Report 99-031

Measures of Deviation for Nonparametric Tests of Regression

Cheikh A.T. Diack

ISSN: 1389-2355

Measures of Deviation for Nonparametric Tests of Regression*

Cheikh A. T. DIACK
EURANDOM, P.O. Box 513, 5600 MB Eindhoven,
The Netherlands. E-mail:diack@eurandom.tue.nl

September 9, 1999

Abstract

Let (X, Y) have regression function $r(x) = E(Y|X = x)$. We consider the problem of nonprametric tests for a variety of hypotheses including specification, significance, monotonicity and convexity. The proposed tests involve an estimator of $r(x)$ and its first and second derivatives. The estimates used are a variant of the Nadaraya-Watson kernel type proposed by Mack and Müller (1989). The asymptotic distribution for these estimates of the maximal deviation from $r^{(p)}(x)$ $(p \geq 0)$ is proved. Limit theorems for quadratic norms for the estimates are also obtained. Using these results we study the behavior of the nonparametric tests.

1 Introduction

Our starting point is a random sample $(X_1, Y_1), ..., (X_n, Y_n)$ from a bivariate population with distribution function $G(x, y)$ and density $g(x, y)$ with respect to the Lebesgue measure. This probability distribution is assumed to be unknown. The paper investigates the tests of hypothesis like specification, significance, monotonicity and convexity for the unknown regression function $r(x) = \mathbb{E}(Y|X = x)$. The literature on nonparametric tests is extensive. Specification tests are proposed by Hausman (1978), Bierens (1982, 1990), Lee (1988), Eubank and Spiegelman (1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White(1995) and Yatchew and Bos (1997). Stoker (1989, 1991) proposes tests of significance. Surveys in testing monotonicity and convexity include Schlee (1980), Yatchew (1992), Yatchew and Bos (1997), Diack and Thomas

**Key words and phrases*: Global measure of deviation, kernel estimates, derivative, nonparametric tests, local alternatives, specification, significance, monontonicity, convexity.

(1998), Bowman and al. (1998), Diack (1998, 1999[7],1999[8]) and Doveh and al.(1999).

In this paper, we continue this effort by proposing new consistent nonparametric tests. The proposed tests are based on the asymptotic distribution of either a normalized maximal deviation quantity or a quadratic norm of the same quantity. They involve an estimator of $r(x)$ and its first and second derivatives. We use the kernel method to estimate the regression function.

The kernel estimate $\hat{r}(x)$ of $r(x)$ (due to Watson 1964 and Nadaraya 1964) is motivated by the formula:

$$r(x) = \left\{ \int yg(x,y)\,dy \right\} / f(x)$$

where f denotes the marginal density of X. The Nadaraya-Watson estimator is defined by:

$$\hat{r}(x) = \left\{ (nh_n)^{-1} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right) \right\} \left\{ (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \right\}^{-1} \tag{1}$$

where h_n is a bandwidth sequence and K is a kernel function. It is natural to consider an estimator of the p-th derivative of $r(x)$ to be the p-th derivative of $\hat{r}(x)$. The difficulty in computating $\frac{d^p}{dx^p}\hat{r}(x)$ alone is incentive enough to seek more convenient forms of estimates for $r^{(p)}(x)$. Thus, we will use the following variant of the Nadaraya-Watson kernel type estimator proposed by Mack and Müller (1989):

$$\hat{r}_{n,p}(x) = \left(nh_n^{p+1}\right)^{-1} \sum_{i=1}^{n} K^{(p)}\left(\frac{x - X_i}{h_n}\right) Y_i / \hat{f}_n(X_i) \tag{2}$$

where

$$\hat{f}_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{3}$$

is the kernel estimate of the marginal density $f(x)$ and $K^{(p)}$ is the p-th derivative of K. Consistency and asymptotic normality of these estimators are proved in Mack and Müller (1989). We extend these results by proving the asymptotic distribution of the maximal deviation from $r^{(p)}$. We also provide limit theorems for quadratic norms of the normalized deviations of the estimates from their expected values. Using a kernel estimate when f is known and the Yang estimate when f is known or unknown, Johnston (1982) gives results about the the asymptotic distribution of the maximal deviation from $r(x)$. Thus our results extend in some sense the results of Johnston (1982).

The remainder of this paper is as follows. In Section 2, we introduce further notation and results about some global measures of deviations. These results are used to provide nonparametric tests in Section 3. We also discuss their consistency and power. Section 4 is for the proofs of our results.

2 Uniform and quadratic confidence bands

In this section, we give uniform confidence bands for the regression function $r(x)$ and its derivatives. To be more precise, we provide the limit distribution of the maximal absolute deviation

$$\sup_{0\leq x\leq 1} \left|\hat{r}_{n,p}(x) - r^{(p)}(x)\right|. \tag{4}$$

A similar result follows if one considers the maximun deviation (rather than absolute deviation) of the regression function. The asymptotic distribution of the functional

$$\int \frac{\left\{\hat{r}_{n,p}(x) - r^{(p)}(x)\right\}^2 f(x)}{s(x)} dx \tag{5}$$

where $s(x) = \mathbb{E}(Y^2|X = x)$, is also evaluated under appropriate conditions as $n \rightarrow \infty$.

Now, we prepare some assumptions which are a restatement of assumptions of Mack and Müller (1989) in this setting. As in Mack and Müller (1989), let the density estimate appearing in (3) incorporate a different bandwidth sequence $\{b_n\}$.

Assumptions on the kernel K:

$(A1)$ The kernel K vanishes outside of $[-1, 1]$.

$(A2)$ $K \in \mathcal{C}^{(p+1)}[-1, 1]$, with $K^{(p)}(-1) = K^{(p)}(1) = 0$. Here $\mathcal{C}^{(p)}[-1, 1]$ denotes the set of all real-valued p times continuously differentiable functions.

$(A3)$ We assume the kernel is of order $k-p$ for some integer $k > p$.
Recall that a kernel K is of order $k-p$ means

$$\int x^j K(x)\, dx = \begin{Bmatrix} 1 & \text{if} & j = 0 \\ 0 & \text{if} & 0 < j < k-p \end{Bmatrix}$$

Assumptions on the marginal density f:

$(A4)$ f is continuous and positive on an open interval containing $[0, 1]$.

$(A5)$ f is $k-p$ continuously differentiable.

Assumptions on the regression function r:

$(A6)$ r is k continuously differentiable on an open interval containing $[0, 1]$.

Assumptions on the bandwiths (h_n and b_n):

($A7$) $b_n \to 0$, $h_n = n^{-\tau}$ with $\frac{1}{2(k-p)+1} < \tau < \frac{1}{2p+1}$.

($A8$) $nb_n^{2(k-p)} h_n \log n \to 0$, $b_n^{-1} h_n \log(1/b_n) \log n \to 0$.

($A8'$) $nb_n^{2(k-p)} h_n^{1/2} \to 0$, $b_n^{-1} h_n^{1/2} \log(1/b_n) \to 0$.

With these assumptions, b_n must converge to zero slower than h_n but still sufficiently fast for ($A8$) to hold. ($A7$) and ($A8$) hold for $b_n = n^{-\beta}$ for all $\beta > \frac{1-\tau}{2(k-p)}$. If ($A7$) holds and $\beta < \tau/2$, then ($A8'$) holds.

Whenever the marginal density is known we do not need to incorporate the bandwidth b_n, thus ($A8$) can be removed and ($A7$) becomes $h_n = n^{-\tau}$ with $\frac{1}{2k+1} < \tau < \frac{1}{2p+1}$.

Because the major difficulty in proving the asymptotic distribution of the maximal deviation from $r^{(p)}$ is the possible unboundedness of Y, we will need some additional hypotheses about Y as in Johnston (1982).

Assumptions on the observable Y:

($A9$) There exists $m > 2$ such that $\mathbb{E}|Y|^m < +\infty$ and $s(x)$ is bounded away from zero.

Let a_n be a sequence of real variable such that

($A10$) $a_n (nh_n)^{-1/2} (\log n)^{5/2} \to 0$, $a_n^{1-m} (nh_n)^{1/2} (\log n)^{1/2} \to 0$.

($A10'$) $a_n n^{-1/2} h_n^{-1} (\log n)^2 \to 0$, $a_n^{1-m} n^{1/2} h_n^{1/4} \to 0$.

($A11$) $(\log n) \sup_{x\in[0,1]} \int_{|y|\geq a_n} y^2 g(x,y)\, dy \to 0$.

($A11'$) $h_n^{-1} \sup_{x\in[0,1]} \int_{|y|\geq a_n} y^2 g(x,y)\, dy \to 0$.

($A12'$) We assume that $\left\{\int_{-a_n}^{a_n} y^2 g(x,y)\, dy\right\}^{1/2}$ (as function of x) has uniformly bounded and continous first derivatives on $[-1,1]$.

Observe that assumption ($A10$) is weaker than the corresponding assumption ($A2$) of Johnston (1982). To approximate the empirical distribution function of a uniformly distributed random variable, Johnston (1982) used a result of Révész (1976) while we use an improved result from Tusnády (1977). The asymptotic distributions for the quadratic norms of the normalized deviations of the estimates will be obtained by substituting ($A8'$), ($A10'$) and ($A11'$) (which are stronger) for ($A8$), ($A10$) and ($A11$) respectively.

Limit Theorems

The asymptotic distribution of the maximal (absolute) deviation is calculated with the aid of the following theorem.

Theorem 1 . *Suppose* (A1)-(A12) *hold and* $h_n = n^{-\tau}\,(0<\tau<1/2)$. *Define*

$$M_{n,p} = \left(nh_n^{2p+1}\right)^{1/2} \sup_{t\in[0,1]} \left|\{f(t)/s(t)\}^{1/2}\left(\hat{r}_{n,p}(t) - r^{(p)}(t)\right)\right|$$

and

$$\widetilde{M}_{n,p} = \left(nh_n^{2p+1}\right)^{1/2} \sup_{t\in[0,1]} \left[\{f(t)/s(t)\}^{1/2}\left\{\hat{r}_{n,p}(t) - r^{(p)}(t)\right\}\right].$$

Then

$$\mathbb{P}\left\{(2\tau\log n)^{1/2}\left[M_{n,p}/\left\{\lambda\left(K^{(p)}\right)\right\}^{1/2} - d_{n,p}\right] < x\right\} \to e^{-2e^{-x}} \tag{6}$$

and

$$\mathbb{P}\left\{(2\tau\log n)^{1/2}\left[\widetilde{M}_{n,p}/\left\{\lambda\left(K^{(p)}\right)\right\}^{1/2} - d_{n,p}\right] < x\right\} \to e^{-e^{-x}} \tag{7}$$

where $\lambda\left(K^{(p)}\right) = \int \left\{K^{(p)}(u)\right\}^2 du$ *and*

$$d_{n,p} = (2\tau\log n)^{1/2} + (2\tau\log n)^{-1/2}\log\left(\frac{\left\{2\lambda\left(K^{(p)}\right)\right\}^{-1}\lambda\left(K^{(p+1)}\right)}{2\pi}\right).$$

A similar theorem is given in Schlee (1982), but in this paper a different version of kernel estimator is used and only the asymptotic distribution of the supremum on a sequence of points is considered. Thus our results are stronger. Furtheremore, **theorem 1** improves the results of Johnston (1982). Indeed, we generalize his results (when the density function f is known) by giving the asymptotic distribution to the supremum of the derivative. On the other hand, **theorem 1** extends the results of Mack and Müller (1989).

The following result is for the quadratic functional (5).

Theorem 2 . *Under the assumptions of Theorem 1 and substituting* (A8′), (A10′) *and* (A11′) *respectively for* (A8), (A10) *and* (A11), *if* $\tau > 1/(2k+1/2)$ *then*

$$h_n^{-1/2}\left[\left(nh_n^{2p+1}\right)\int \frac{\left\{\hat{r}_{n,p}(x) - r^{(p)}(x)\right\}^2 f(x)}{s(x)}dx - \lambda\left(K^{(p)}\right)\right] \tag{8}$$

is asymptotically normally distributed with mean zero and variance

$$2\int\left\{\int K^{(p)}(x+y)\,K^{(p)}(x)\,dx\right\}^2 dy$$

as $n\to\infty$.

All proofs are given in Section 4.

3 Application to nonparametric tests

The statistical interest of **theorem 1** is twofold: not only this is a convenient way to obtain a confidence band for $r^{(p)}$, but it also provides a way for nonparametric tests of hypotheses including specification, significance, monotonicity and convexity. It would be hard to obtain an explicit confidence band from **theorem 2**, nevertheless, one can use it to construct nonparametric tests. In this section, we present consistent nonparametric tests. We prove these tests have asymptotic powers for some local alternatives. Note that for pratical applications, one would estimate both $s(x)$ and $f(x)$. This can be done by using kernel type estimates. Thus, we will suppose $s(x)$ and $f(x)$ are known.

3.1 Hypotheses and test statistics

Specification test:

A standard procedure to approximate r is to specify a parametric model. In this case, r is assumed to belong to a parametric family of known real functions (that is $\mathbb{E}(Y|X) = m(X,\theta)$ where $\theta \in \Theta$, is an unknown vector parameter). Thus, there are two competing models. Therefore a specification test may provide a way to prevent wrong conclusions.

The null hypothesis to be tested is that the parametric model is correct. That we can describe in the following form

$$H_\circ : \mathbb{P}\left[r(X_i) = m(X_i, \theta_\circ)\right] = 1 \text{ for some } \theta_\circ \in \Theta.$$

In what follows, we will assume that Θ is a subset of $\mathbb{R}^q$ and $m(x,\theta)$ is linear in θ. That is

$$m(x,\theta) = \theta^1 N_1(x) + \cdots + \theta^q N_q(x)$$

where $\theta = \left(\theta^1, \cdots, \theta^q\right)^t$ and $\{N_i(x)\}$ is a sequence of basis functions. Two important cases are the polynomial regression (with $N_i(x) = x^{i-1}$) and the spline regression (with $N_i(x)$ given by the basis splines depending on chosen knots and a given order).

The following corollary of **theorem 1** is of some interest.

Corollary 3 . *Let $\hat{\theta}$ denote an estimator consistent for $\theta_\circ$. Define*

$$M_{n,\hat{\theta}} = (nh_n)^{1/2} \sup_{t\in[0,1]} \left| \{f(t)/s(t)\}^{1/2} \left[\hat{r}_{n,0}(t) - m\left(t,\hat{\theta}\right)\right]\right|.$$

*Under the assumptions of **theorem 1** if the null hypothesis is true and if*

$$\sup_{t\in[0,1]} \left|m(t,\theta) - m\left(t,\hat{\theta}\right)\right| = o_p\left(\{n \log n\}^{-1/2}\right)$$

*then one can substitute $M_{n,\hat{\theta}}$ for $M_{n,0}$ in **theorem 1**.*

This corollary follows from **theorem 1** quite easily. Under some mild regularity conditions, such estimators as least squares, generalized method of moments or adaptive efficient weighted estimators satisfy the assumption required by this corollary.

As a consequence of this corollary, a consistent test against a nonparametric model may be based on the statistic

$$T_{n,\hat{\theta}} = (2\tau \log n)^{1/2} \left[M_{n,\hat{\theta}} / \{\lambda(K)\}^{1/2} - d_{n,0} \right].$$

Thus, the null hypothesis is rejected at level α if

$$T_{n,\hat{\theta}} \geq \log \left(-\frac{2}{\log(1-\alpha)} \right).$$

In the case when the regression function is linear (that is the case whenever the second derivative of the regression function is identical to zero), a consistent test against the nonparametric model can be based on the statistic $T_{n,2}$ given by

$$T_{n,2} = (2\tau \log n)^{1/2} \left[\sup_{t\in[0,1]} |\Upsilon_{n,2}(t)| / \left\{\lambda\left(K^{(2)}\right)\right\}^{1/2} - d_{n,2} \right]$$

where $\Upsilon_{n,2}(t) = (nh_n^5)^{1/2} \{f(t)/s(t)\}^{1/2} (\hat{r}_{n,2}(t))$ with the same cutoff point as $T_{n,\hat{\theta}}$. A similar test (of linearity) was used by Schlee (1980).

In the same spirit as above, and using now **theorem 2,** one can also construct a consistent test based on the statistic

$$T^1_{n,\hat{\theta}} = h_n^{-1/2} \left[(nh_n) \int \frac{\left\{\hat{r}_{n,0}(x) - m\left(x, \hat{\theta}\right)\right\}^2 f(x)}{s(x)} dx - \lambda(K) \right],$$

but we will need a stronger condition about $\hat{\theta}$. Namely, we assume that

$$\sup_{t\in[0,1]} \left| m(t,\theta) - m\left(t,\hat{\theta}\right) \right| = o_p\left(\{nh_n\}^{-1/2} \right). \quad (9)$$

It follows from **theorem 2** that as $n \to \infty, T^1_{n,\hat{\theta}}$ is asymptotically normally distributed with mean zero and variance

$$2 \int \left\{ \int K^{(p)}(x+y) K^{(p)}(x) dx \right\}^2 dy.$$

The statistic $T^1_{n,\hat{\theta}}$ is similar to Härdle and Mammen's test (cf. Härdle and Mammen 1993). Their test is based on a modification of the squared deviation between the Watson-Nadraya kernel estimator of the regression function and $m\left(t,\hat{\theta}\right)$.

Here, as in the case of linearity, a competitive statistic is

$$T_{n,2}^{1} = h_n^{-1/2}\left[\left(nh_n^5\right)\int\frac{\{\hat{r}_{n,2}(x)\}^2 f(x)}{s(x)}dx - \lambda\left(K^{(2)}\right)\right].$$

Significance test:

To test $H_0 : r = r_0$ against an unrestricted alternative, it is natural to use the statistic $T_{n,0}$ given by

$$T_{n,0} = (2\tau\log n)^{1/2}\left[\sup_{t\in[0,1]}|\Upsilon_{n,0}(t)|/\{\lambda(K)\}^{1/2} - d_{n,0}\right]$$

with $\Upsilon_{n,0}(t) = (nh_n)^{1/2}\{f(t)/s(t)\}^{1/2}(\hat{r}_{n,0}(t) - r_0(t))$ and reject the null hypothesis for large values. So the null hypothesis is rejected at level α if

$$T_{n,0} \geq \log\left(-\frac{2}{\log(1-\alpha)}\right).$$

The test statistic provided by **theorem 2** is defined by

$$T_{n,0}^{1} = h_n^{-1/2}\left[(nh_n)\int\frac{\{\hat{r}_{n,0}(x) - r_0\}^2 f(x)}{s(x)}dx - \lambda(K)\right].$$

Monotonicity test:

We consider testing whether the regression function is monotonically increasing. The regression function is nondecreasing if its first derivative is non-negative. Then the test statistic can be based on the greatest discrepancy between the estimate of the first derivative of the regression function and zero. Therefore, monotonicity is rejected at level α when

$$\widetilde{T}_{n,1} \geq \log\left(-\frac{1}{\log(1-\alpha)}\right)$$

where $\widetilde{T}_{n,1}$is given by

$$\widetilde{T}_{n,1} = (2\tau\log n)^{1/2}\left[\sup_{t\in[0,1]}\Upsilon_{n,1}(t)/\{\lambda(K')\}^{1/2} - d_{n,1}\right]$$

and

$$\Upsilon_{n,1}(t) = \left(nh_n^3\right)^{1/2}\{f(t)/s(t)\}^{1/2}(\hat{r}_{n,1}(t)).$$

Convexity test:

As in the monotone case, we reject the convexity of the regression function when

$$\widetilde{T}_{n,2} \geq \log\left(-\frac{1}{\log(1-\alpha)}\right)$$

with

$$\widetilde{T}_{n,2} = (2\tau \log n)^{1/2}\left[\sup_{t\in[0,1]} \Upsilon_{n,2}(t) / \left\{\lambda\left(K^{(2)}\right)\right\}^{1/2} - d_{n,2}\right].$$

Notice the statistic reflects the greatest discrepancy between the estimate of the second derivative of the regression function and zero. Similar statistics for monotonicity and convexity tests are obtained by Schlee (1980), but in a more restrictive setting since he considered the supremum of $\Upsilon_{n,p}(t)$ only on a sequence of points on $[0,1]$.

3.2 Asymptotic power

Here we focus our attention on the specification test, although similar results are valid for the other tests.
To make a local power calculation for the tests of the null hypotheses described above, we need to consider the behavior of the different statistics (calculated under a fixed but unknown point $r_0 = m(., \theta_\circ)$ of the null) for a sequence of alternatives of the form

$$r_n(x) = m(x, \theta_\circ) + \gamma_n \varphi(x),$$

where r_n satisfies $(A6)$, $\varphi(.)$ is a known function and γ_n is a sequence of real variables converging to zero.

Theorem 4 . (1) : *Let* $\gamma_n = n^{-1/2+\tau/2}\{2\tau \log n\}^{-1/2}$. *Under the assumptions of* ***corollary 3,***

$$\mathbb{P}\left(T_{n,\hat{\theta}} < x\right) \rightarrow \exp\left(-\psi(\varphi)\exp(-x)\right)$$

where

$$\psi(\varphi) = \int_0^1 \left\{e^{\varphi(t)/\{m(x,\theta_\circ)\lambda(K)\}^{1/2}} + e^{-\varphi(t)/\{m(x,\theta_\circ)\lambda(K)\}^{1/2}}\right\} dt.$$

(2) : *Under the assumptions of* ***theorem 2,*** *if* (9) *is true and if* $\gamma_n = n^{-1/2+\tau/4}$ *then* $T^1_{n,\hat{\theta}}$ *is asymptotically normally distributed with mean*

$$\int \frac{\varphi^2(x) f(x)}{s(x)} dx$$

and variance

$$2\int\left\{\int K^{(p)}(x+y)\,K^{(p)}(x)\,dx\right\}^2 dy.$$

(1) follows readily from theorem A1 of Bickel and Rosemblatt (1973) while (2) is obtained after staightforward calculations.

We see that $T^1_{n,\hat{\theta}}$ is more powerful than $T_{n,\hat{\theta}}$ for large samples, however $T_{n,\hat{\theta}}$ may well be preferable for moderate sample sizes and some alternatives. One can prove similar results for $T_{n,2}$ and $T^1_{n,2}$, but these tests are less powerful than $T_{n,\hat{\theta}}$ for large samples. For all these tests, it would be desirable to study their small sample behavior through Monte Carlo simulations.

4 Proofs

Obtaining the limit theorems 1 and 2 is a conceptually simple extension of Bickel and Rosenblatt 's (1973) results. Here, we adapt the proof of Johnston (1982) for regression estimate when the density function f is known. The major technical difficulty in adapting Johnston's (1982) proof in our case is the unknown f. However, we prove under the conditions assumed that is plausible to replace $\hat{f}_n$ by f in (2).

4.1 Proof of theorem 1

Proof. let $r_{n,p}$ defined by

$$r_{n,p}(x) = \left(nh_n^{p+1}\right)^{-1}\sum_{i=1}^{n} K^{(p)}\left(\frac{x-X_i}{h_n}\right) Y_i/f(X_i)$$

We also make the definitions

$$U_{n,p}(t) = \left(nh_n^{2p+1}\right)^{1/2}\{f(t)/s(t)\}^{1/2}\left(\hat{r}_{n,p}(t) - r^{(p)}(t)\right)$$

and

$$U_{1n,p}(t) = \left(nh_n^{2p+1}\right)^{1/2}\{f(t)/s(t)\}^{1/2}\left(r_{n,p}(t) - r^{(p)}(t)\right).$$

First of all, we prove that it is reasonable to replace $U_{n,p}$ by $U_{1n,p}$.

- **replacement cost:**

  Let $\|V\| = \sup\{|V(t)| : 0\leq t\leq 1\}$.

$$\|U_{n,p} - U_{1n,p}\| = \left(nh_n^{2p+1}\right)^{1/2} \sup_{0\le t\le 1} \left|\{f(t)/s(t)\}^{1/2} (\hat{r}_{n,p}(t) - r_{n,p}(t))\right|$$

Thus,

$$\|U_{n,p} - U_{1n,p}\| \le \left(nh_n^{2p+1}\right)^{1/2} \|f/s\|^{1/2} \|\hat{r}_{n,p} - r_{n,p}\|.$$

As in lemma 1 of Mack and Müller (1989) and under the conditions assumed, we have

$$\begin{aligned} \|U_{n,p} - U_{1n,p}\| &= \mathcal{O}_p\left((nh_n)^{1/2}\left\{\left[(nb_n)^{-1}\log 1/b_n\right]^{1/2} + b_n^{k-p}\right\}\right) \\ &= o_p\left(\{\log n\}^{-1/2}\right) \quad \text{by} \quad (A8). \end{aligned}$$

Now, as in Johnston (1982), we consider a truncated version of $U_{1n,p}$.

- **Truncated version of $U_{1n,p}(.)$**

Define

$$r_{a_n,p}(x) = \left(nh_n^{p+1}\right)^{-1} \sum_{i=1}^{n} K^{(p)}\left(\frac{x - X_i}{h_n}\right) V_i 1_{\{|V_i|\le a_n\}}$$

where $V_i = Y_i/f(X_i)$ and $1_{\{.\}}$ is the indicator on a set.

We also define

$$U_{2n,p}(t) = \left(nh_n^{2p+1}\right)^{1/2} \{f(t)/s(t)\}^{1/2} (r_{a_n,p}(t) - \mathbb{E}r_{a_n,p}(t)).$$

Thus

$$\left\|U_{1n,p} - U_{2n,p}\right\| \le \left(nh_n^{2p+1}\right)^{1/2} \|f/s\|^{1/2} \left\|r_{n,p} - r_{a_n,p} - \left(r^{(p)} - \mathbb{E}r_{a_n,p}\right)\right\|.$$

Then, using (2.12) and (3.1) of Mack and Müller (1989) and $(A6), (A10)$ we see that

$$\begin{aligned} \left\|U_{1n,p} - U_{2n,p}\right\| &= \mathcal{O}_p\left(\left\{nh_n^{2k+1}\right\}^{1/2} + a_n^{1-m}\{nh_n\}^{1/2}\right) \qquad (10) \\ &= o_p\left(\{\log n\}^{-1/2}\right). \end{aligned}$$

Next, we approximate $U_{2n,p}$ by an appropriate Brownian bridge. It is convenient to introduce the two-dimensional empirical process $Z_n(x,v)$ based on the sample $\{(X_i,V_i), i=1,...,n\}$ and given by

$$Z_n(x,v) = n^{1/2}\{F_n(x,v) - F(x,v)\},$$

with F and F_n the distribution function of (X,V) and the empirical distribution function of $\{(X_i,V_i), i=1,...,n\}$ respectively. Define Ψ to be the transformation of (X,V) to a uniform random variable on $[0,1]^2$, given by $\Psi(u,v) = (F_X(u), F_{V|X}(v))$, where F_X and $F_{V|X}$ are the distribution function of X and conditional distribution function of V given X respectively.

- **Approximation by Brownian bridge**

We can write

$$U_{2n,p}(t) = \{f(t)/s(t)\}^{1/2} h_n^{-1/2} \iint_{|v|\leq a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dZ_n(x,v)$$

Now, from formula (2.9) and (2.10) of Mack and Müller (1989) we have

$$\begin{aligned} U_{2n,p}(t) &= \{f(t)/s(t)\}^{1/2} h_n^{-1/2} \iint_{|v|\leq a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dB_n(\Psi(x,v)) \\ &\quad + \mathcal{O}_p\left(a_n\{nh_n\}^{-1/2}\{\log n\}^2\right), \end{aligned} \tag{11}$$

where B_n is the two-dimensional Brownian bridge given by $B_n(u,s) = W_n(u,s) - usW_n(1,1)$ and W_n is the two dimensional Wiener process.

Let us define

$$U_{3n,p}(t) = \{f(t)/s(t)\}^{1/2} h_n^{-1/2} \iint_{|v|\leq a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dB_n(\Psi(x,v)).$$

By $(A10)$ we see that $(\log n)^{1/2}\|U_{n,p}\|$ and $(\log n)^{1/2}\|U_{3n,p}\|$have the same distribution. Define now,

$$s_n(t) = \mathbb{E}\left(Y^2 1_{\{|Y|\leq a_n\}}|X=t\right)$$

and

$$U_{4n,p}(t) = \{s_n(t)/s(t)\}^{-1/2} U_{3n,p}(t).$$

Under the conditions assumed, one can prove as in lemma A.4 of Johnston (1982), that

$$\|U_{3n,p} - U_{4n,p}\| = o_p\left(\{\log n\}^{-1/2}\right). \tag{12}$$

On the other hand,

$$U_{4n,p}(t) = \{f(t)/s_n(t)\}^{1/2} h_n^{-1/2} \iint_{|v|\le a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dB_n(\Psi(x,v)).$$

This latter formula can be rewritten in the following form

$$U_{4n,p}(t) = \{\vartheta_n(t) f(t)\}^{-1/2} h_n^{-1/2} \iint_{|v|\le a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dB_n(\Psi(x,v))$$

where $\vartheta_n(t) = \mathbb{E}\left(V^2 1_{\{|V|\le a_n\}} | X = t\right)$.
Define

$$U_{5n,p}(t) = \{\vartheta_n(t) f(t)\}^{-1/2} h_n^{-1/2} \iint_{|v|\le a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dW_n(\Psi(x,v))$$

Using the results of Tusnády (1977), $(A10)$ and integration by parts (see formula $(A.1.1)$ of Johnston 1982), one can prove

$$\begin{aligned} \left\|U_{5n,p} - U_{4n,p}\right\| &= \mathcal{O}_p\left(a_n \{nh_n\}^{-1/2} \{\log n\}^2\right) \\ &= o_p\left(\{\log n\}^{-1/2}\right). \qquad (13) \end{aligned}$$

Now, it is easy to see that $U_{5n,p}$ has the same covariance function as $U_{6n,p}$ defined by

$$U_{6n,p}(t) = \{\vartheta_n(t) f(t) h_n\}^{-1/2} \int \{\vartheta_n(x) f(x)\}^{1/2} K^{(p)}\left(\frac{t-x}{h_n}\right) dW(x)$$

where W is Brownian motion on $(-\infty, +\infty)$. Moreover, since $U_{5n,p}$ and $U_{6n,p}$ are both Gaussian, then they have the same distribution.
Now, define

$$U_{7n,p}(t) = h_n^{-1/2} \int K^{(p)}\left(\frac{t-x}{h_n}\right) dW(x).$$

We obtain, as Johnston (1982) (see lemma A.6 therein), that

$$\left\|U_{6n,p} - U_{7n,p}\right\| = \mathcal{O}_p\left(h_n^{1/2}\right) = o_p\left(\{\log n\}^{-1/2}\right).$$

Thus, to prove the theorem we only need to determine the asymptotic distribution of the (absolute) supremum $V_{a_n,2p}^{\mathbb{E}}$ which is given in the following lemma. This lemma is easily derived from Bickel and Rosemblatt (1973).

Lemma 5 *(Bickel and Rosemblatt 1973). Let d_n and $\lambda\left(K^{(p)}\right)$ be as in* ***theorem 1*** *and let $h_n = n^{-\tau}$ $(0 < \tau < 1/2)$. Define*

$$\varrho_{n,p}(t) = h_n^{-1/2} \int K^{(p)} \left(\frac{t-x}{h_n} \right) dW_n(x).$$

Then

$$\mathbb{P}\left\{ (2\tau \log n)^{1/2} \left[\sup_{t\in[0,1]} \left|\varrho_{n,p}(t)\right| / \left\{\lambda\left(K^{(p)}\right)\right\}^{1/2} - d_{n,p} \right] < x \right\} \longrightarrow e^{-2e^{-x}}$$

and

$$\mathbb{P}\left\{ (2\tau \log n)^{1/2} \left[\sup_{t\in[0,1]} \varrho_{n,p}(t) / \left\{\lambda\left(K^{(p)}\right)\right\}^{1/2} - d_{n,p} \right] < x \right\} \longrightarrow e^{-e^{-x}}$$

with $d_{n,p}$ as in ***theorem 1.***

■

4.2 Proof of theorem 2

Proof. The following lemma is an immediate consequence of theorem 4.1 of Bickel and Rosemblatt (1973).

Lemma 6 *Let define ζ_n by*

$$\zeta_n(t) = h_n^{-1/2} \int \{\vartheta(x) f(x)\}^{1/2} K^{(p)} \left(\frac{t-x}{h_n} \right) dW(x)$$

with $\vartheta(t) = \mathbb{E}(V^2|X=t)$ and $V = Y/f(X)$. We assume that ω is an integrable piecewise continuous and bounded function. Under the assumptions of ***theorem 2,***

$$h_n^{-1/2} \quad \int \{\zeta_n(t)\}^2 \omega(t)\, dt - \lambda\left(K^{(p)}\right) \int \vartheta(t) f(t) \omega(t)\, dt \bigg]$$

is asymptotically normally distributed with mean zero and variance

$$\left[2 \int \left\{ \int K^{(p)}(x+y) K^{(p)}(x)\, dx \right\}^2 dy \right] \int \{\vartheta(t) f(t) \omega(t)\}^2 dt$$

as $n \longrightarrow \infty$.

Define

$$\zeta_{1n}(t) = h_n^{-1/2} \int \{\vartheta_n(x) f(x)\}^{1/2} K^{(p)}\left(\frac{t-x}{h_n}\right) dW(x).$$

Lemma 6 is valid for ζ_{1n}. Besides, one can set $\omega(x) = \{\vartheta(x) f(x)\}^{-1}$. Now using $(A11')$ and applying **lemma 6**, straightforward calculations show

$$h_n^{-1/2}\left[\int \frac{\{\zeta_{1n}(t)\}^2 f(t)}{s(t)} dt - \lambda\left(K^{(p)}\right)\right] \tag{14}$$

is asymptotically normally distributed with mean zero and variance

$$2\int\left\{\int K^{(p)}(x+y) K^{(p)}(x)\, dx\right\}^2 dy$$

as $n \to \infty$. Moreover, we have noticed that ζ_{1n} has the same distribution as ζ_{2n} defined by

$$\zeta_{2n}(t) = h_n^{-1/2} \iint_{|v| \leq a_n} v K^{(p)}\left(\frac{t-x}{h_n}\right) dW_n(\Psi(x,v)).$$

So ζ_{2n} can be substituted for ζ_{1n} in (14). Define

$$\zeta_{3n}(t) = h_n^{-1/2} \iint_{|v| \leq a_n} v K^{(p)}\left(\frac{t-x}{h_n}\right) dB_n(\Psi(x,v)).$$

To prove that one can substitute ζ_{3n} for ζ_{1n} in (14) ,we will apply the following remark from Bickel and Rosemblatt (1973).

Remark 7 *if $\{g_n\}$ is a sequence of functionals on $D[0,1]$ (cf.Billingsley 1968) satisfying Lipschitz conditions*

$$|g_n(x) - g_n(y)| \leq \gamma_n \|x - y\|$$

and A_n, B_n are stochastic processes realizable in D such that $\|A_n - B_n\| = o_p(1/\gamma_n)$, then $g_n(A_n)$ converges in law if and only if $g_n(B_n)$ does, and to the same limit.

We will apply this proposition to the functional

$$h_n^{-1/2} \quad \int \{\zeta_n(t)\}^2 \omega(t)\, dt - \lambda\left(K^{(p)}\right)\Big]$$

with $\gamma_n = h_n^{-1/2}$.

As in (13) and using $(A10')$, we see that

$$\|\zeta_{3n} - \zeta_{2n}\| = o_p\left(h_n^{1/2}\right).$$

Using (12) and $(A10')$ one can prove in a similar way

$$\|\zeta_{3n} - \zeta_{4n}\| = o_p\left(h_n^{1/2}\right),$$

where ζ_{4n} is defined by

$$\zeta_{4n}(t) = h_n^{-1/2} \iint_{|v|\leq a_n} vK^{(p)}\left(\frac{t-x}{h_n}\right) dZ_n(x,v).$$

We finish by applying the same rule and using respectively (11) and (10) with the assumptions $(A10')$ and $(A8')$. ■

References

[1] P.J. Bickel and M. Rosenblatt. On some global measures of deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095, 1973.

[2] H.J. Bierens. Consistent model specification tests. *Journal of Econometrics*, 20:105–134, 1982.

[3] H.J. Bierens. A consistent conditional moment test of functional form. *Econometrica*, 58:1443–58, 1990.

[4] P. Billingsley. *Convergence of probabilty measures.* Wiley, New York, 1968.

[5] M.C. Bowman, A.W. Jones and I. Gijbels. Testing monotonicity of regression. *Journal of computational and Graphical Statistics*, 7(4):489–500, 1998.

[6] C.A.T. Diack. Sur la convergence des tests de schlee et yatchew. *submitted*, 1998.

[7] C.A.T. Diack. A consistent nonparametric test of the convexity of regression based on least squares splines. *Technical Report of EURANDOM*, 001-99(submitted), 1999.

[8] C.A.T. Diack. Two consistent nonparmetric tests of the monotonicity of regression. *Technical Report of EURANDOM*, 003-99(Submitted), 1999.

[9] C.A.T. Diack and C. Thomas-Agnan. A nonparametric test of the non-convexity of regression. *Journal of Nonparametric Statistics*, 49(4):335–362, 1998.

[10] A. Doveh, E. Shapiro and P.D. Feigin. Testing of monotonicity in regression models. *Prepints*, 1999.

[11] R.L. Eubank and C.H. Spiegelman. Testing the goodness of fit of a linear model via nonparametric regression techniques. *American Statistical Association JASA, Theory and methods*, 85(410):387–392, June 1990.

[12] W. Hardle and E. Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.

[13] J. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–72, 1978.

[14] Y Hong and H. White. Consistent specification testing via nonparametric series regression. *Econometrica*, 63:1133–1160, 1995.

[15] G.J. Johnston. Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of multivariate analysis*, 12:402–414, 1982.

[16] B.J. Lee. *A nonparametric model specification test using a kernel regression method.* Ph.D. dissertation, University of Wisconsin-Madison, 1988.

[17] Y.P. Mack and H.G. Muller. Derivative estimation in nonparametric regression with random predictor variable. *Sankhya: The Indian Journal of Statistics*, 51, Serie A, Pt. 1:59–72, 1989.

[18] E. Nadaraya. On estimating regression. *Theory of Proba. and Appl.*, 9:141–142, 1964.

[19] P. Revesz. On strong approximation of the multidimensional empirical process. *The Annals of Probability*, 4(5):729–743, 1976.

[20] W. Schlee. Nonparametric test of the monotony and convexity of regression. *Nonparametric Statistical Inference*, 2:823–836, 1980.

[21] T. Stoker. Tests of additive derivative constraints. *Review of Economic Studies*, 56:535–552, 1989.

[22] T. Stoker. *Lectures on semiparametric econometrics.* Core Foundation Louvain-La-Neuve, 1991.

[23] G. Tusnady. A remark on the approximation of the sample d.f. in the multidimensional case. *Period. Math. Hungar.*, 8:53–55, 1977.

[24] G.S. Watson. Smooth regression analysis. *Sankhya, Ser. A Math. Sci.*, 26:359–372, 1964.

[25] J.M. Wooldridge. A test for functional form against nonparametric alternatives. *Econometrics Theory*, 8:452–475, 1992.

[26] A. Yatchew and L. Bos. Nonparametric regression and testing in economic models. *J. of Quantitative Economics*, 13:81–131, 1997.

[27] A.J. Yatchew. Nonparametric regression tests based on least squares. *Econometrics Theory*, 8:435–451, 1992.

[28] J.X. Zheng. A consistent test of functional form via nonparametric estimation techniques. *manuscrit*, 1996.