GEORGIAN MATHEMATICAL JOURNAL: Vol. 6, No. 6, 1999, 501-516

LIMIT DISTRIBUTION OF THE MEAN SQUARE DEVIATION OF THE GASSER–MÜLLER NONPARAMETRIC ESTIMATE OF THE REGRESSION FUNCTION

R. ABSAVA AND E. NADARAYA

Abstract. Asymptotic distribution of the mean square deviation of the Gasser–Müller estimate of the regression curve is investigated. The testing of hypotheses on regression function is considered. The asymptotic behaviour of the power of the proposed criteria under contiguous alternatives is studed.

Let $\{Y_{nk},\ k=1,\dots,n\}_{n\geq 1}$ be a sequence of arrays of random variables defined as follows:

$$Y_{nk}=g(x_{nk})+Z_{nk},\quad x_{nk}=\frac{k}{n},\quad k=1,\dots,n,\quad n\geq 1,$$

where $g(x), x\in[0,1]$, is the unknown real-valued function to be estimated by given observations Y_{nk}; Z_{nk}, $k=1,\dots,n$, is a sequence of arrays of independent random variables identically distributed in each array such that $EZ_{nk}=0$, $EZ_{nk}^2=\sigma^2$, $k=1,\dots,n$, $n\geq 1$.

Let us consider the estimate of the function $g(x)$ from [1]:

$$g_n(x)=\sum_{i=1}^n W_{ni}(x)Y_{ni},\quad x\in[0,1],$$

where

$$W_{ni}=b_n^{-1}\int_{\Delta_i}K\Big(\frac{x-t}{b_n}\Big)dt,\quad \Delta_i=\Big[\frac{i-1}{n},\frac{i}{n}\Big].$$

Here $K(x)$ is some kernel, b_n is a sequence of positive numbers tending to 0.

Assume that the kernel $K(x)$ has a compact support and satisfies the conditions:

1991 *Mathematics Subject Classification.* 62G07.

Key words and phrases. Limit distribution, consistency, power of criteria, contiguous alternatives.

1°. $\operatorname{supp}(K) \subset [-\tau, \tau]$, $0 < \tau < \infty$, $\sup |K| < \infty$, $\int K(u)du = 1$, $K(-x) = K(x)$,
2°. $\int K(u)u^j du = 0$, $0 < j < s$, $\int K(u)u^s du \neq 0$,
3°. $K(x)$ has a bounded derivative in $R = (-\infty, \infty)$.

Denote by F_s the family of regression functions $g(x)$, $x \in [0,1]$, having derivatives of order up to s ($s \geq 2$), $g^{(s)}(x)$ is continuous. Note that $\sum_{i=1}^n W_{ni}(x) = 1$ for $x \in \Omega_n(\tau) = [\tau b_n, 1-\tau b_n]$ and $Eg_n(x) = g(x) + O(b_n^2)$ if the kernel $K(x)$ satisfies condition 1° and $g(x) \in F_2$. On the other hand, $\sum_{i=1}^n W_{ni}(x) \neq 1$ for $x \in [0, \tau b_n) \cup (1-\tau b_n, 1]$ and it may happen that $Eg_n(x) \not\to g(x)$, for example, $Eg_n(0) \to \frac{g(0)}{2}$ (or $Eg_n(1) \to \frac{g(1)}{2}$). If the estimate $g_n(x)$ is divided by $\sum_{i=1}^n W_{ni}(x)$, then the proposed estimate $\widetilde{g}_n(x)$ becomes asymptotically unbiased and, moreover, $E\widetilde{g}_n(x) = g(x) + O(b_n)$ for $x \in [0, \tau b_n) \cup (1-\tau b_n, 1]$. Hence the asymptotic behaviour of the estimate $g_n(x)$ near the boundary of the interval [0, 1] is worse than within the interval $\Omega_n(\tau)$. A phenomenon of such kind is known in the literature as the boundary effect of the estimator $g_n(x)$ (see, for examle, [2]). It would be interesting to investigate the limit behaviour of the distribution of the mean square deviation of $g_n(x)$ from $g(x)$ on the interval $\Omega_n(\tau)$ and this is the aim of the present article.

The method of proving the statements given below is based on the functional limit theorems for semimartingales from [3].

We will use the notation:

$$U_n = nb_n \int_{\Omega_n(\tau)} (g_n(x) - Eg_n(x))^2 dx, \quad \sigma_n^2 = 4\sigma^4 (nb_n)^2 \sum_{k=2}^{n} \sum_{i=1}^{k-1} Q_{ik}^2(n),$$

$$Q_{ij} \equiv Q_{ij}(n) = \int_{\Omega_n(\tau)} W_{ni}(x) W_{nj}(x) dx,$$

$$\eta_{ik} \equiv \eta_{ik}(n) = 2nb_n Q_{ik} Z_{ni} Z_{nk} \sigma_n^{-1},$$

$$\xi_{nk} = \sum_{i=1}^{k-1} \eta_{ik}, \quad k = 2, \dots, n, \quad \xi_{n1} = 0, \quad \xi_{nk} = 0, \ \ k > n,$$

$$\mathcal{F}_k^{(n)} = \sigma(\omega : Z_{n1}, Z_{n2}, \dots, Z_{nk}),$$

where $\mathcal{F}_k^{(n)}$ is the σ-algebra generated by the random variables $Z_{n1}, Z_{n2}, \dots, Z_{nk}$, $\mathcal{F}_0^{(n)} = (\varnothing, \Omega)$.

Lemma 1 ([4], p. 179). *The stochastic sequence* $(\xi_{nk}, \mathcal{F}_k^{(n)})_{k \geq 1}$ *is a martingale-difference.*

Lemma 2. *Suppose the kernel* $K(x)$ *satisfies conditions* 1° *and* 3°. *If*

$nb_n^2 \to \infty$, *then*

$$b_n^{-1}\sigma_n^2 \to 2\sigma^4 \int_{-2\tau}^{2\tau} K_0^2(u)\,du, \tag{1}$$

and

$$EU_n = \sigma^2 \int K^2(u)du + O(b_n) + O\Big(\frac{1}{nb_n}\Big), \tag{2}$$

where $K_0 = K * K$ (*the symbol* $*$ *denotes the convolution*).

Proof. It is not difficult to see that

$$\sigma_n^2 = 2\sigma^4(nb_n)^2\Big[\sum_{k=1}^n\sum_{i=1}^n Q_{ki}^2 - \sum_{i=1}^n Q_{ii}^2\Big] = d_1(n) + d_2(n). \tag{3}$$

Here by the definition of Q_{ki} we have

$$d_1(n) = 2\sigma^4 n^2 b_n^{-2}\sum_{i,j}\Big(\int_{\Delta_i}\int_{\Delta_j}\Psi_n(t_1,t_2)dt_1dt_2\Big)^2,$$

where

$$\Psi_n(t_1,t_2) = \int_{\Omega_n(\tau)} K\Big(\frac{x-t_1}{b_n}\Big)K\Big(\frac{x-t_2}{b_n}\Big)\,dx.$$

Using the mean value theorem, we can rewrite $d_1(n)$ as

$$d_1(n) = 2\sigma^4(nb_n)^{-2}\sum_{i,j}\Psi_n^2(\theta_i^{(1)},\theta_j^{(2)}),$$

with $\theta_i^{(1)} \in \Delta_i$, $\theta_i^{(2)} \in \Delta_j$.

Let $P(x)$ be the uniform distribution function on $[0,1]$ and $P_n^{(i)}(x)$ the empirical distribution function of the "sample" $\theta_1^{(i)},\dots,\theta_n^{(i)}$, $i = 1, 2$, i.e., $P_n^{(i)}(x) = n^{-1}\sum_{k=1}^n I_{(-\infty,x)}(\theta_k^{(i)})$, where $I_A(\cdot)$ is the indicator of the set A. Then $d_1(n)$ can be written as the integral

$$d_1(n) = 2\sigma^4 b_n^{-2}\int_0^1\int_0^1 \Psi_n^2(x,y)dP_n^{(1)}(x)dP_n^{(2)}(y). \tag{4}$$

It is easy to check that

$$\Big|\int_0^1\int_0^1 \Psi_n^2(x,y)dP_n^{(1)}(x)dP_n^{(2)}(y) - \int_0^1\int_0^1 \Psi_n^2(x,y)dP(x)dP(y)\Big| \le I_1 + I_2,$$

where

$$I_1 = \Big|\int_0^1\int_0^1 \Psi_n^2(x,y)dP_n^{(2)}(y)[dP_n^{(1)}(x) - dP(x)]\Big|,$$

$$I_2 = \Big| \int_0^1 \int_0^1 \Psi_n^2(x,y) dP(x) [dP_n^{(2)}(y) - dP(y)] \Big|.$$

Furthermore, the integration by parts of the internal integral in I_1 yields

$$I_1 \le 2 \int_0^1 dP_n^{(2)}(y) \int_0^1 \Big(|P_n^{(1)}(x) - P(x)|\, |\Psi_n(x,y)| b_n^{-1} \times \\ \times \int_{\Omega_n(\tau)} \Big| K^{(1)}\Big(\frac{t-x}{b_n}\Big)\Big| \Big| K\Big(\frac{t-y}{b_n}\Big)\Big| dt \Big) dx. \tag{5}$$

Since $\sup\limits_{0\le x\le 1} |P_n^{(i)} - P(x)| \le \frac{2}{n}$ and $|\Psi_n(x,y)| \le c b_n$, we have

$$I_1 = O(b_n/n).$$

Here and in what follows c is the positive constant varying from one formula to another.

In the same manner we may show that

$$I_2 = O(b_n/n).$$

Therefore

$$d_1(n) = 2\sigma^4 b_n^{-2} \int_0^1 \int_0^1 \Psi_n^2(t_1,t_2) dt_1 dt_2 + O\Big(\frac{1}{nb_n}\Big). \tag{6}$$

It is easy to show also that

$$d_1(n) = 2\sigma^4 \int_{\Omega_n(\tau)} \int_{\Omega_n(\tau)} \Big(\int_{(x-1)b_n^{-1}}^{x b_n^{-1}} K(u) K\Big(\frac{x-y}{b_n} - u\Big) du \Big)^2 dx dy + O\Big(\frac{1}{nb_n}\Big).$$

Since $[-\tau, \tau] \subset [\frac{x-1}{b_n}, \frac{x}{b_n}]$ for all $x \in \Omega_n(\tau)$, we have

$$d_1(n) = 2\sigma^4 \int_{\Omega_n(\tau)} \int_{\Omega_n(\tau)} K_0^2\Big(\frac{x-y}{b_n}\Big) dx\, dy + O\Big(\frac{1}{nb_n}\Big).$$

But it is easy to see that

$$b_n^{-1} d_1(n) = 2\sigma^4 \int_0^1 \Big(\int_{\tau - y/b_n}^{(1-y)/b_n - \tau} K_0^2(u)\, du \Big) dy + O(b_n) + O\Big(\frac{1}{nb_n}\Big).$$

Therefore

$$b_n^{-1} d_1(n) \to 2\sigma^4 \int K_0^2(u)\, du. \tag{7}$$

We can directly verify that

$$b_n^{-1} |d_2(n)| \le n^2 b_n^{-3} \sum_{i=1}^n \int_{\Delta_i} \int_{\Delta_i} \int_{\Delta_i} \int_{\Delta_i} \Big(\int_{\Omega_n(\tau)} \Big| K\Big(\frac{x-t_1}{b_n}\Big) K\Big(\frac{x-t_2}{b_n}\Big)\Big| dx \times$$

$$\times \int_{\Omega_n(\tau)} \Big| K\Big(\frac{y-t_3}{b_n}\Big) K\Big(\frac{y-t_4}{b_n}\Big)\Big| dy \Big) dt_1\, dt_2\, dt_3\, dt_4 \le c\frac{1}{nb_n}. \tag{8}$$

Statement (1) immediately follows directly from (3), (7), and (8).

Let us now show that (2) holds. We have

$$Dg_n(x) = \frac{\sigma^2}{nb_n^2} \int_0^1 K^2\Big(\frac{x-t}{b_n}\Big)\, dP_n(x),$$

where

$$P_n(x) = n^{-1} \sum_{k=1}^n I_{(-\infty,x)}(\theta_k), \quad \theta_k \in \Delta_k, \quad k = 1, \ldots, n.$$

Applying the same argument as in deriving (6), we find

$$Dg_n(x) = \frac{\sigma^2}{nb_n^2} \int_0^1 K^2\Big(\frac{x-t}{b_n}\Big)\, dt + O\Big(\frac{1}{(nb_n)^2}\Big). \tag{9}$$

Furthermore, taking into account that $[-\tau, \tau] \subset [\frac{x-1}{b_n}, \frac{x}{b_n}]$ for all $x \in \Omega_n(\tau)$ by (9) we can write

$$Dg_n(x) = \frac{\sigma^2}{nb_n} \int K^2(u)\, du + O((nb_n)^{-2}).$$

Thus

$$EU_n = \sigma^2 \int K^2(u)\, du + O(b_n) + O\Big(\frac{1}{nb_n}\Big). \quad \square$$

Theorem 1. *Suppose the kernel $K(x)$ satisfies conditions $1°$ and $3°$. If $\sup_{n\ge1} EZ_{n1}^4 < \infty$ and $nb_n^2 \to \infty$, then*

$$b_n^{-1/2}(U_n - \sigma^2\theta_1)\sigma^{-2}\theta_2^{-1} \xrightarrow{d} \xi,$$

where

$$\theta_1 = \int K^2(u)\, du, \quad \theta_2^2 = 2\int K_0^2(u)\, du.$$

By the symbol $\xrightarrow{d}$ we denote the convergence in distribution, ξ is a random variable distributed according to the standard normal distribution.

Proof. Note that

$$\frac{U_n - EU_n}{\sigma_n} = H_n^{(1)} + H_n^{(2)},$$

where

$$H_n^{(1)} = \sum_{k=1}^n \xi_{nk}, \quad H_n^{(2)} = \frac{nb_n}{\sigma_n} \sum_{i=1}^n (Z_{ni}^2 - EZ_{ni}^2) Q_{ii}.$$

$H_n^{(2)}$ converges to zero in probability. Indeed,

$$DH_n^{(2)} = \frac{(nb_n)^2}{\sigma_n^2} \sum_{i=1}^n EZ_{ni}^4 Q_{ii}^2 \le \sup_{n\ge1} EZ_{n1}^4 \frac{(nb_n)^2}{\sigma_n^2} \sum_{i=1}^n Q_{ii}^2 = \sup_{n\ge1} EZ_{n1}^4 \frac{|d_2(n)|}{\sigma_n^2}.$$

This and (8) yield $DH_n^{(2)} = O(1/(n\sigma_n^2)) = O(1/(nb_n))$. Hence $H_n^{(2)} \xrightarrow{P} 0$. (By the symbol $\xrightarrow{P}$ we denote the convergence in probability).

Now let us prove the convergence $H_n^{(1)} \xrightarrow{d} \xi$. For this it is sufficient verify the validity of Corollaries 2 and 6 of Theorem 2 from [3]. We will show that the asymptotic normality conditions from the corresponding statement are fulfilled by the sequence $\{\xi_{nk}, \mathcal{F}_k^{(n)}\}_{k\ge1}$ which is a square-integrable martingale-difference according to Lemma 1.

A direct calculation shows that $\sum_{k=1}^n E\xi_{nk}^2 = 1$. Now the asymptotic normality will take place if for $n \to \infty$ we have

$$\sum_{k=1}^n E[\xi_{nk}^2 I(|\xi_{nk}| \ge \epsilon)\ \mathcal{F}_{k-1}^{(n)}] \xrightarrow{P} 0 \tag{10}$$

and

$$\sum_{k=1}^n \xi_{nk}^2 \xrightarrow{P} 1. \tag{11}$$

But, as shown in [3], the validity of (11) together with the condition $\sup\limits_{1\le k\le n} |\xi_{nk}| \xrightarrow{P} 0$ implies the statement (10) as well.

Since for $\epsilon > 0$

$$P\Big\{ \sup_{1\le k\le n} |\xi_{nk}| \ge \epsilon \Big\} \le \epsilon^{-4} \sum_{k=1}^n E\xi_{nk}^4,$$

to prove $H_n^{(1)} \xrightarrow{d} \xi$ we have to verify only (11), since relation (12) given below is fulfiled.

Now we will establish $\sum_{k=1}^n \xi_{nk}^2 \xrightarrow{P} 1$. It is sufficient to verify that

$$E\Big(\sum_{k=1}^n \xi_{nk}^2 - 1 \Big)^2 \to 0$$

as $n \to \infty$, i.e., due to $\sum_{k=1}^n E\xi_{nk}^2 = 1$

$$E\Big(\sum_{k=1}^n \xi_{nk}^2 \Big)^2 = \sum_{k=1}^n E\xi_{nk}^4 + 2 \sum_{1\le k_1<k_2\le n} E\xi_{nk_1}^2 \xi_{nk_2}^2 \to 1.$$

First show that $\sum_{k=1}^{n} E\xi_{nk}^4 \to 0$ as $n \to \infty$. By virtue of the definitions of ξ_{nk} and η_{ij} we can write

$$\sum_{k=1}^{n} E\xi_{nk}^4 = I_n^{(1)} + I_n^{(2)},$$

where

$$I_n^{(1)} = \frac{16(nb_n)^4}{\sigma_n^4} \sum_{k=2}^{n} EZ_{nk}^4 \sum_{j=1}^{k-1} EZ_{nj}^4 Q_{jk},$$
$$I_n^{(2)} = \frac{48(nb_n)^4}{\sigma_n^4} \sum_{k=2}^{n} \sum_{i \neq j}^{k-1} EZ_{ni}^2 EZ_{nj}^2 Q_{ik}^2 Q_{jk}^2.$$

On the other hand, since $\sup_{n \geq 1} EZ_{n1}^4 < \infty$, $|Q_{ij}| \leq cb_n^{-1}n^{-2}$ and $b_n^{-1}\sigma_n^2 \to \sigma^4\theta_2^2$, we have $I_n^{(1)} = O((nb_n)^{-2})$, $I_n^{(2)} = O(1/(n\sigma_n^4)) = O(1/(nb_n^2))$. Hence

$$\sum_{k=1}^{n} E\xi_{nk}^4 \to 0, \qquad n \to \infty. \tag{12}$$

Now let us establish that

$$2 \sum_{1 \leq k_1 < k_2 \leq n} E\xi_{nk_1}^2 \xi_{nk_2}^2 \to 1$$

as $n \to \infty$. The definition of ξ_{ni} yields

$$\xi_{nk_1}^2 \cdot \xi_{nk_2}^2 = B_{k_1k_2}^{(1)} + B_{k_1k_2}^{(2)} + B_{k_1k_2}^{(3)} + B_{k_1k_2}^{(4)},$$

where

$$B_{k_1k_2}^{(1)} = \sigma_2(k_1)\sigma_2(k_2), B_{k_1k_2}^{(2)} = \sigma_2(k_1)\sigma_1(k_2),$$
$$B_{k_1k_2}^{(3)} = \sigma_1(k_1)\sigma_2(k_2), B_{k_1k_2}^{(4)} = \sigma_1(k_1)\sigma_1(k_2),$$
$$\sigma_1(k) = \sum_{i \neq j} \eta_{ik}\eta_{jk}, \qquad \sigma_2(k) = \sum_{i=1}^{k-1} \eta_{ik}^2.$$

Therefore

$$2 \sum_{1 \leq k_1 < k_2 \leq n} E\xi_{nk_1}^2 \xi_{nk_2}^2 = \sum_{i=1}^{4} A_n^{(i)},$$

where

$$A_n^{(i)} = 2 \sum_{1 \leq k_1 < k_2 \leq n} EB_{k_1k_2}^{(i)}, \quad i = 1, 2, 3, 4.$$

Let us consider the term $A_n^{(3)}$. According to the definition of η_{ij} we have

$$E\eta_{ik_2}^2\eta_{sk_1}\eta_{tk_1} = 0, \quad s \neq t, \ \ k_1 < k_2.$$

Thus

$$A_n^{(4)} = 0. \tag{13}$$

To estimate the term $A_n^{(2)}$, note that $\sup\limits_{n\geq 1} EZ_{n1}^4 < \infty$ and $|Q_{ij}| \leq cb_n^{-1}n^{-2}$. So we obtain

$$|A_n^{(2)}| \leq c\frac{nb_n^2}{\sigma_n^4}\sum_{k_1=2}^{n}\sum_{i=1}^{k_1-1} Q_{ik_1}^2 \leq c\frac{1}{n\sigma_n^2} = O\Big(\frac{1}{nb_n}\Big). \tag{14}$$

Now we will verify that $A_n^{(1)} \to 1$ as $n \to \infty$. For this let us represent $A_n^{(1)}$ in the form

$$A_n^{(1)} = D_n^{(1)} + D_n^{(2)},$$

where

$$D_n^{(1)} = 2\sum_{1\leq k_1<k_2\leq n}\Big(\sum_{i=1}^{k_1-1} E\eta_{ik_1}^2\Big)\Big(\sum_{j=1}^{k_2-1} E\eta_{jk_2}^2\Big),$$

$$D_n^{(2)} = 2\Big(\sum_{k_1<k_2} B_{k_1k_2}^{(1)} - \sum_{k_1<k_2}\Big(\sum_{i=1}^{k_1-1} E\eta_{ik_1}^2\Big)\Big(\sum_{j=1}^{k_2-1} E\eta_{jk_2}^2\Big)\Big).$$

From the definition of σ_n^2 it follows that

$$D_n^{(1)} = 1 - \sum_{k=2}^{n}\Big(\sum_{i=1}^{k-1} E\eta_{ik}^2\Big)^2,$$

where

$$\sum_{k=2}^{n}\Big(\sum_{i=1}^{k-1} E\eta_{ik}^2\Big)^2 \leq c\Big(\frac{nb_n}{\sigma_n}\Big)^4\sum_{k=2}^{n}\Big(\sum_{i=1}^{k-1} Q_{ik}^2\Big)^2 \leq c\frac{1}{n\sigma_n^4} = O\Big(\frac{1}{nb_n^2}\Big).$$

Therefore

$$D_n^{(1)} \to 1 \quad \text{as} \quad n \to \infty. \tag{15}$$

Next we will show that $D_n^{(2)} \to 0$ as $n \to \infty$. It is easy to verify that

$$D_n^{(2)} = 2\sum_{k_1<k_2}\Big[\sum_{i=1}^{k_1-1}\operatorname{cov}(\eta_{ik_1}^2, \eta_{ik_2}^2) + \sum_{i=1}^{k_1-1}\operatorname{cov}(\eta_{ik_1}^2, \eta_{k_1k_2}^2)\Big].$$

But

$$E\eta_{ik_1}^2\eta_{ik_2}^2 \le c\Big(\frac{nb_n}{\sigma_n}\Big)^4 Q_{ik_1}^2 Q_{ik_2}^2 \le c\,\frac{1}{n^4\sigma_n^4}.$$

Similarly,

$$E\eta_{ij}^2 = O(n^{-2}\sigma_n^{-2}).$$

Therefore

$$\operatorname{cov}(\eta_{ik_1}^2, \eta_{ik_2}^2) = O((n\sigma_n)^{-4}). \tag{16}$$

Furthermore, since $\sum_{1\le k_1<k_2\le n}(k_1-1) = O(n^3)$, equation (16) implies

$$D_n^{(2)} = O\Big(\frac{1}{n\sigma_n^4}\Big) = O\Big(\frac{1}{nb_n^2}\Big). \tag{17}$$

Thus, according to (15) and (17),

$$A_n^{(1)} = 1 + O\Big(\frac{1}{nb_n^2}\Big). \tag{18}$$

Finally, we will show that $A_n^{(4)} \to 0$ as $n \to \infty$. Again, by the definition of η_{ij} we can write

$$\begin{aligned}
|A_n^{(4)}| &= 4\Big|\sum_{k_1<k_2}\sum_{t<s}^{k_1-1} E\eta_{sk_1}\eta_{tk_1}\eta_{sk_2}\eta_{tk_2}\Big| \le \\
&\le c\Big(\frac{nb_n}{\sigma_n}\Big)^4\Big[\Big|\sum_{s,t,k_1,k_2} Q_{sk_1}Q_{sk_2}Q_{tk_1}Q_{tk_2}\Big| + \\
&\qquad + \Big|\sum_{k,s,t} Q_{ks}^2 Q_{kt}^2\Big| + \Big|\sum_{k,s,t} Q_{kt}Q_{st}Q_{ks}Q_{ss}\Big|\Big] = \\
&= c\Big(\frac{nb_n}{\sigma_n}\Big)^4\big[|E_n^{(1)}| + |E_n^{(2)}| + |E_n^{(3)}|\big].
\end{aligned} \tag{19}$$

By the same argument as in (4), it can be shown that

$$\begin{aligned}
E_n^{(1)} &= n^{-7}b_n^{-8}\sum_{s,t,k}\Psi_n(\theta_s^{(1)}, \theta_k^{(2)})\Psi_n(\theta_t^{(1)}, \theta_k^{(2)}) \times \\
&\times \int_0^1 \Psi_n(\theta_s^{(1)}, u)\Psi_n(\theta_t^{(1)}, u)\,dP_n^{(1)}(u).
\end{aligned}$$

Hence, integrating by parts and taking into account that $\sup|\Psi_n(t_1,t_2)| \le cb_n$ and $\sup|K^{(1)}(u)| < \infty$, we obtain

$$E_n^{(1)} = n^{-7}b_n^{-8}\int_0^1 \sum_{s,t,k}\Psi_n(\theta_s^{(1)}, \theta_k^{(2)})\Psi_n(\theta_t^{(1)}, \theta_k^{(2)}) \times$$

$$\times\Psi_n(\theta_s^{(1)},u)\Psi_n(\theta_t^{(1)},u))du = O\Big(\frac{1}{n^5b_n^4}\Big). \tag{20}$$

In the same manner, we can rewrite (20) as

$$E_n^{(1)} = n^{-4}b_n^{-8}\int_0^1\int_0^1\int_0^1\int_0^1 \Psi_n(z,u)\Psi_n(z,t)\Psi_n(y,u)\times \\ \times\Psi_n(y,t)\,du\,dt\,dy\,dz + O(n^{-5}b_n^{-4}).$$

It is easy to verify that

$$n^{-4}b_n^{-8}\int_0^1\int_0^1\int_0^1\int_0^1 |\Psi_n(z,u)\Psi_n(z,t)\Psi_n(y,u)\Psi_n(y,t)|\,du\,dt\,dy\,dz \le \\ \le cn^{-4}b_n^{-2}\int_{\Omega_n(\tau)}\int_{\Omega_n(\tau)}\Big|K_0\Big(\frac{x-v}{b_n}\Big)\Big|\,dx\,dv \le cn^{-4}b_n^{-1}.$$

Thus

$$\Big(\frac{nb_n}{\sigma_n}\Big)^4 E_n^{(1)} = O(b_n) + O\Big(\frac{1}{nb_n^2}\Big). \tag{21}$$

Furthermore, it is not difficillt to show

$$(nb_n)^4\sigma_n^{-4}E_n^{(2)} = O\Big(\frac{1}{nb_n^2}\Big)$$

and

$$(nb_n)^4\sigma_n^{-4}E_n^{(3)} = O\Big(\frac{1}{nb_n^2}\Big). \tag{22}$$

Therefore (19), (21), and (22) imply

$$A_n^{(4)} \to 0 \quad \text{as} \quad n\to\infty. \tag{23}$$

Combining relations (12), (13), (14), (18), and (23), we obtain

$$E\Big(\sum_{k=1}^n \xi_{nk}^2 - 1\Big)^2 \to 0 \quad \text{as} \quad n\to\infty.$$

Therefore

$$\frac{U_n - EU_n}{\sigma_n} \xrightarrow{d} \xi.$$

But, due to Lemma 2, we have $EU_n = \sigma^2\theta_1 + O(b_n) + O((nb_n)^{-1})$ and $b_n^{-1}\sigma_n^2 \to \sigma^4\theta_2^2$, and hence we obtain

$$b_b^{-1/2}\Big(\frac{U_n - \sigma^2\theta_1}{\sigma^2\theta_2}\Big) \xrightarrow{d} \xi. \quad \square$$

Let us denote by

$$T_n = nb_n \int_{\Omega_n(\tau)} [g_n(x) - g(x)]^2 dx.$$

Theorem 2. *Suppose $g(x) \in F_s$, $s \geq 2$ and $K(x)$ satisfies conditions $1^\circ - 3^\circ$. If, in addition, $\sup\limits_{n\geq 1} EZ_{n1}^4 < \infty, nb_n^2 \to \infty$ and $nb_n^{2s} \to 0$, then*

$$b_n^{-1/2}(T_n - \sigma^2\theta_1)\sigma^{-2}\theta_2^{-1} \xrightarrow{d} \xi.$$

Proof. Note that

$$T_n = U_n + d_n^{(1)} + d_n^{(1)},$$

where

$$d_n^{(1)} = 2nb_n \int_{\Omega_n(\tau)} [g_n(x) - Eg_n(x)]\,[Eg_n(x) - g(x)]dx,$$

$$d_n^{(2)} = nb_n \int_{\Omega_n(\tau)} [Eg_n(x) - g(x)]^2 dx.$$

It is easy to verify that

$$Eg_n(x) - g(x) = b_n^{-1} \sum_{i=1}^{n} \int_{\Delta_i} K\Big(\frac{x-t}{b_n}\Big) dx\, g\Big(\frac{i}{n}\Big) - g(x) =$$

$$= b_n^{-1} \int_0^1 K\Big(\frac{x-t}{b_n}\Big)\,[\widetilde{g}_n(t) - g(t)]dt +$$

$$+ b_n^{-1} \int_0^1 K\Big(\frac{x-t}{b_n}\Big)\,[g(t) - g(x)]dt, \quad x \in \Omega_n(\tau), \tag{24}$$

with

$$\widetilde{g}_n(x) = \sum_{i=1}^{n} g\Big(\frac{i}{n}\Big)\, I_{\Delta_i}(x), \quad \Delta_i = \Big[\frac{i-1}{n}, \frac{i}{n}\Big].$$

But

$$\sup_{0\leq x\leq 1} |\widetilde{g}_n(x) - g(x)| \leq \max_{1\leq i\leq n} \sup_{x\in\Delta_i} \Big|g\Big(\frac{i}{n}\Big) - g(x)\Big| =$$

$$= \max_{1\leq i\leq n} \sup_{x\in\Delta_i} \Big|\frac{i}{n} - x\Big|\, |g'(\tau_i)| = O\Big(\frac{1}{n}\Big), \quad \tau_i \in \Delta_i.$$

Therefore the second term in (24) is $O(\frac{1}{n})$. Since $g(x) \in F_s$ and $K(x)$ satisfies conditions 1°–3°, we have

$$\sup_{x\in\Omega_n(\tau)} \Big|\int_0^1 K\Big(\frac{x-t}{b_n}\Big)\,[g(t) - g(x)]dt\Big| =$$

$$= \frac{b_n^s}{(s-1)!} \sup_{x\in\Omega_n(\tau)} \left| \int_{-\tau}^{\tau} \int_0^1 (1-t)^{s-1} g^{(s)}(x+tub_n) u^s K(u)du \right| \le cb_n^s. \quad (25)$$

Thus from (24) and (25) it follows that

$$\sup_{x\in\Omega_n(\tau)} |Eg_n(x) - g(x)| \le c\Big(b_n^s + \frac{1}{n}\Big). \tag{26}$$

Therefore

$$b_n^{-1/2} d_n^{(2)} \le c\big(nb_n^{2s}\sqrt{b_n} + b_n^{s+1/2} + n^{-1}\sqrt{b_n}\big). \tag{27}$$

On the other hand, (9) and (26) yield

$$b_n^{-1/2} E|d_n^{(1)}| \le nb_n^{1/2} \bigg\{ \int_{\Omega_n(\tau)} E(g_n(x) - Eg_n(x))^2 dx \times$$
$$\times \int_{\Omega_n(\tau)} (Eg_n(x) - g(x))^2 dx \bigg\}^{1/2} \le c\sqrt{n} b^s. \tag{28}$$

Finally, the statement of Theorem 2 follows directly from Theorem 1, (27) and (28). □

Using Theorem 2, it is easy to solve the problem of testing the hypothesis about $g(x)$. Given σ^2, it is required to verify the hypothesis

$$H_0 : g(x) = g_0(x), \quad x \in \Omega_n(\tau).$$

The critical region is defined approximately by the inequality

$$T_n^0 = nb_n \int (g_n(x) - g_0(x))^2 dx \ge q_n(\alpha), \tag{29}$$

where

$$q_n(\alpha) = \sigma^2\big(\theta_1 + \lambda_\alpha \sqrt{b_n}\theta_2\big), \quad \theta_1 = \int K^2(u)du, \quad \theta_2^2 = 2\int K_0^2(u)du,$$

and λ_α is the quantile of level $1-\alpha$ of the standard normal distribution, i.e.,

$$\Phi(\lambda_\alpha) = 1-\alpha, \quad \Phi(u) = (2\pi)^{-1/2} \int_{-\infty}^{u} \exp\Big(-\frac{x^2}{2}\Big)dx.$$

Suppose now that σ^2 is unknown. We will use the $b_n^{-1/2}$-consistent estimate of variance σ^2 (see, for example, [5]):

$$S_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2, \quad Y_i \equiv Y_{ni}.$$

Indeed,

$$ES_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} E\epsilon_i^2 + \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \Big[g\Big(\frac{i-1}{n}\Big) - g\Big(\frac{i}{n}\Big)\Big]^2,$$

where $\epsilon_i = Z_{i+1} - Z_i$, $Z_i \equiv Z_{ni}$. Moreover, since $E\epsilon_i^2 = 2\sigma^2$, we can write

$$ES_n^2 = \sigma^2 + \frac{1}{2(n-1)} \frac{1}{n^2} \sum_{i=1}^{n-1} (g^{(1)}(\tau_i))^2 = \sigma^2 + O(n^{-2}).$$

To calculate the variance, it is sufficient to note that

$$E(S_n^2 - \sigma^2)^2 = \frac{1}{4(n-1)^2} \sum_{i=1}^{n-1} E\epsilon_i^4 + O(n^{-2}) - \sigma^2 + \frac{1}{4(n-1)^2} \sum_{i\neq j} E\epsilon_i^2 E\epsilon_j^2 =$$

$$= \frac{1}{4(n-1)^2} \sum_{i=1}^{n-1} E\epsilon_i^4 + \frac{(n-2)(n-3)}{4(n-1)^2} (2\sigma^2)^2 - \sigma^4 + O(n^{-1}) =$$

$$= \frac{1}{4(n-1)^2} \sum_{i=1}^{n-1} E\epsilon_i^4 + O(n^{-1}).$$

Since $\sup\limits_{n\geq 1} EZ_{n1}^4 < \infty$, this yields $E(S_n^2 - \sigma^2)^2 = O(n^{-1})$. Therefore

$$b_n^{-1/2}(S_n^2 - \sigma^2) \xrightarrow{P} 0.$$

Corollary. *Under the conditions of Theorem* 2

$$b_n^{-1/2}\Big(\frac{T_n^0 - S_n^2\theta_1}{S_n^2\theta_2}\Big) \xrightarrow{d} \xi.$$

This corollary enables us to construct a test for verifying

$$H_0 : g(x) = g_0(x).$$

The critical region is defined approximately by the inequality $T_n^0 \geq \widetilde{q}_n(\alpha)$, where $\widetilde{q}_n(\alpha)$ is obtained from $q_n(\alpha)$ by using S_n^2 instead of σ^2.

Consider now the asymptotic properties of test (29) (i.e., the asymptotic behaviour of the power function as $n \to \infty$). First, let us study the question whether the corresponding test is consistent.

Theorem 3. *Under the conditions of Theorem* 2

$$\Pi_n(g_1) = P_{H_1}(T_n^0 \geq q_n(\alpha)) \to 1, \quad n \to \infty,$$

i.e., the test defined by (29) *is consistent under any alternatives*

$$H_1 : g(x) = g_1(x) \neq g_0(x), \quad \Delta = \int_0^1 (g_1(x) - g_0(x))^2 dx > 0.$$

Proof. Denote

$$Z_n(g_1) = b_n^{-1/2}\left(nb_n \int_{\Omega_n(\tau)} (g_n(x) - g_1(x))^2 dx - \sigma^2\theta_1\right)\sigma^{-2}\theta_2^{-1}.$$

It is easy to show that

$$\Pi_n(g_1) = P_{H_1}\big\{Z_n(g_1) \geq -nb_n^{1/2}(\theta_2^{-1}\sigma^{-2}\Delta + o_p(1))\big\}.$$

Since $Z_n(g_1)$ is asymptoticaly normally distributed with parameters $(0,1)$ under hypothesis H_1, $nb_n^{1/2} \to \infty$ and $\Delta > 0$, we have $\Pi_n(g_1) \to 1$ as $n \to \infty$. $\square$

Thus under any fixed alternative the power of test (29) tends to 1 as $n \to \infty$. Nevertheless, if the alternative hypothesis varies with n and becomes "closer" to the null hypothesis H_0, the power of the test may not converge to 1 depending on the rate at which the alternative approaches the null hypothesis. In our case the sequence of "close" alternatives has the form

$$H_n : \widetilde{g}_n(x) = g_0(x) + \gamma_n\phi(x) + o(\gamma_n). \tag{30}$$

Theorem 4. *Suppose $g_0(x)$ and $\varphi(x)$ are from F_s, but $K(x)$ satisfies coditions 1°–3° and $\sup\limits_{n\geq 1} EZ_{n1}^4 < \infty$. If $b_n = n^{-\delta}$, $\gamma_n = n^{-1/2+\delta/4}$, $1/2s < \delta < 1/2$, then under alternatives H_n the statistic*

$$b_n^{-1/2}(T_n^0 - \theta_1\sigma^2)\sigma^{-2}\theta_2^{-1}$$

has the limiting normal distribution with parameters $(\frac{1}{\sigma^2\theta_2}\int_0^1 \varphi^2(x)dx, 1)$.

Proof. Let us represent T_n^0 as the sum

$$T_n^0 = nb_n \int_{\Omega_n(\tau)} (g_n(x) - \widetilde{g}_n(x))^2 dx + nb_n \int_{\Omega_n(\tau)} (\widetilde{g}_n(x) - g_0(x))^2 dx +$$
$$+2nb_n \int_{\Omega_n(\tau)} (g_n(x) - \widetilde{g}_n(x))(\widetilde{g}_n(x) - g_0(x))dx = T_n^1 + A_1(n) + A_2(n).$$

It is easy to check that

$$b_n^{-1/2}A_1(n) = \int_0^1 \varphi^2(u)du + O(n^{-\delta}). \tag{31}$$

Let us introduce the random variable

$$d_n = \int_{\Omega_n(\tau)} (g_n(x) - E_1 g_n(x))\varphi(x)\,dx.$$

Here E_1 denotes the mathematical expectation under the hypothesis H_n. We can derive the inequality

$$b_n^{-1/2} E|A_2(n)| \le n b_n^{1/2} \gamma_n \left[E|d_n| + \int_{\Omega_n(\tau)} |E_1 g_n(x) - \widetilde{g}_n(x)| \varphi(x)\, dx \right] =$$
$$= n b_n^{1/2} \gamma_n E|d_n| + O(n \gamma_n b_n^{s+1/2}).$$

But

$$E|d_n| \le \sigma \left\{ \sum_{i=1}^{n} \left(b_n^{-1} \int_{\Omega_n(\tau)} \left(\int_{\Delta_i} K\Big(\frac{x-t}{b_n}\Big) dt \right)) \varphi(x)\, dx \right)^2 \right\}^{1/2} \le c n^{-1/2}.$$

Therefore

$$b_n^{-1/2} E|A_2(n)| \le c\big(n^{-\delta/4} + n^{\frac{2-(4s+1)\delta}{4}}\big) \to 0. \tag{32}$$

Referring to the proof of Theorem 2 it is easy to verify that $b_n^{-1/2}(T_n^0 - \sigma^2\theta_1)\sigma^{-2}\theta_2^{-1}$ is asymptotically normally distributed with parameters (0,1). Hence, from (31) and (32) we conclude that the theorem is valid. $\square$

Remark 1. It follows from Theorem 4 that more closer alternatives of form (30) (i.e., under $\gamma_n n^{1/2-\delta/4} \to 0$) are not distinguished from H_0 by this test (i.e., $P_{H_n}(T_n^0 \ge q_n(\alpha)) \to \alpha$), and for more remote alternatives (i.e., under $\gamma_n n^{1/2-\delta/4} \to \infty$) the corresponding test preserves the consistency property (i.e., $P_{H_n}(T_n^0 \ge q_n(\alpha)) \to 1$).

Thus the local behaviour of the power $P_{H_n}(T_n^0 \ge q_n(\alpha))$ is

$$P_{H_n}(T_n \ge q_n(\alpha)) \to 1 - \Phi\left(\lambda_\alpha - \frac{1}{\sigma^2\theta_2} \int_0^1 \varphi^2(u)\, du \right). \tag{33}$$

Since $\int_0^1 \varphi^2(u) du > 0$ and is equal to zero if and only if $\varphi(u) = 0$, it follows from (33) that the test for testing the hypothesis $H_0 : g(x) = g_0(x)$ against the alternative of form (30) is asymptotically strictly unbiased.

Remark 2. Theorem 4 is analogous to Theorem 6.1 from the book by J. D. Hart [6] with the only difference that [6] deals with the statistic based on the Priestley–Chao estimate [7].

Remark 3. If in the interval $[0,1]$ we choose points x_{nk} in a more general manner, i.e.,

$$\int_0^{x_{nk}} p(u)\, du = k/n, \quad k = 1, \dots, n,$$

where $p(x) > 0$, $x \in [0,1]$, is the probability density satisfying certain conditions of smoothness, then Theorems 1–4 remain valid provided that the parameters θ_1 and θ_2 are changed appropriately.

Remark 4. If instead of $K(x)$ we consider its modification $K_{q,r}(x)$ from [2], then we can give the hypothesis H_0 on the entire interval $[0,1]$. For the corresponding estimate we have $g_n^*(x) \equiv g_n(x)$, $x \in \Omega_n$, while the relation

$$nb_n^{1/2}\left(\int_0^1 (g_n^* - g)^2 dx - \int_{\Omega_n} (g_n - g)^2 dx\right) \xrightarrow{P} 0$$

will be proved in our forthcoming paper.

Remark 5. If in Theorem 4 we set $\delta = \frac{2}{2s+1}$, then $\gamma_n = n^{-\frac{s}{2s+1}}$. By Yu. Ingster's results [8] the test T_n is minimax consistent with respect to alternatives of form (30).

Acknowledgement

The authors express their thanks to the referee for his careful reading of the paper and comments that led to adding Remarks 2–5.

References

1. T. Gasser and H. Müller, Kernel estimation of regression functions. *Smoothing technique for curve estimation. Eds. T. Gasser and M. Rosenblatt. Lecture Notes in Math.* 757, 23–68, *Springer, Berlin etc.,* 1979.
2. J. Hart and T. Wehrly, Kernel regression when the boundary is large. *J. Amer. Statist. Assoc.* **87**(1992), No. 420, 1018–1022.
3. R. S. Lipster and A. N. Shiryaev, A functional limit theorem for semimartingales. (Russian) *Teor. Veroyatnost. i Primenen.* **25**(1980), No. 4, 683–703.
4. E. A. Nadaraya, Nonparametric estimation of probarility densities and regression curves. *Kluver Academic Publishers, Dordrecht,* 1989.
5. K. A. Brownlee, Statistical theory and methodology in science and engineering. *Wiley, New York,* 1960.
6. J. D. Hart, Nonparametric smoothing and lack-of-fit tests. *Springer, New York,* 1997.
7. M. B. Priestley and M. T. Chao, Nonparametric function fitting. *J. Roy. Stat. Soc. Ser. B* **34**(1972), 385–392.
8. Yu. I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives, I. *Math. Methods Statist.* **2**(1993), No. 2, 84–114.

(Received 20.05.1998)

Authors' address:
I. Javakhishvili Tbilisi State University
2, University St., Tbilisi 380043
Georgia