) K_{ij}$$

is an estimator of $F(y | X_i)$. A similar empirical process has been used by Andrews (1997) for specification testing of a parametric conditional distribution function. A test statistic is based on some functional of $\sqrt{n}\hat{Q}_n$. Using similar arguments and conditions in Theorem 1, we can obtain the same asymptotic expansions in (6) and (7), substituting U_i by $U_i(y) = 1(Y_i \leq y) - F(y | X_i)$. From this expansion, the asymptotic distribution of the test statistic can be derived using similar arguments as Andrews (1997). As in section 3, a bootstrap test can be implemented based on a bootstrap sample of the nonparametric residuals $\hat{U}_i(y) = 1(Y_i \leq y) - \hat{F}(y | X_i)$ and the asymptotic expansion. The resulting bootstrap test statistic is

$$\bar{Q}_n^*(y, w) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(X_i) \hat{U}_i^*(y) \left[1(W_i \leq w) - \hat{\phi}(X_i, w)\right],$$

where $\hat{U}_i^*(y) = \hat{U}_i(y) V_i$, and V_i holds conditions in A7. Then, a bootstrap test statistic can be based on some functional of $\sqrt{n}\bar{Q}_n^*$. Using arguments in Theorem 3, it can be showed that, uniformly in (y, w),

$$\bar{Q}_n^*(y, w) = \frac{1}{n} \sum_{i=1}^{n} f(X_i) U_i^*(y) \left[1(W_i \leq w) - \phi(X_i, w)\right] + o_{p^*}\left(n^{-1/2}\right),$$

where $U_i^*(y) = U_i(y) V_i$. From this expansion, it seems relatively straightforwadly to prove that a test based on some functional of $\sqrt{n}\bar{Q}_n^*$ is consistent. Alternatively, a bootstrap analog of $\hat{Q}_n$ can be constructed from a bootstrap sample $\{(Y_i^*, X_i), i = 1, ..., n\}$, where Y_i^* are generated from the estimated conditional distribution $\hat{F}(\cdot \mid X_i)$, see e.g. Cao-Abad and González-Manteiga (1993).

Gonzalo and Linton (1996) have proposed a test of conditional independence restrictions which does not need to use smoothing techniques. They also propose a bootstrap test, but consistency is not proved.

MATHEMATICAL APPENDIX

Through the appendix we use the following notation $m_i = m(X_i)$, $f_i = f(X_i)$, $\phi_i(w) = \phi(X_i, w)$, $r_i(z) = r(X_i, z)$, $g_i = g(W_i)$, $\sigma_i^2 = E(U_i^2 \mid W_i)$, $\hat{f}_i = \hat{f}(X_i)$, $\hat{m}_i = \hat{m}(X_i)$, $\tilde{g}_i = n^{-1}h^{-q}\sum_{j=1}^{n} g_j K_{ij}/\hat{f}_i$, $\tilde{m}_i = n^{-1}h^{-q}\sum_{j=1}^{n} m_j K_{ij}/\hat{f}_i$, $\tilde{U}_i = n^{-1}h^{-q}\sum_{j=1}^{n} U_j K_{ij}/\hat{f}_i$, $\Pi_i(w) = 1(W_i \leq w)$ and $E_{ijkl}(\cdot) = E(\cdot \mid X_i, X_j, X_k, X_l)$, $E^*(\cdot) = E(\cdot \mid \mathcal{Y}_n)$. The proofs apply some lemmas proved in the Lemmata.

PROOF OF THEOREM 1

Thorough the proofs, we use the fact that for each $i = 1, ...n$, $\Pi_i(w) \leq 1$, $\phi_i(w) \leq 1$ and $\hat{\phi}_i(w) \leq 1$, uniformly in w. Since the kernel is symmetric, $\sqrt{n}\hat{T}_n(w)$ is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \hat{f}_i \left(\Pi_i(w) - \hat{\phi}_i(w)\right) \tag{16}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Pi_i(w)\hat{f}_i(m_i - \tilde{m}_i). \tag{17}$$

Thus, (6), in Theorem 1, follows by Markov's inequality, after showing that, uniformly in w, $E\left((17)^2\right) = o(1)$, which is proved from

$$E\left[\Pi_i(w)^2 \hat{f}_1^2 (m_1 - \tilde{m}_1)^2\right] = o(1) \tag{18}$$

$$nE\left\{\hat{f}_2\hat{f}_1\Pi_1(w)\Pi_2(w)(m_1 - \tilde{m}_1)(m_2 - \tilde{m}_2)\right\} = o(1). \tag{19}$$

Define $r_i = (m_1 - m_i)K_{1i}$ and $r = E_1(r_2)$. On one hand, uniformly in w, (18) is bounded by

$$\begin{aligned} E\left[\hat{f}_1^2 (m_1 - \tilde{m}_1)^2\right] &\leq \frac{1}{n^2h^{2q}} E\left[\left(\sum_{i=1}^{n}(m_1 - m_i)K_{1i}\right)^2\right] &(20)\\ &\leq \frac{2}{nh^q}E\left[(r_2 - r_1)^2\right] + \frac{2}{h^{2q}}E\left(r^2\right) &(21)\\ &= O\left(\frac{1}{nh^q} + h^{2\eta}\right), \end{aligned}$$

since $E\left[(r_2 - r)^2\right] \leq E\left(r_2^2\right) = O(h^q)$ by Lemma 3, and $E\left(r^2\right) = O\left(h^{2(\eta+q)}\right)$ by Lemma 5. On the other hand, (19) is bounded, uniformly in w, by an expression proportional to

$$\frac{1}{nh^{2q}}\left|E\left[(m_1 - m_2)^2 K_{12}^2\right]\right| + \frac{1}{h^{2q}}\left|E\left\{(m_1 - m_2)K_{12}E_{12}\left[(m_2 - m_3)K_{23}\right]\right\}\right| + \frac{n}{h^{2q}}\left\{E\left[E_1\left[(m_1 - m_2)K_{12}\right]\right]\right\}^2,$$

where, the first term in the last expression is $O\left((nh^q)^{-1}\right)$ by Lemma 3, the second one $O(h^\eta)$ by Lemma 3,5 dominance convergence and Lemma 3, and the last term is $O\left(nh^{2\eta}\right)$ by Lemma 5.

Now, we prove (7) in Theorem 1. Notice that (16) is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i f_i\left(\Pi_i(w)-\phi_i(w)\right) \tag{22}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\left(\hat{f}_i-f_i\right)\left(\Pi_i(w)-\phi_i(w)\right) \tag{23}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\hat{f}_i\left(\phi_i(w)-\hat{\phi}_i(w)\right) \tag{24}$$

Thus, (7) follows by Markov's inequality, after showing that, uniformly in w, $E\left[(23)^2\right]=o(1)$ and $E\left[(24)^2\right]=o(1)$. On one hand,

$$E\left((23)^2\right)\leq E\left[\sigma_1^2\left(\hat{f}_1-f_1\right)^2\right]=O\left(\frac{1}{nh^q}+h^{2\lambda}\right), \tag{25}$$

applying lemmas 2 and 3 and using the same argument applied in the proof of the convergence of the second term in (20). On the other hand,

$$E\left((24)^2\right)\leq E\left[\sigma_1^2\hat{f}_1^2\left(\phi_1(w)-\hat{\phi}_1(w)\right)^2\right]=o(1), \tag{26}$$

applying Lemma 6 and using the same argument applied in the proof of the convergence of the second term in (20). ■

PROOF OF THEOREM 2

Given Theorem 1, it sufices to prove that $n^{-1/2}\sum_{i=1}^{n}U_i f_i\left[\Pi_i(w)-\phi_i(w)\right]$ converges weakly to $T_\infty^1(w)$, which follows using standard invariance principle arguments, as in the proof of Stute (1997) Theorem 1.1. ■

PROOF OF COROLLARY 1

Remark that in Theorem 1 we show that $\sqrt{n}\hat{T}_n(w)=(16)+(17)$, where $(17)=o_p(1)$, uniformly in w, under H_0, H_{1n} and H_1. Then, (9) follows applying Theorem 2 and the continuous mapping theorem. In order to prove (10), notice that under H_{1n},

$$\begin{aligned}(16) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i-E\left(U_i|W_i\right)\right)\hat{f}_i\left(\Pi_i(w)-\hat{\phi}_i(w)\right)+\frac{1}{n}\sum_{i=1}^{n}g_i\hat{f}_i\left(\Pi_i(w)-\hat{\phi}_i(w)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i-E\left(U_i|W_i\right)\right)f_i\left(\Pi_i(w)-\phi_i(w)\right)+\frac{1}{n}\sum_{i=1}^{n}g_i f_i\left(\Pi_i(w)-\phi_i(w)\right)+o_p(1),\end{aligned}$$

using same arguments as in the proof of (7) in Theorem 1. Then, applying Theorem 2 and the Law of large numbers, we obtain that $\sqrt{n}\hat{T}_n$ converges weakly to $T_\infty+S$, and (10) follows applying the continuous mapping theorem. Finally, (11) follows noticing that, by Theorem 1, uniformly in w,

$$\begin{aligned}\hat{T}_n(w) &= \frac{1}{n}\sum_{i=1}^{n} U_i f_i\left[\Pi_i(w)-\phi_i(w)\right]+o_p\left(n^{-1/2}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - E\left(Y_i|\,W_i\right)\right] f_i\left[\Pi_i(w)-\phi_i(w)\right]+\frac{1}{n}\sum_{i=1}^{n}\left[E\left(Y_i|\,W_i\right)-m_i\right] f_i\left[\Pi_i(w)-\varphi_i(w)\right]+o_p\left(n^{-1/2}\right)\end{aligned}$$

where, under H_1, by the law of large numbers, the first term of the last expression is $o_p(1)$ and the second one one diverges to infinity in probability, uniformly in w. ■

PROOF OF THEOREM 3

Since $\sqrt{n}\hat{T}_n^*(w) - n^{-1/2}\sum_{i=1}^{n} U_i^* f_i\left[\Pi_i(w)-\phi_i(w)\right]$ is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{U}_i^* - U_i^*\right)\hat{f}_i\left[\Pi_i(w)-\hat{\phi}_i(w)\right] \tag{27}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^*\hat{f}_i\left[\phi_i(w)-\hat{\phi}_i(w)\right] \tag{28}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i^*\left(\hat{f}_i - f_i\right)\left[\Pi_i(w)-\phi_i(w)\right], \tag{29}$$

it sufices to show that, uniformly in w, $E^*\left((27)^2\right) = o_p(1)$, $E^*\left((28)^2\right) = o_p(1)$ and $E^*\left((29)^2\right) = o_p(1)$. First, uniformly in w,

$$\begin{aligned}E^*\left((27)^2\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_i - m_i\right)^2\hat{f}_i^2\left[\Pi_i(w)-\hat{\phi}_i(w)\right]^2 \\ &\le \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}_i - m_i\right)^2\hat{f}_i^2 \\ &\le \frac{2}{n}\sum_{i=1}^{n}\left(m_i - \tilde{m}_i\right)^2\hat{f}_i^2 + \frac{2}{n}\sum_{i=1}^{n}\tilde{U}_i^2\hat{f}_i^2 \\ &= O\left(\left(nh^q\right)^{-1}+h^{2\eta}\right),\end{aligned} \tag{30}$$

uniformly in w, by Markov's inequality, after applying (20) and

$$E\left[\tilde{U}_1^2\hat{f}_1^2\right] = \frac{1}{nh^{2q}}E\left[\sigma_1^2 K_{12}^2\right] = O\left(\frac{1}{nh^q}\right),$$

by Lemma 3,

$$\begin{aligned}E^*\left[(28)^2\right] &= \frac{1}{n}\sum_{i=1}^{n} U_i^2\hat{f}_i^2\left[\phi_i(w)-\hat{\phi}_i(w)\right]^2\\ &= o_p(1),\end{aligned}$$

uniformly in w, by Markov's inequality, after applying (26), and

$$E^*\left[(29)^2\right] = \frac{1}{n}\sum_{i=1}^{n} U_i^2\left(\hat{f}_i - f_i\right)^2\left[\Pi_i(w) - \phi_i(w)\right]^2$$
$$\leq \frac{1}{n}\sum_{i=1}^{n} U_i^2\left(\hat{f}_i - f_i\right)^2$$
$$= O\left((nh^q)^{-1} + h^{2\lambda}\right),$$

uniformly in w, by Markov's inequality, after applying (25). Notice that the result holds under the null and under the alternative. ■

PROOF OF COROLLARY 3

By Theorem 3, $\sqrt{n}\bar{T}_n^*(w)$ and $n^{-1/2}\sum_{i=1}^{n} U_i^* f_i\left[\Pi_i(w) - \phi_i(w)\right]$ converges to the same limiting process. Then, the Corollary is proved using standard invariance principle arguments, conditional on $\mathcal{Y}_n$, as in Stute et al (1998) theorem. ■

PROOF OF THEOREM 4

Notice that

$$\sqrt{n}\hat{T}_n^*(w) = \sqrt{n}\bar{T}_n^*(w) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{m}_i\hat{f}_i\left[\Pi_i(w) - \hat{\phi}_i(w)\right] \quad (31)$$
$$= \sqrt{n}\bar{T}_n^*(w) + (17) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{m}_i - m_i)\hat{f}_i\left[\Pi_i(w) - \hat{\phi}_i(w)\right], \quad (32)$$

where, in Theorem 1, we proved that (17)$= o_p(1)$. The third term in (32) is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{m}_i - m_i)\hat{f}_i\left(f_i - \hat{f}_i\right)\left[\Pi_i(w) - \hat{\phi}_i(w)\right]}{f_i} \quad (33)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{m}_i - m_i)\hat{f}_i^2\left[\Pi_i(w) - \phi_i(w)\right]}{f_i} \quad (34)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{m}_i - m_i)\hat{f}_i^2\left[\phi_i(w) - \hat{\phi}_i(w)\right]}{f_i}. \quad (35)$$

Thus, it sufices to prove that the three terms in the last expression are $o_p(1)$ uniformly in w. The proof is quite lengthy, thus, we simplify notation by calling $\phi_i(w) = \phi_i$ and $\hat{\phi}_i(w) = \hat{\phi}_i$. By Cauchy-Swartz's inequality and A8, uniformly in w,

$$E(|(33)|) \leq C\sqrt{n}\left\{E\left[(\hat{m}_1 - m_1)^2\hat{f}_1^2\right]E\left[\left(f_i - \hat{f}_i\right)^2\right]\right\}^{1/2}$$
$$= O\left(\frac{1}{\sqrt{n}h^q} + \sqrt{n}h^{\eta+\lambda}\right),$$

by (30) and (25). Uniformly in w,

$$
\begin{aligned}
E\left((34)^2\right) &\leq C\left|E\left[(\hat{m}_1 - m_1)^2 \hat{f}_1^4\right]\right| && (36)\\
&\leq \frac{C}{n^4h^{4q}}\left|E\left(\sum_{k=2}^{n} U_k^2 K_{1k}^2 \sum_{i=2}^{n} K_{1i} \sum_{j=2}^{n} K_{1j}\right)^2\right| && (37)\\
&\quad + \frac{C}{n^4h^{4q}}\left|E\left[\left(\sum_{k=2}^{n}(m_k - m_1) K_{1k}\right)^2 \sum_{i=2}^{n} K_{1i} \sum_{j=2}^{n} K_{1j}^2\right]\right| && (38)
\end{aligned}
$$

where (37) is bounded, uniformly in w, by an expression proportional to

$$
\begin{aligned}
&\frac{1}{n^3h^{4q}}\left|E\left(\sigma_2^2 K_{12}^4\right)\right| + \frac{1}{n^2h^{4q}}\left|E\left[\sigma_2^2 K_{12}^3 E_{12}(K_{13})\right]\right| + \frac{1}{n^2h^{4q}}\left|E\left[\sigma_2^2 K_{12}^2 E_{12}\left(K_{13}^2\right)\right]\right|\\
&+\frac{1}{nh^{4q}}\left|E\left\{\sigma_2^2 K_{12}^2 E_{12}\left[E_{123}(K_{14}) K_{13}\right]\right\}\right|\\
=\;& O\left(\frac{1}{nh^q}\right),
\end{aligned}
$$

applying lemmas 2, 3 and dominance convergence; and (38) is, uniformly in w, bounded by an expression proportional to

$$
\begin{aligned}
&\frac{1}{n^3h^{4q}}\left|E\left[(m_2 - m_1)^2 K_{12}^4\right]\right| + \frac{1}{n^2h^{4q}}\left|E\left\{E_{12}[K_{13}](m_2 - m_1)^2 K_{12}^3\right\}\right|\\
&+\frac{1}{n^2h^{4q}}\left|E\left\{E_{12}\left[(m_2 - m_1)^2 K_{12}^2\right] K_{13}^2\right\}\right| + \frac{1}{n^2h^{4q}}\left|E\left\{E_{12}\left[(m_3 - m_1) K_{13}^2\right](m_2 - m_1) K_{12}^2\right\}\right|\\
&+\frac{1}{nh^{4q}}\left|E\left\{E_{12}\left[E_{123}(K_{14}) K_{13}\right](m_2 - m_1)^2 K_{12}^2\right\}\right| + \frac{1}{n^2h^{4q}}\left|E\left\{E_{12}\left[(m_3 - m_1) K_{13}\right](m_2 - m_1) K_{12}^3\right\}\right|\\
&+\frac{1}{nh^{4q}}\left|E\left\{E_{12}\left[E_{123}(K_{14})(m_3 - m_1) K_{13}\right](m_2 - m_1) K_{12}^2\right\}\right|\\
&+\frac{1}{h^{4q}}\left|E\left\{E_{14}\left(E_{145}\left[E_{1345}\left((m_2 - m_1) K_{12}\right)(m_3 - m_1) K_{13}\right] K_{15}\right) K_{14}\right\}\right|\\
=\;& o(1),
\end{aligned}
$$

since the first seven terms, in the last expression, are $O\left(h^2 (nh^q)^{-1}\right)$ by Lemmas 2, 6 and dominance convergence, and the last term is an $O(h^\eta)$ by Lemmas 3, 5 and dominance convergence. Second, uniformly in w, $E\left((35)^2\right)$ is bounded by

$$
CE\left[(\hat{m}_1 - m_1)^2 \hat{f}_1^4\right] + Cn\left|E\left[(\hat{m}_1 - m_1) \hat{f}_1^2 \left(\phi_1 - \hat{\phi}_1\right)(\hat{m}_2 - m_2) \hat{f}_2^2 \left(\phi_2 - \hat{\phi}_2\right)\right]\right|, \qquad (39)
$$

where the first term in (39) is the right side if (36) which is $O\left((nh^q)^{-1} + h^\eta\right)$ and the second term is bounded, uniformly in w, by an expression proportional to

$$
\frac{n}{h^{2q}}\left|E\left[(\hat{m}_1 - m_1) \hat{f}_1 (\phi_1 - \phi_3) K_{13} (\hat{m}_2 - m_2) \hat{f}_2 (\phi_2 - \phi_4) K_{24}\right]\right| \qquad (40)
$$

$$+\frac{1}{h^{2q}}\left|E\left[(\hat{m}_1-m_1)\hat{f}_1(\phi_1-\phi_2)K_{12}(\hat{m}_2-m_2)\hat{f}_2(\phi_2-\phi_3)K_{23}\right]\right| \tag{41}$$

$$+\frac{1}{nh^{2q}}\left|E\left[(\hat{m}_1-m_1)\hat{f}_1(\hat{m}_2-m_2)\hat{f}_2(\phi_1-\phi_2)^2K_{12}^2\right]\right| \tag{42}$$

Thus the proof of the Theorem is concluded by showing that, uniformly in w, (40) , (41) and (42) are $o(1)$. First, (40) is bounded, uniformly in w, by an expression proportional to

$$\begin{aligned}
&\frac{1}{nh^{4q}}\left|E\left[(m_3-m_1)K_{13}^2(m_3-m_2)K_{23}(\phi_1-\phi_3)(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{1}{nh^{4q}}\left|E\left[(m_1-m_2)^2K_{12}^2(\phi_1-\phi_3)K_{13}(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{1}{nh^{4q}}\left|E\left[(m_3-m_1)K_{13}^2(m_4-m_2)K_{24}^2(\phi_1-\phi_3)(\phi_2-\phi_4)\right]\right| \\
&+\frac{1}{nh^{4q}}\left|E\left[\sigma_3^2K_{13}^2K_{23}(\phi_1-\phi_3)(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{1}{h^{4q}}\left|E\left[(m_5-m_1)K_{15}(m_5-m_2)K_{25}(\phi_1-\phi_3)K_{13}(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{1}{h^{4q}}\left|E\left[\sigma_5^2K_{15}K_{25}(\phi_1-\phi_3)K_{13}(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{1}{h^{4q}}\left|E\left[(m_3-m_1)K_{13}^2(m_5-m_2)K_{25}(\phi_1-\phi_3)(\phi_2-\phi_4)K_{24}\right]\right| \\
&+\frac{n}{h^{4q}}\left|E\left[(m_5-m_1)K_{15}(m_6-m_2)K_{26}(\phi_1-\phi_3)K_{13}(\phi_2-\phi_4)K_{24}\right]\right|
\end{aligned}$$

$$\begin{aligned}
\leq\ & \frac{1}{nh^{4q}}E\left\{E_{23}\left[E_{123}(|K_{24}|)\,|m_3-m_1|\,K_{13}^2\right]|m_3-m_2|\,|K_{23}|\right\} \\
&+\frac{1}{nh^{4q}}\left|E\left\{E_{12}\left[E_{123}(|K_{24}|)\,|K_{13}|\right](m_1-m_2)^2K_{12}^2\right\}\right| \\
&+\frac{1}{nh^{4q}}\left\{E\left[|m_3-m_1|\,K_{13}^2\right]\right\}^2 \\
&+\frac{1}{nh^{4q}}E\left\{E_{23}\left[E_{123}(|K_{24}|)K_{13}^2\right]\sigma_3^2\,|K_{23}|\right\} \\
&+\frac{1}{h^{4q}}E\left\{\left(E_5\left[|m_5-m_1|\,|K_{15}|\,|E_{51}((\phi_1-\phi_3)K_{13})|\right]\right)^2\right\} \\
&+\frac{1}{h^{4q}}E\left\{\sigma_5^2\left(E_5\left[|K_{15}|\,|E_{51}((\phi_1-\phi_3)K_{13})|\right]\right)^2\right\} \\
&+\frac{1}{h^{4q}}E\left[|m_3-m_1|\,K_{13}^2\right]E\left\{|E_{24}\left[(m_5-m_2)K_{25}\right]|K_{24}||\right\} \\
&+\frac{n}{h^{4q}}\left(E\left\{|E_{13}\left[(m_5-m_1)K_{15}\right]|K_{13}||\right\}\right)^2 \\
=\ & o(1),
\end{aligned}$$

since the first and second terms, in the last expression, are $O\left(h^2(nh^q)^{-1}\right)$ and the third term $O\left(h^2(nh^{2q})^{-1}\right)$ by lemmas 2, 6 and dominance convergence; the fourth term is $O\left((nh^q)^{-1}\right)$ by lemmas 2,3 and dominance convergence; the fifth and sixth terms are $o(1)$ by lemmas 2,7 and dominance convergence, the seventh term is $O\left(h^{\eta-q+1}\right)=o\left(\sqrt{n}h^q\right)$ by Lemmas 2,5,6 and dominance convergence; and the eighth term is $O\left(nh^{2\eta}\right)$ by Lemmas 2,5 and dominance convergence. Second, (41) is, uniformly w, bounded by

$$
\begin{aligned}
& \frac{1}{n^2h^{4q}}\left|E\left[(m_3-m_1)K_{13}(m_3-m_2)K_{23}^2(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)\right]\right| \\
& +\frac{1}{n^2h^{4q}}\left|E\left[(m_2-m_1)^2K_{12}^3(\phi_1-\phi_2)(\phi_2-\phi_3)K_{23}\right]\right| \\
& +\frac{1}{n^2h^{4q}}\left|E\left[(m_2-m_1)K_{12}^2(m_3-m_2)K_{23}^2(\phi_1-\phi_2)(\phi_2-\phi_3)K_{23}\right]\right| \\
& +\frac{1}{n^2h^{4q}}\left|E\left[\sigma_3^2K_{13}K_{23}^2(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)\right]\right| \\
& +\frac{1}{nh^{4q}}\left|E\left[(m_4-m_1)K_{14}(m_4-m_2)K_{24}(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)K_{23}\right]\right| \\
& +\frac{1}{nh^{4q}}\left|E\left[(m_3-m_1)K_{13}(m_4-m_2)K_{24}(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)K_{23}\right]\right| \\
& +\frac{1}{nh^{4q}}\left|E\left[(m_4-m_1)K_{14}(m_3-m_2)K_{23}^2(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)\right]\right| \\
& +\frac{1}{nh^{4q}}\left|E\left[\sigma_4^2K_{14}K_{24}(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)K_{23}\right]\right| \\
& +\frac{1}{h^{4q}}\left|E\left[(m_4-m_1)K_{14}(m_5-m_2)K_{25}(\phi_1-\phi_2)K_{12}(\phi_2-\phi_3)K_{23}\right]\right| \\
\leq\ & \frac{1}{n^2h^{4q}}E\left\{E_{23}\left[|m_3-m_1|\,|K_{13}|\,|K_{12}|\right]|m_3-m_2|\,K_{23}^2\right\} \\
& +\frac{1}{n^2h^{4q}}E\left\{E_{12}\left(|K_{23}|\right)(m_2-m_1)^2|K_{12}|^3\right\} \\
& +\frac{1}{n^2h^{4q}}\left\{E_2\left[(m_2-m_1)^2K_{12}^2\right]\right\}^2 \\
& +\frac{1}{n^2h^{4q}}E\left\{E_{23}\left[|K_{13}|\,|K_{12}|\right]\sigma_3^2K_{23}^2\right\} \\
& +\frac{1}{nh^{4q}}E\left\{E_{14}\left[E_{124}\left(|K_{13}|\right)|m_4-m_2|\,|K_{24}|\,|K_{12}|\right]|m_4-m_1|\,|K_{14}|\right\} \\
& +\frac{1}{nh^{4q}}E\left\{E_{23}\left[E_{21}\left(|m_3-m_1|\,|K_{13}|\right)|K_{12}|\right]E_{23}\left[|m_4-m_2|\,|K_{24}|\right]|K_{23}|\right\} \\
& +\frac{1}{nh^{4q}}E\left\{E_{12}\left[|m_4-m_1|\,|K_{14}|\right]E_{12}\left(|m_3-m_1|\,K_{13}^2\right)|K_{12}|\right\} \\
& +\frac{1}{nh^{4q}}E\left\{E_{14}\left[E_{124}\left(|K_{13}|\right)|K_{24}|\,|K_{12}|\right]\sigma_4^2\,|K_{14}|\right\} \\
& +\frac{1}{h^{4q}}E\left\{E_{123}\left[|m_4-m_1|\,|K_{14}|\right]\left|E_{123}\left[(m_5-m_2)K_{25}\right]\right|\,|K_{23}|\,|K_{12}|\right\} \\
=\ & o(1),
\end{aligned}
$$

since, the first three terms, in the last expression, are $O\left(h^2\left(n^2h^{2q}\right)^{-1}\right)$ by lemmas 2,6 and dominance convergence, the fourth term is $O\left(\left(n^2h^{2q}\right)^{-1}\right)$ by lemmas 3,6 and dominance convergence, the fifth, sixth and seventh terms are $O\left(h^2\left(nh^q\right)^{-1}\right)$ by lemmas 2,6 and dominance convergence, the eighth term is $O\left((nh^q)^{-1}\right)$ by Lemmas 2,3 and dominance convergence, and the nineth term is $O\left(h^{\eta+1}\right)$ by lemmas 2,5,6 and dominance convergence. Finally (42) is bounded, uniformly in w, by an expression proportional to

$$\begin{aligned}&\frac{1}{n^3h^{4q}}E\left[(m_1-m_2)^2K_{12}^4(\phi_1-\phi_2)^2\right]\\&+\frac{1}{n^2h^{4q}}\left|E\left[(m_3-m_1)K_{13}(m_3-m_2)K_{23}(\phi_1-\phi_2)^2K_{12}^2\right]\right|\\&+\frac{1}{n^2h^{4q}}\left|E\left[(m_2-m_1)^2K_{12}^3(m_3-m_2)K_{23}(\phi_1-\phi_2)^2\right]\right|\\&+\frac{1}{n^2h^{4q}}\left|E\left[\sigma_3^2K_{13}K_{23}(\phi_1-\phi_2)^2K_{12}^2\right]\right|\\&+\frac{1}{nh^{4q}}\left|E\left[(m_3-m_1)K_{13}(m_4-m_2)K_{24}(\phi_1-\phi_2)K_{12}^2\right]\right|\end{aligned}$$

since the first term, in the last expression, is $O\left(h^2\left(n^3h^{3q}\right)^{-1}\right)$ by Lemma 6, the second term is $O\left(h^2\left(n^2h^{2q}\right)^{-1}\right)$ by Lemma 6 and dominance convergence, the third term is $O\left(h^{\eta+1}\left(n^2h^{2q}\right)^{-1}\right)$ by lemmas 5,6 and dominance convergence, the fourth term is $O\left(\left(n^2h^{2q}\right)^{-1}\right)$ by lemmas 2,3 and dominance convergence, and the fifth term is $O\left(h^{\eta+1}\left(nh^q\right)^{-1}\right)$.

PROOF OF COROLLARY 3

It follows applying Theorem 4 and Corollary 2.

LEMMATA

The next five lemmas are the Lemmata in Robinson (1988). Lemmas 6 and 7 are straightforwardly proved from the previous lemmas.

Lemma 1 *Let* $\sup_u|k(u)|+\int_{-\infty}^{\infty}\left|u^\lambda k(u)\right|du<\infty$, *for some* $\lambda\geq 0$. *Then uniformly in* x

$$\int_{-\infty}^{\infty}\|y-z\|^\lambda\left|K\left(\frac{y-z}{h}\right)\right|dy\leq Ch^{q+\lambda}.$$

Proof.- Robinson (1988) Lemma 1.

Lemma 2 *Let* $\sup_x f(x)<\infty$, $\sup_u|k(u)|+\int_{-\infty}^{\infty}|k(u)|\,du<\infty$. *Then uniformly in* x

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

Proof.- Robinson (1988) Lemma 2.

Lemma 3 *Let* $\sup_x f(x)<\infty$, $E\left[|s(X)|\right]<\infty$, $\sup_u|k(u)|+\int_{-\infty}^{\infty}|k(u)|\,du<\infty$. *Then uniformly in* x

$$E\left|s(X)K\left(\frac{X-x}{h}\right)\right|\leq Ch^q.$$

Proof.- Robinson (1988) Lemma 3.

Lemma 4 *For λ satisfying $l-1<\lambda\le l$, where $l\ge 1$ is an integer, let $f\in\mathcal{G}_{\lambda}^{\infty}$, $k\in\mathcal{K}_l$. Then uniformly in x*

$$E\left\{K\left(\frac{X-x}{h}\right)-h^{q}f(x)\right\}\le Ch^{q+\lambda}.$$

Proof.- Robinson (1988) Lemma 4.

Lemma 5 *For λ,μ satisfying $l-1<\lambda\le l$, $m-1<\mu\le m$, where $l\ge 1$, $m\ge 1$ are integers, let $f\in\mathcal{G}_{\lambda}^{\infty}$ and $s\in\mathcal{G}_{\mu}^{\alpha}$ some $\alpha>0$, $k\in\mathcal{K}_{l+m-1}$. Then uniformly in x*

$$E\left\{[s(X)-s(x)]K\left(\frac{X-x}{h}\right)\right\}\le CS(x)h^{q+\min(\lambda+1,\mu)},$$

where $E[|S(X)|^{\alpha}]<\infty$.

Proof.- Robinson (1988) Lemma 5.

Lemma 6 *For λ,μ satisfying $l-1<\lambda\le l$, $m-1<\mu\le m$, where $l\ge 1$, $m\ge 1$ are integers, let $f\in\mathcal{G}_{\lambda}^{\infty}$ and $s\in\mathcal{G}_{\lambda}^{\alpha}$ some $\alpha>0$. $k\in\mathcal{K}_l$. Then uniformly in x, for $\gamma\le\alpha$,*

$$E\left\{|s(X)-s(z)|^{\gamma}\left|K\left(\frac{X-x}{h}\right)\right|\right\}\le CS(x)h^{q+\gamma},\tag{43}$$

where $S(x)=S_1(x)^{\gamma}+|s(x)|^{\gamma}+E[|s(X)|^{\gamma}]$ and $\sup_{y\in S_{x\rho}}|s(y)-s(x)|/\|y-x\|\le S_1(x)$.

Proof.- The left side of (43) is bounded by

$$\begin{aligned}&CS_1(x)^{\gamma}\int_{S_{x\rho}}\|y-x\|^{\gamma}\left|K\left(\frac{y-x}{h}\right)\right|dy+C\int_{\overline{S}_{x\rho}}\|y-x\|^{q+\gamma}f(y)\,|s(y)-s(x)|\left|K\left(\frac{y-x}{h}\right)\right|dy\\ \le\;& Ch^{q+\gamma}S_1(x)+Ch^{q+\gamma}\left[|s(x)|^{\gamma}+E(|s(X)|^{\gamma})\right]\sup_{u}\left\{|u|^{q+\gamma}|k(u)|^{q}\right\}.\end{aligned}$$

Lemma 7 *If $k\in\mathcal{K}_2$ and $r(\cdot,w)\in\mathcal{G}_{v}^{\infty}$ some $v>0$, uniformly in w, then uniformly in w and s.*

$$E\left\{[\phi(s,w)-\phi(X,w)]K\left(\frac{X-s}{h}\right)\right\}=o(h^{q}),\tag{44}$$

Proof.- The left side of (44) is equal to

$$E\left\{1(X\le x)(r(s,z)-r(X,z))K\left(\frac{X-s}{h}\right)\right\}\tag{45}$$

$$+r(s,z)E\left\{[1(s\le x)-1(X\le x)]K\left(\frac{X-s}{h}\right)\right\},\tag{46}$$

where, uniformly in w and s,

$$(45) \leq E\left\{\left|\left|r(s,z) - r(X,z)\right| K\left(\frac{X-s}{h}\right)\right|\right\} \leq Ch^{q+1} \tag{47}$$

by Lemma 6, using the fact that, since $r(x,\cdot) \in \mathcal{G}_v^\infty$, $\sup_x S(x) < \infty$, and (46) is equal to $r(s,x)$ times

$$\begin{aligned}
& 1(s \leq x) \int_{-\infty}^{\infty} K\left(\frac{s-u}{h}\right)[f(s) - f(u)]\,du && (48)\\
& + \int_{-\infty}^{x} K\left(\frac{s-u}{h}\right)[f(u) - f(s)]\,du && (49)\\
& -f(s)\left\{\int_{-\infty}^{x} K\left(\frac{s-u}{h}\right)du - h^q 1(s \leq x)\right\} && (50)\\
= \;& o(h^q),
\end{aligned}$$

since $(48) \leq Ch^{q+\lambda}$ by Lemma 4,

$$(49) \leq \int_{-\infty}^{\infty}\left|K\left(\frac{s-u}{h}\right)\right|\left|f(s) - f(u)\right|du \leq Ch^{q+1},$$

as in the proof of Lemma 6, and $(50) = o_p(1)$ using the fact that,

$$\lim_{n\to\infty} \frac{1}{h^q}\int_{-\infty}^{x} K\left(\frac{s-u}{h}\right)du = \lim_{n\to\infty}\int_{-\infty}^{(x-s)/h} K(u)\,du = \begin{cases} 1 \text{ if } s \leq x \\ 0 \text{ if } s > x \end{cases}.$$

REFERENCES

[1] ANDREWS, D.W.K. (1997): "A Conditional Kolmogorov Test," *Econometrica*, **65**, 1097-1128.

[2] BIERENS, H. (1982): "Consistent Model Specification Test," *Journal of Econometrics*, **20**, 105-134.

[3] BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, **58**, 1443-1458.

[4] BIERENS, H. AND W. PLOBERGER (1997): "Asymptotic theory of integrated conditional moment test," *Econometrica*, **65**, 1153-1174.

[5] BRUNK, H.D (1970): "Estimation by isotonic regression," in *Nonparametric Techniques in Statisitical Inference* (M.L. Puri ed.), 177-197. Cambridge University Press.

[6] CAO-ABAD, R. AND W. GONZALEZ-MANTEIGA (1993): "Bootstrap methods in regression smoothing," *Nonparametric Statistics*, **2**, 379-388.

[7] CRAMÉR, H. (1928): "On the composition of elementary errors," *Skandinavisk Aktuarietidskrift, Vol. 11* (1928), pp. 13-74, 141-180.

[8] DE JONG, R.M. (1996): "The Bierens test under data dependence," *Journal of Econometrics*, **72**, 1-32.

[9] DELGADO, M.A., M.A. DOMINGUEZ AND P. LAVERGNE (1997): "Consistent specification testing of nonlinear in variables symultaneous equations models," manuscript.

[10] EUBANK, R. AND S. SPIEGELMAN (1990): "Testing the goodness of fit of a linear model via nonparametric regression techniques," *Journal of the American Statatistical Association*, **85**, 387-392.

[11] FAN, Y. AND Q. LI (1996): "Consistent model especification tests: omitted variables and semiparametric functional forms," *Econometrica*, **64**, 865-890.

[12] GOZALO, P.L. (1993): "A consistent model especification test for nonparametric estimation of regression function models," *Econometric Theory*, **9**, 451-477.

[13] GOZALO, P.L. AND O.B. LINTON (1996): "Conditional independence restrictions: testing and estimation," Yale University, Cowless Foundation Discussion Paper 1140.

[14] HANSEN, B. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, **64**, 413-430.

[15] HÄRDLE, W. AND E. MAMMEN (1993): "Comparing nonparametric versus parametric regression fits," *The Annals of Statistics*, **21**, 1926-1947.

[16] HÄRDLE, W. AND J.S MARRON (1991): "Bootstrap simultaneous error bars for nonparametric regression," *The Annals of Statistics*, **19**, 778-796.

[17] HECKMAN, N.E. (1986): "Spline smoothing in a partly linear model," *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.

[18] HONG-ZHY, A AND C. BING (1991): "A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series," *International Statistical Review*, **59**, pp. 287-307.

[19] HONG, Y. AND H. WHITE (1995): "Consistent specification testing via nonparametric series regression," *Econometrica*, **63**, 1133-1159.

[20] HOROWITZ, J. L. AND W. HÄRDLE (1994): "Testing a parametric model against a nonparametric alternative," *Econometric Theory,* **10**, 821-848.

[21] KOLMOGOROV, A.N. (1933): "Sulla determinazione empirica delle leggi di probability," *Giorn. Ist. Ital. Attuari,* **4**, 1-11.

[22] LAVERGNE, P. AND Q. VUONG (1996): "Nonparametric selection of regressors: the nonnested case," *Econometrica*, **64**, 207-219.

[23] LEWBEL, A. (1995): "Consistent nonparametric hypothesis tests with an application to Slutsky symmetry," *Journal of Econometrics,* **67**, 379-401.

[24] POWELL, J.L., J.H. STOCK AND T.M. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica*, **57**, 1403-1430..

[25] ROBINSON, P.M. (1988): "Root-n-consistent semiparametric regression," *Econometrica*, **56**, 931-954.

[26] ROBINSON, P.M. (1989): "Hypothesis testing in semiparametric and nonparametric models for econometric time series," *Review of Economic Studies,* **58**, 437-453.

[27] ROBINSON, P.M. (1991): "Consistent nonparametric entropy-based testing," *Review of Economic Studies*, **58**, 437-453.

[28] ROSENBLATT, M. (1975): "A Quadratic Measure of Deviations of Two-dimensional Density Estimates an a Test of Independence," *Annals of Statistics*, **3**, pp. 1-14.

[29] SMIRNOV, N.V. (1936): "Sur la distribution de ω^2," *C. R. Acad. Sci. Paris.,* **202**, 449.

[30] SPECKMAN, P. (1988): "Kernel smoothing in partially linear models," *Journal of the Royal Statistical Society, Series B,* **50**, 413-446.

[31] STUTE, W. (1997): "Nonparametric Model Checks for Regression," *Annals of Statistics,* **25**, 613-641.

[32] STUTE, W., W.G. GONZALEZ-MANTEIGA AND M. PRESEDO (1998): "Bootstrap Approximations in Model Checks for Regression," *Journal of American Statistics Association*, **93**, 141-149.

[33] SUE, J.Q. AND L.J. WEI (1991): "A lack of fit test for the mean function in a generalized linear model," *Journal of the American Statistical Association,* **86**, pp. 420-426.

[34] VON MISES, R. (1931): *Wahrscheinlichkitsrechnung*, Deuticke, Vienna.

[35] WANG, Y-J. AND D.W.K. ANDREWS (1993): "Tests of specification for parametric and semiparametric models," *Journal of Econometrics*, **57**, 277-318.

[36] WU, C.F.J. (1986): "Jacknife, bootstrap and other resampling methods in regression analysis," *Annals of Statistics*, **14**, 1261-1350.

[37] ZHENG, X. (1996): "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics*, **75**, 263-289.

TABLE 1

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

$m(x) = 1 + x$

$p = 1$									
	α	$n = 50$				$n = 100$			
	0.1	0.098				0.103			
Asymptotic Wald	0.05	0.054				0.052			
	0.01	0.016				0.013			
		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.151	0.126	0.105	0.101	0.138	0.119	0.112	0.113
Bootstrap Analog	0.05	0.080	0.062	0.050	0.050	0.075	0.067	0.056	0.055
	0.01	0.011	0.010	0.007	0.006	0.018	0.011	0.011	0.011
	0.1	0.160	0.131	0.111	0.142	0.144	0.122	0.114	0.146
Approx. Bootstrap	0.05	0.088	0.065	0.059	0.075	0.079	0.068	0.060	0.076
	0.01	0.014	0.012	0.011	0.014	0.019	0.013	0.012	0.018
$p = 2$									
	α	$n = 50$				$n = 100$			
	0.1	0.126				0.105			
Asymptotic Wald	0.05	0.070				0.053			
	0.01	0.021				0.016			
		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.164	0.126	0.100	0.102	0.157	0.120	0.100	0.093
Bootstrap Analog	0.05	0.072	0.049	0.038	0.038	0.072	0.055	0.047	0.045
	0.01	0.009	0.003	0.004	0.003	0.014	0.011	0.095	0.008
	0.1	0.175	0.132	0.109	0.120	0.162	0.125	0.102	0.103
Approx. Bootstrap	0.05	0.080	0.055	0.043	0.056	0.080	0.057	0.050	0.058
	0.01	0.015	0.006	0.004	0.009	0.015	0.012	0.010	0.010

TABLE 1 (Cont.)

$m(x) = 1 + \sin(8x)$

$p = 1$

	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.096				0.102			
	0.05	0.053				0.051			
	0.01	0.010				0.012			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.153	0.123	0.135	0.354	0.138	0.119	0.125	0.429
	0.05	0.081	0.062	0.069	0.173	0.075	0.067	0.062	0.242
	0.01	0.010	0.010	0.008	0.023	0.018	0.011	0.011	0.037
Approx. Bootstrap	0.1	0.165	0.145	0.190	0.437	0.144	0.122	0.162	0.598
	0.05	0.089	0.074	0.101	0.242	0.079	0.068	0.083	0.363
	0.01	0.015	0.010	0.017	0.048	0.019	0.013	0.018	0.086

$p = 2$

	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.121				0.105			
	0.05	0.068				0.058			
	0.01	0.014				0.015			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.171	0.127	0.109	0.173	0.154	0.116	0.109	0.206
	0.05	0.072	0.053	0.047	0.077	0.074	0.056	0.054	0.104
	0.01	0.008	0.004	0.004	0.013	0.014	0.012	0.009	0.016
Approx. Bootstrap	0.1	0.182	0.145	0.145	0.222	0.160	0.124	0.129	0.279
	0.05	0.085	0.066	0.070	0.108	0.081	0.060	0.065	0.151
	0.01	0.014	0.005	0.011	0.020	0.014	0.012	0.014	0.034

TABLE 1 (Cont.)

$m(x) = 1 + \sin(10x)$

$p = 1$									
	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.099				0.106			
	0.05	0.052				0.054			
	0.01	0.012				0.011			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.153	0.129	0.152	0.269	0.142	0.117	0.126	0.406
	0.05	0.079	0.063	0.063	0.131	0.076	0.062	0.062	0.197
	0.01	0.011	0.010	0.010	0.016	0.017	0.013	0.013	0.037
Approx. Bootstrap	0.1	0.163	0.148	0.195	0.308	0.145	0.125	0.171	0.504
	0.05	0.089	0.075	0.093	0.167	0.081	0.068	0.080	0.285
	0.01	0.015	0.014	0.013	0.031	0.018	0.014	0.021	0.063
$p = 2$									
	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.132				0.111			
	0.05	0.076				0.055			
	0.01	0.019				0.015			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.168	0.120	0.113	0.160	0.153	0.115	0.107	0.097
	0.05	0.070	0.054	0.044	0.074	0.075	0.058	0.052	0.088
	0.01	0.009	0.004	0.005	0.007	0.013	0.011	0.009	0.011
Approx. Bootstrap	0.1	0.181	0.141	0.145	0.181	0.161	0.128	0.132	0.246
	0.05	0.086	0.064	0.069	0.097	0.081	0.064	0.064	0.119
	0.01	0.013	0.005	0.009	0.012	0.014	0.012	0.013	0.024

TABLE 2

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

$\gamma = 5$

$p = 1$									
	α	$n = 50$				$n = 100$			
Asymptotic t-ratio	0.1	0.980				1.000			
	0.05	0.963				1.000			
	0.01	0.878				0.997			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.955	0.965	0.961	0.923	1.000	1.000	1.000	0.999
	0.05	0.912	0.926	0.921	0.854	0.998	1.000	1.000	0.998
	0.01	0.675	0.736	0.735	0.634	0.988	0.992	0.990	0.982
Approx. Bootstrap	0.1	0.962	0.967	0.962	0.933	1.000	1.000	1.000	1.000
	0.05	0.922	0.929	0.921	0.870	0.998	1.000	1.000	0.998
	0.01	0.749	0.760	0.729	0.621	0.990	0.991	0.990	0.984

$p = 2$									
	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.971				1.000			
	0.05	0.944				1.000			
	0.01	0.824				0.992			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.797	0.802	0.780	0.714	0.975	0.980	0.979	0.969
	0.05	0.656	0.671	0.657	0.583	0.945	0.950	0.953	0.933
	0.01	0.315	0.349	0.348	0.303	0.815	0.836	0.831	0.788
Approx. Bootstrap	0.1	0.814	0.810	0.789	0.736	0.977	0.980	0.979	0.973
	0.05	0.686	0.684	0.659	0.594	0.948	0.953	0.954	0.937
	0.01	0.378	0.375	0.355	0.284	0.838	0.843	0.834	0.793

TABLE 2 (Cont.)

$\gamma = 8$

	$p = 1$								
	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.199				0.276			
	0.05	0.126				0.178			
	0.01	0.048				0.067			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.675	0.665	0.656	0.642	0.959	0.968	0.973	0.971
	0.05	0.478	0.471	0.452	0.436	0.893	0.893	0.899	0.905
	0.01	0.156	0.153	0.148	0.134	0.567	0.577	0.585	0.554
Approx. Bootstrap	0.1	0.697	0.678	0.679	0.688	0.962	0.970	0.975	0.977
	0.05	0.509	0.490	0.478	0.504	0.898	0.899	0.910	0.921
	0.01	0.184	0.168	0.169	0.174	0.589	0.592	0.596	0.618
	$p = 2$								
	α	$n = 50$				$n = 100$			
Asymptotic Wald	0.1	0.200				0.239			
	0.05	0.122				0.154			
	0.01	0.048				0.047			
Bootstrap Analog		$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$	$C = 0.25$	$C = 0.5$	$C = 1$	$C = 2$
	0.1	0.436	0.399	0.364	0.351	0.747	0.740	0.735	0.722
	0.05	0.257	0.222	0.203	0.194	0.556	0.535	0.522	0.502
	0.01	0.044	0.039	0.038	0.038	0.197	0.176	0.167	0.155
Approx. Bootstrap	0.1	0.465	0.416	0.383	0.382	0.757	0.753	0.743	0.742
	0.05	0.283	0.236	0.203	0.211	0.576	0.550	0.531	0.535
	0.01	0.068	0.048	0.038	0.042	0.216	0.186	0.171	0.171

TABLE 2 (Cont.)

$\gamma = 10$

$p=1$									
	α	$n=50$				$n=100$			
Asymptotic Wald	0.1	0.096				0.096			
	0.05	0.054				0.052			
	0.01	0.013				0.012			
Bootstrap Analog		$C=0.25$	$C=0.5$	$C=1$	$C=2$	$C=0.25$	$C=0.5$	$C=1$	$C=2$
	0.1	0.473	0.447	0.410	0.410	0.813	0.821	0.814	0.796
	0.05	0.268	0.250	0.237	0.237	0.627	0.623	0.615	0.599
	0.01	0.059	0.058	0.050	0.050	0.252	0.250	0.237	0.232
Approx. Bootstrap	0.1	0.498	0.468	0.433	0.433	0.823	0.827	0.824	0.833
	0.05	0.294	0.267	0.259	0.259	0.664	0.631	0.626	0.640
	0.01	0.075	0.062	0.057	0.057	0.269	0.263	0.251	0.271

	$p=2$								
	α	$n=50$				$n=100$			
Asymptotic Wald	0.1	0.130				0.105			
	0.05	0.073				0.057			
	0.01	0.016				0.019			
Bootstrap Analog		$C=0.25$	$C=0.5$	$C=1$	$C=2$	$C=0.25$	$C=0.5$	$C=1$	$C=2$
	0.1	0.290	0.244	0.213	0.205	0.496	0.455	0.433	0.411
	0.05	0.152	0.126	0.104	0.095	0.305	0.272	0.252	0.235
	0.01	0.020	0.017	0.015	0.013	0.073	0.061	0.059	0.054
Approx. Bootstrap	0.1	0.311	0.256	0.225	0.235	0.509	0.462	0.443	0.436
	0.05	0.172	0.139	0.113	0.123	0.324	0.282	0.259	0.268
	0.01	0.033	0.021	0.019	0.023	0.080	0.067	0.063	0.068