



**HAL**  
open science

# Asymptotically efficient estimates for nonparametric regression models

Leonid Galtchouk, Sergei Pergamenshchikov

► **To cite this version:**

Leonid Galtchouk, Sergei Pergamenshchikov. Asymptotically efficient estimates for nonparametric regression models. 2004. hal-00129706

**HAL Id: hal-00129706**

**<https://hal.science/hal-00129706>**

Preprint submitted on 8 Feb 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Asymptotically efficient estimates for nonparametric regression models.

**L. Galtchouk<sup>1</sup>, S. Pergamenshchikov<sup>2</sup>**

<sup>1</sup>IRMA, Département de Mathématiques,  
Université de Strasbourg  
7 rue René Descartes  
Strasbourg Cedex France

e-mail: [galtchou@math.u-strasbg.fr](mailto:galtchou@math.u-strasbg.fr)

<sup>2</sup>Laboratoire de Mathématiques Raphael Salem,  
UMR CNRS 6085, Site Colbert, UFR Sciences,  
Université de Rouen, F76 821 Mont St Aignan, Cedex, France.  
e-mail: [Serge.Pergamenchikov@univ-rouen.fr](mailto:Serge.Pergamenchikov@univ-rouen.fr)

## **Abstract**

The paper deals with estimating problem of regression function at a given state point in nonparametric regression models with gaussian noises and with nongaussian noises having unknown distribution. An asymptotically efficient kernel estimator is constructed for a local minimax risk.

**Key words:** asymptotical efficiency, kernel estimates, minimax, nonparametric regression, local risk.

**AMS (1991) Subject Classification :** primary 62G07; secondary 62G20

# 1 Introduction

The problem of minimax estimation of nonparametric regression function has been studied in a number of papers [see, for example, Ibragimov and Hasminskii (1979), Gunst and Mason (1980), Golubev (1992), Donoho, Johnstone, Kerkyacharian and Picard (1995), Barron, Birgé and Massart (1999), Efroimovich (1999), Nemirovskii (2000), Kalifa and Mallat (2003) and references cited therein].

This paper deals with nonparametric estimation of the regression function  $S$  at a given point  $z_0$ , i.e. the value  $S(z_0)$ , when the smoothness of  $S$  is known. In earlier published papers concerning estimation at a given point only the order of the optimal rate for a minimax risk was found but the sharp constant was not known.

Note that in estimation problems for nonparametric regression exact asymptotics have been found only in some cases : Nusbaum (1985), Golubev and Nussbaum (1992) for a  $L_2$ -risk and the regression function belonging to some Sobolev space, Korostelev (1993) for the uniform norm and a Hölder regression.

In this paper we find the sharp asymptotical lower bound for a local risk and prove that a kernel estimate is asymptotically efficient for Hölder classes with a known smoothness. We make use of the approach developed by Galtchouk and Pergamenshchikov (2002).

The main difficulty in obtaining the exact constant in the lower bound for the risk in the problem of estimating a function at a given point is bound with the traditional definition of risk (see, [11]) which does not allow one to perform the additional limit transition in the Hajek -Le Cam inequality. We would remind that the key idea in deriving the lower bound for the risks is to replace the supremum over the whole Hölder class of functions with that over some its parametric subclass and to apply the Hajek - Le Cam inequality for parametric problems. However in order to get the exact constant in the lower bound the additional transition to the limit is needed with the Hölder constant for this subclass tending to infinity. It cannot be done if one uses the traditional definition of risk which assumes that both the Hölder class and the Hölder constant are fixed. In this paper we propose to change the traditional definition of risk by employing some parametric family with an arbitrary large Hölder's constant, that is we introduce a set of functions for which only the smoothness order is fixed and the set may include the functions with arbitrary large constants. To this end we come to the integral description of vicinities based on the weak Hölder condition and which contain all Hölder's functions with a fixed smoothness order  $\beta$ . This means that the optimal estimation rate for such functions is  $n^{-\beta/(2\beta+1)}$ ,

where  $n$  is a number of observations. For the modified definition of risk we find a constant which as well as in the parametric case equals the mean of the loss function computed for some gaussian random variable. This constant fully coincides with the corresponding normalized limiting expectation of the loss function calculated for the stochastic term in the expansion of the kernel estimate with the indicator function  $\chi_{\{(-1,1)\}}$  as a kernel. This property is connected with the fact that the variance of the stochastic term of the indicator kernel estimate attains its minimal values in the class of all kernel estimates.

In other words, in order to prove the optimality property for the kernel estimate one has to show that the principal term in the asymptotic representation is the stochastic one. There arises a question if one can exclude the influence of the approximating term. Two reasons make it possible. First we consider a local risk in which the supremum is taken over a special vicinity with the centre having the smoothness order slightly higher than the assumed one  $\beta$  (see below the definition of "zero-constant" Hölder condition in section 4). Second, the complementary term to the central function, which has the smoothness order  $\beta$ , is very small and in the limit as the radius of the vicinity tends to zero the approximating term vanishes and does not influence on the limiting constant for the risk of the kernel estimate and the stochastic term turns out to be the principal one. Notice that by the similar way the efficient estimator is constructed by Belitser (2000) for the density estimation problem at a given point.

The paper is organized as follows. In the next section we formulate the problem and give definitions of a function class, vicinities and a local risk. In section 3 the sharp lower bound for the local minimax risk over all estimates is established. In section 4 we prove that the upper bound of the local risk for the kernel estimate coincides with the lower bound, i.e. this estimate is asymptotically efficient. In section 5 the asymptotical efficiency is proved in the case of nongaussian noises with unknown distributions. Appendix contains some technical results.

## 2 Statement of the problem

Consider the following regression model

$$y_k = S(x_k) + \xi_k, \quad 1 \leq k \leq n, \quad (2.1)$$

where  $S(\cdot)$  is an unknown function,  $x_k = k/n$  and  $\{\xi_k\}$  are i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ .

The problem is to estimate the function  $S(\cdot)$  at a fixed point  $z_0 \in [0, 1]$ , i.e. the value  $S(z_0)$  by observations  $\{y_k, 1 \leq k \leq n\}$ .

Assume that the unknown function  $S = S(x)$ ,  $x \in \mathbf{R}$ , belongs to class

$$\mathcal{H}(\beta) = \bigcup_{K>0} \mathcal{H}(K, \beta), \quad (2.2)$$

$\mathcal{H}(K, \beta)$  is Hölder's class with a constant  $K$  and exponent  $\beta = 1 + \alpha$ ,  $\alpha \in (0, 1)$ .

To solve this problem we will make use of a special local risk. First we construct the vicinities.

For each function  $S_0 \in \mathcal{H}(\beta)$  and  $0 < \delta < 1$  we introduce a family of vicinities  $\mathcal{U}_{\delta, L}(S_0)$  in  $\mathcal{H}(\beta)$  as

$$\mathcal{U}_{\delta, L, n}(S_0) = \left\{ S(\cdot) : S(x) = S_0(x) + D(x), D \in \mathbf{C}_1, \sup_{x \in \mathbf{R}} |\dot{D}(x)| \leq L, \left| \int_{-1}^1 (D(z_0 + hz) - D(z_0)) dz \right| \leq \delta/n^{\beta/(2\beta+1)} \right\}, \quad (2.3)$$

where  $L$  is a given positive constant. Notice that the class of functions  $\mathcal{U}_{\delta, L}(S_0)$  comprises, for example, a set of functions  $S$  with

$$D(x) = \varphi_n^{-1} V_\nu \left( \frac{x - z_0}{h} \right), \quad (2.4)$$

where  $\varphi_n = n^{\beta/(2\beta+1)}$ ,

$$V_\nu(x) = \nu^{-1} \int_{-\infty}^{\infty} \tilde{Q}_\nu(u) g \left( \frac{u - x}{\nu} \right) du, \quad (2.5)$$

$$\tilde{Q}_\nu(u) = \mathbf{1}_{\{|u| \leq 1-2\nu\}} + 2\mathbf{1}_{\{1-2\nu \leq |u| \leq 1-\nu\}},$$

$$g(z) = \begin{cases} c \exp\{-(1-z^2)^{-1}\}, & |z| \leq 1, \\ 0, & |z| > 1, \end{cases}$$

$0 < \nu < 1/4$ , a positive constant  $c$  is such that  $\int_{-1}^1 g(z) dz = 1$ . For any  $0 < \nu < 1/4$ , one has

$$V_\nu(0) = 1, \quad \text{and} \quad \int_{-1}^1 V_\nu(x) dx = 2.$$

Therefore,

$$\int_{-1}^1 (D(z_0 + hz) - D(z_0)) dz = 0.$$

Since

$$|\dot{D}(x)| = \varphi_n^{-1} h^{-1} |\dot{V}_\nu(\frac{x - z_0}{h})| \leq n^{-\alpha/(2\beta+1)} \nu^{-2} \dot{g}^*, \quad \dot{g}^* = \sup_x |\dot{g}(x)|,$$

the function  $S = S_0 + D$  belongs to  $\mathcal{U}_{\delta,L}(S_0)$ , if one chooses  $\delta$  for each  $n \geq 1$  such that

$$n^{-\beta/(2\beta+1)} \nu^{-1} g^* + n^{-\alpha/(2\beta+1)} \nu^{-2} \dot{g}^* < \delta, \quad g^* = \sup_x g(x).$$

**Remark 2.1.** To explain definition (2.3) of the vicinity  $\mathcal{U}_{\delta,L}$  we introduce the following definition.

A function  $D$  is said to satisfy the weak Hölder condition with an exponent  $\beta$  and a constant  $\delta$  at a point  $z_0$  if

$$\left| \int_{-1}^1 (D(z_0 + hz) - D(z_0)) dz \right| \leq \delta h^\beta.$$

In particular, for  $\beta = 1 + \alpha$ ,  $0 < \alpha < 1$ , the relation

$$\int_{-1}^1 (D(z_0 + hz) - D(z_0)) dz = \int_{-1}^1 \left( \int_{z_0}^{z_0 + hz} (\dot{D}(t) - \dot{D}(z_0)) dt \right) dz$$

shows that if  $D$  is differentiable and the derivative satisfies the usual Hölder condition with exponent  $\alpha$ ,  $0 < \alpha < 1$ , and a constant  $\delta$ , then  $D$  satisfies the weak Hölder condition with the exponent  $\beta = 1 + \alpha$  and the same constant  $\delta$ . Therefore the vicinity  $\mathcal{U}_{\delta,L}$  in (2.3) comprises functions  $S = S_0 + D$  with  $D$  satisfying the weak Hölder condition with exponent  $\beta = 1 + \alpha$  and constant  $\delta$  at the point  $z_0$ .

For any estimate  $\tilde{S}_n(z_0)$  of  $S(z_0)$  we define the local risk as

$$\mathcal{R}_{\delta,n}(\tilde{S}_n, S_0) = \sup_{S \in \mathcal{U}_{\delta,L}(S_0)} \mathbf{E}_S \psi_n(\tilde{S}_n, S), \quad (2.6)$$

where the expectation  $\mathbf{E}_S$  is taken with respect to the law  $\mathbf{P}_S$  in (2.1),

$$\psi_n(\tilde{S}_n, S) = \varphi_n |\tilde{S}_n(z_0) - S(z_0)|, \quad \varphi_n = n^{\beta/(2\beta+1)}.$$

**Remark 2.2.** It is easy to see that the vicinity  $\mathcal{U}_{\delta,L}(S_0)$  contains the functions of the form  $S = S_0 + D$ , where  $D$  belongs to Hölder classes  $\mathcal{H}(K, \beta)$  with  $K \leq \delta$ . As is well known (see, [11]) the rate of convergence is equal to  $\varphi_n$  for Hölder's class with exponent  $\beta$ . By this reason the local risk is defined by (2.6).

**Remark 2.3.** *As we will see in Section 3 in order to derive the lower bound for the risk one has roughly speaking to proceed to the limit in the minimax risk with respect to Hölder's constant as  $K \rightarrow \infty$ . It will be observed that the common definition of risk (see, [11]) when Hölder class  $\mathcal{H}(K, \beta)$  is fixed, does not permit this limit transition. Therefore in this paper we apply the definition of vicinity (2.3) according to which a vicinity contains a family of functions  $S = S_0 + D$  with  $D$  from (2.4) with an arbitrary large Hölder constant.*

### 3 Lower bound

**Theorem 3.1.** *Suppose that  $S_0 \in \mathcal{H}(\beta)$ . Then for any  $0 < \delta < 1$*

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{S}} \mathcal{R}_{\delta, n}(\tilde{S}, S_0) \geq \mathbf{E}|\eta|, \quad \eta \sim \mathcal{N}(0, \sigma^2/2).$$

**Proof.** Denote

$$S_u(x) = S_0(x) + \frac{u}{\varphi_n} V_\nu \left( \frac{x - z_0}{h} \right),$$

where

$$h = n^{-1/(2\beta+1)}, \quad \beta = 1 + \alpha,$$

the function  $V_\nu(\cdot)$  is defined in (2.5),  $u$  is a real number. For any  $b > 0$  and  $\delta > 0$ , there exists  $n_{b, \delta} > 0$  such that  $S_u \in \mathcal{U}_{\delta, L}(S_0)$  provided that  $|u| \leq b$  and  $n \geq n_{b, \delta}$ .

For  $n \geq n_{b, \delta}$ , we have the inequalities

$$\begin{aligned} \mathbf{R}_{\delta, n}(\tilde{S}, S_0) &\geq \sup_{|u| \leq b} \mathbf{E}_{S_u} \psi_n(\tilde{S}, S_u) \\ &\geq \frac{1}{2b} \int_{-b}^b \mathbf{E}_{S_u} \psi_{a, n}(\tilde{S}, S_u) du := I_n(a, b), \end{aligned} \quad (3.1)$$

where

$$\psi_{a, n}(\tilde{S}, S_u) = v_a(\varphi_n(\tilde{S}_n(z_0) - S_u(z_0))), \quad v_a(x) = |x| \wedge a, \quad a > 0.$$

Notice that measure  $\mathbf{P}_{S_u}$  is equivalent to measure  $\mathbf{P}_{S_0}$  and in this case the Radom-Nikodym derivative can be written as

$$\begin{aligned} \rho_n(u) &= \frac{d\mathbf{P}_{S_u}}{d\mathbf{P}_{S_0}} = \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^n ((y_k - S_u(x_k))^2 - (y_k - S_0(x_k))^2)\right\} \\ &= \exp\left\{-u \varsigma_n \eta_n - \frac{u^2}{2} \varsigma_n^2\right\}, \end{aligned} \quad (3.2)$$

where

$$\eta_n = \frac{1}{\sigma^2 \varsigma_n \varphi_n} \sum_{k=1}^n V_\nu \left( \frac{x_k - z_0}{h} \right) \xi_k, \quad \varsigma_n^2 = \frac{1}{\sigma^2 \varphi_n^2} \sum_{k=1}^n V_\nu^2 \left( \frac{x_k - z_0}{h} \right).$$

Notice that  $\eta_n \sim \mathcal{N}(0, 1)$  for any  $n \geq 1$  and

$$\lim_{n \rightarrow \infty} \varsigma_n^2 = \lim_{n \rightarrow \infty} \frac{1}{n \sigma^2} \sum_{k=1}^n \frac{1}{h} V_\nu^2 \left( \frac{x_k - z_0}{h} \right) = \frac{1}{\sigma^2} \int_{-1}^1 V_\nu^2(u) du := \sigma_\nu^2.$$

Therefore we can write that

$$\rho_n(u) = \exp \left\{ -u \sigma_\nu \eta_n - \frac{u^2 \sigma_\nu^2}{2} + r_n \right\},$$

where

$$\mathbf{P}_{S_0} - \lim_{n \rightarrow \infty} r_n = 0.$$

Now for the term  $I_n(a, b)$  in (3.1) one has the representation

$$\begin{aligned} I_n(a, b) &= \frac{1}{2b} \int_{-b}^b \mathbf{E}_{S_0} \psi_{a,n}(\tilde{S}, S_u) \rho_n(u) du \\ &\geq \frac{1}{2b} \int_{-b}^b \mathbf{E}_{S_0} \mathbf{1}_{B_d} \psi_{a,n}(\tilde{S}, S_u) \varrho_n(u) du + \delta_n(a, b) \\ &:= J_n(a, b) + \delta_n(a, b), \end{aligned}$$

where  $B_d = \{|\eta_n| \leq d\}$ ,  $d = \sigma_\nu(b - \sqrt{b})$ ,  $b > 1$ ,  $\varrho_n(u) = e^{u \sigma_\nu \eta_n - u^2 \sigma_\nu^2 / 2}$ ,

$$\begin{aligned} \delta_n(a, b) &= \frac{1}{2b} \int_{-b}^b \mathbf{E}_{S_0} \mathbf{1}_{B_d} \psi_{a,T}(\tilde{S}, S_u) \theta_n(u) du, \\ \theta_n(u) &= \rho_n(u) - \varrho_n(u). \end{aligned}$$

Note that the sequence  $\{\rho_n(u), n \geq 1\}$  is uniformly integrable. Indeed,

$$\mathbf{E}_{S_0} \rho_n(u) = 1,$$

for  $n \geq 1$ , since  $\rho_n(u)$  is a density, and

$$\rho_n(u) \xrightarrow[n \rightarrow \infty]{} \rho_\infty(u) = e^{u \sigma_\nu \eta - u^2 \sigma_\nu^2 / 2}, \quad \eta \sim \mathcal{N}(0, 1)$$

with  $\mathbf{E} \rho_\infty(u) = 1$  for all  $|u| \leq b$ .

Since  $\varrho_n(u)$  is bounded on  $B_d$  we obtain that

$$\lim_{n \rightarrow \infty} \delta_n(a, b) = 0.$$



Rewrite  $\varrho_n(u)$  as

$$\varrho_n(u) = \zeta_n e^{-\sigma_\nu^2(u-\tilde{\eta}_n)^2/2}, \quad \zeta_n = e^{\eta_n^2/2}, \quad \tilde{\eta}_n = \eta_n/\sigma_\nu.$$

Then we have

$$\begin{aligned} J_n(a, b) &= \frac{1}{2b} \int_{-b}^b \mathbf{E}_{S_0} \mathbf{1}_{B_d} \psi_{a,n}(\tilde{S}, S_u) \varrho_n(u) du \\ &= \mathbf{E}_{S_0} \mathbf{1}_{B_d} \zeta_n \frac{1}{2b} \int_{-b}^b v_a(u - c_n) \exp\left\{-\frac{\sigma_\nu^2}{2}(u - \tilde{\eta}_n)^2\right\} du \\ &= \mathbf{E}_{S_0} \mathbf{1}_{B_d} \zeta_n \frac{1}{2b} \int_{-b-\tilde{\eta}_n}^{b-\tilde{\eta}_n} v_a(u - c_n + \tilde{\eta}_n) \exp\left\{-\frac{\sigma_\nu^2}{2}u^2\right\} du \\ &\geq \mathbf{E}_{S_0} \mathbf{1}_{B_d} \zeta_n \frac{1}{2b} \int_{-\sqrt{b}}^{\sqrt{b}} v_a(u - c_n + \tilde{\eta}_n) \exp\left\{-\frac{\sigma_\nu^2}{2}u^2\right\} du, \end{aligned}$$

where  $c_n = \varphi_n(\tilde{S}_n(z_0) - S_0(z_0))$ .

Applying the Anderson inequality (see Ibragimov and Hasminskii (1981) Ch.II, Lemma 10.1) yields

$$J_n(a, b) \geq \mathbf{E}_{S_0} \mathbf{1}_{B_d} \zeta_n \frac{1}{2b} \int_{-\sqrt{b}}^{\sqrt{b}} v_a(u) \exp\left\{-\frac{\sigma_\nu^2}{2}u^2\right\} du.$$

Taking into account that

$$\mathbf{E}_{S_0} \mathbf{1}_{B_d} \zeta_n = 2\sigma_\nu(b - \sqrt{b})/\sqrt{2\pi}$$

one gets

$$\lim_{a \rightarrow \infty} \liminf_{n \rightarrow \infty} I_n(a, b) \geq \frac{b - \sqrt{b}}{b} \frac{\sigma_\nu}{\sqrt{2\pi}} \int_{-\sqrt{b}}^{\sqrt{b}} |u| \exp\left\{-\frac{u^2 \sigma_\nu^2}{2}\right\} du.$$

Limiting  $b \rightarrow \infty$  and taking into account that  $\sigma_\nu^2 \rightarrow 2/\sigma^2$  as  $\nu \rightarrow 0$  yield the desired result.  $\square$

## 4 Upper bound

In this section we will find the upper bound for risk (2.6) for a kernel estimate

$$\hat{S}_n(z_0) = \left( \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) \right)^{-1} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) y_k, \quad (4.1)$$

with  $Q(\cdot) = \chi_{[-1,1]}$  and  $h = n^{-1/(2\beta+1)}$ .

We need the following definition.

**Definition** A function  $S$  is said to satisfy the "zero-constant" Hölder condition with an exponent  $\gamma > 0$  at the point  $z_0$  if

$$\lim_{y \rightarrow z_0} \frac{S(y) - S(z_0)}{|y - z_0|^\gamma} = 0.$$

**Example.** Let us consider the following function

$$S(y) = \int_0^y L_0(z) \sin(\sqrt{|z|}) dz, \quad L_0(z) = \frac{1}{1 + |\ln|z||}.$$

The derivative of this function satisfies the "zero-constant" Hölder condition with the exponent  $\gamma = 1/2$ , i.e. this function belongs to a Hölder class with exponent  $3/2$ , but it does not belong to any Hölder class with an exponent greater than  $3/2$ .

Furthermore, all functions which satisfy the Hölder condition with exponent  $\gamma' > \gamma$ , satisfy the "zero-constant" Hölder condition with the exponent  $\gamma$ .

**Theorem 4.1.** Let  $S_0 \in \mathcal{H}(\beta)$ ,  $\beta = 1 + \alpha$ ,  $0 < \alpha < 1$ . If the derivative  $\dot{S}_0$  at a point  $z_0$  satisfies the "zero-constant" Hölder condition with exponent  $\alpha$ , then the kernel estimator (4.1) satisfies the following inequality :

$$\limsup_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathcal{R}_{\delta,n}(\hat{S}_n, S_0) \leq \mathbf{E}|\eta|, \quad \eta \sim \mathcal{N}(0, \sigma^2/2).$$

**Proof.** The estimate error is

$$\hat{S}_n(z_0) - S(z_0) = B_n + \frac{1}{\sqrt{q_n}} \zeta_n, \quad (4.2)$$

where

$$\begin{aligned} q_n &= \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right), \\ B_n &= \frac{1}{q_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (S(x_k) - S(z_0)), \\ \zeta_n &= \frac{1}{\sqrt{q_n}} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) \xi_k. \end{aligned}$$

Let us show that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{S \in \mathcal{U}_{\delta,L}(S_0)} \varphi_n |B_n| = 0. \quad (4.3)$$

We obtain that

$$B_n = a_n + b_n, \quad (4.4)$$

where

$$a_n = \frac{1}{q_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (S_0(x_k) - S_0(z_0)),$$

$$b_n = \frac{1}{q_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (D(x_k) - D(z_0)).$$

Denoting

$$r(y) = S_0(y) - S_0(z_0) - \dot{S}_0(z_0)(y - z_0)$$

one has

$$a_n = \dot{S}_0(z_0) \frac{1}{q_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (x_k - z_0)$$

$$+ \frac{1}{q_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) r(x_k) := \dot{S}_0(z_0) a_{1,n} + a_{2,n}.$$

Taking into account that

$$\lim_{n \rightarrow \infty} \frac{q_n}{\varphi_n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \frac{1}{h} Q\left(\frac{x_k - z_0}{h}\right) = \int_{-1}^1 Q(u) du = 2, \quad (4.5)$$

one has

$$\varphi_n a_{1,n} = \left(\frac{1}{2} + o(1)\right) \frac{1}{\varphi_n} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (x_k - z_0)$$

$$= \left(\frac{1}{2} + o(1)\right) \frac{nh^2}{\varphi_n} \sum_{k=1}^n Q(u_k) u_k (u_k - u_{k-1})$$

$$:= \left(\frac{1}{2} + o(1)\right) \frac{nh^2}{\varphi_n} I_n,$$

where  $u_k = (x_k - z_0)/h$ . Here

$$I_n = \sum_{k=1}^n Q(u_k) u_k (u_k - u_{k-1}) = \sum_{k=k_*}^{k^*} u_k (u_k - u_{k-1}),$$

where  $k_* = [nz_0 - hn] + 1$ ,  $k^* = [nz_0 + hn]$ . Therefore

$$I_n = \frac{1}{(nh)^2} \sum_{k=k_*}^{k^*} (k - nz_0) = \frac{1}{2(nh)^2} (k^* - k_* + 1)(k^* + k_* - 2nz_0).$$

Thus

$$|I_n| \leq \frac{2hn + 1}{(nh)^2},$$

which implies that

$$\lim_{n \rightarrow \infty} \varphi_n a_{1,n} = 0.$$

Let us check that

$$\lim_{n \rightarrow \infty} \varphi_n a_{2,n} = 0.$$

We have

$$\begin{aligned} a_{2,n} &= \left(\frac{1}{2} + o(1)\right) \frac{1}{\varphi_n^2} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) r(x_k) \\ &= \left(\frac{1}{2} + o(1)\right) \sum_{k=1}^n Q(u_k) r(z_0 + hu_k) (u_k - u_{k-1}). \end{aligned}$$

Taking into account that

$$\lim_{n \rightarrow \infty} \frac{r(z_0 + h)}{h^\beta} = 0,$$

we obtain that

$$\lim_{n \rightarrow \infty} \varphi_n a_{2,n} = 0.$$

Let us consider now the second term in the sum for  $B_n$  in (4.4). One has

$$\begin{aligned} b_n &= \left(\frac{1}{2} + o(1)\right) \frac{1}{nh} \sum_{k=1}^n Q\left(\frac{x_k - z_0}{h}\right) (D(x_k) - D(z_0)) \\ &= \left(\frac{1}{2} + o(1)\right) \sum_{k=1}^n Q(u_k) (D(z_0 + hu_k) - D(z_0)) (u_k - u_{k-1}) \\ &= \left(\frac{1}{2} + o(1)\right) \int_{-1}^1 (D(z_0 + hu) - D(z_0)) du + \left(\frac{1}{2} + o(1)\right) \Delta_n, \end{aligned}$$

where  $u_k = (x_k - z_0)/h$  and

$$\begin{aligned} \Delta_n &= \sum_{k=1}^n Q(u_k) (D(z_0 + hu_k) - D(z_0)) (u_k - u_{k-1}) - \int_{-1}^1 (D(z_0 + hu) - D(z_0)) du \\ &= \sum_{k=k_*}^{k^*} \int_{u_{k-1}}^{u_k} (D(z_0 + hu_k) - D(z_0 + hu)) du \\ &\quad - \int_{u_{k_*}}^1 (D(z_0 + hu) - D(z_0)) du + \int_{u_{k_*-1}}^{-1} (D(z_0 + hu) - D(z_0)) du. \end{aligned}$$

From here we obtain that

$$|\Delta_n| \leq c/n$$

for some positive constant  $c$ . Therefore by the definition of the vicinity in (2.3)

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \varphi_n |b_n| = 0.$$

By (4.5) and in view of  $\zeta_n \sim \mathcal{N}(0, \sigma^2)$ , we get the desired result.  $\square$

## 5 Case of unknown noise distribution.

Let us consider now the model (2.1) with unknown distribution of the noise  $\{\xi_k, k \geq 1\}$ . We will suppose that the density  $p$  of this distribution exists and  $p \in \mathcal{P}_{\epsilon, M}$ , where

$$\mathcal{P}_{\epsilon, M} = \left\{ p : \int_{-\infty}^{+\infty} x p(x) dx = 0, \int_{-\infty}^{+\infty} x^2 p(x) dx = \sigma^2, \int_{-\infty}^{+\infty} |x|^{2+\epsilon} p(x) dx \leq M \right\}, \quad (5.1)$$

where  $\epsilon$  and  $M$  are fixed positive constants.

In this case we will use the following risk :

$$\mathcal{R}_n^*(\tilde{S}_n, S_0) = \sup_{p \in \mathcal{P}_{\epsilon, M}} \sup_{S \in \mathcal{U}_{\delta, L}(S_0)} \mathbf{E}_S \psi_n(\tilde{S}_n, S), \quad (5.2)$$

where  $\psi_n(\tilde{S}_n, S) = \varphi_n |\tilde{S}_n(z_0) - S(z_0)|$ ,  $\varphi_n = n^{\beta/(2\beta+1)}$ . Notice that by theorem 3.1 the lower bound for the risk (5.2) is

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{S}} \mathcal{R}_{\delta, n}^*(\tilde{S}, S_0) \geq \mathbf{E}|\eta|, \quad \eta \sim \mathcal{N}(0, \sigma^2/2). \quad (5.3)$$

The upper bound is given in the following theorem.

**Theorem 5.1.** *Let  $S_0 \in \mathcal{H}(\beta)$ ,  $\beta = 1 + \alpha$ ,  $0 < \alpha < 1$ . If derivative  $\dot{S}_0$  at the point  $z_0$  satisfies the "zero-constant" Hölder condition with an exponent  $\alpha$ , then the kernel estimator (4.1) satisfies the following inequality :*

$$\limsup_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathcal{R}_{\delta, n}^*(\hat{S}_n, S_0) \leq \mathbf{E}|\eta|, \quad \eta \sim \mathcal{N}(0, \sigma^2/2).$$

**Proof.** Taking into account that  $B_n$  in (4.2) does not depend on the distribution of  $(\xi_k)$  it suffices to show that

$$\lim_{n \rightarrow \infty} \sup_{p \in cP_{\epsilon, M}} |\mathbf{E} |\zeta_n| - \mathbf{E} |\zeta_\infty|| = 0, \quad (5.4)$$

where  $\zeta_\infty \sim \mathcal{N}(0, \sigma^2)$ . One has

$$\limsup_{n \rightarrow \infty} \sup_{p \in cP_{\epsilon, M}} \mathbf{E} \zeta_n^2 = \sigma^2 < \infty,$$

that is the sequence  $\{\zeta_n\}$  is uniformly integrable. From here by lemma 6.2 we come to (5.4).  $\square$

## 6 Appendix

We will make use of the following result.

**Lemma 6.1.** (cf. [6], p. 90-91) *Let  $0 < \delta < 1$  and  $r > 0$ . Assume that  $u_k, k = 0, 1, \dots$  is a martingale difference sequence satisfying*

$$|u_k| \leq \delta$$

and

$$\sum_{i=0}^{\infty} \mathbf{E}(u_i^2 | \mathcal{F}_i) \geq r.$$

Let

$$\tau = \inf \left\{ n : \sum_{i=1}^n \mathbf{E}(u_i^2 | \mathcal{F}_i) > r \right\}.$$

*There exists a function  $\rho : (0, +\infty) \rightarrow [0, 2]$  not depending on distribution of the martingale difference, such that  $\lim_{x \rightarrow 0} \rho(x) = 0$  and*

$$\sup_{x \in \mathbf{R}} \left| \mathbf{P} \left( \sum_{k=1}^{\tau} u_k \leq x \right) - \Phi(x/r^{1/2}) \right| \leq \rho(\delta/r^{1/2}),$$

where  $\Phi$  is the standard normal distribution function.

**Lemma 6.2.** *The sequence  $\{\zeta_n\}$  defined by (4.2) satisfies the following limiting property*

$$\zeta_n \implies \zeta_\infty \sim \mathcal{N}(0, \sigma^2)$$

uniformly in  $p \in \mathcal{P}_{\epsilon, M}$ .

**Proof.** First, notice that we can represent  $\zeta_n$  as

$$\zeta_n = \sum_{k=1}^n u_k, \quad u_k = \frac{1}{\sqrt{q_n}} Q\left(\frac{x_k - z_0}{h}\right) \xi_k.$$

Write  $\xi_k = \xi'_k + \xi''_k$ , where

$$\begin{aligned} \xi'_k &= \xi_k \chi_{\{|\xi_k| \leq q_n^{1/4}\}} - \mathbf{E} \xi_1 \chi_{\{|\xi_1| \leq q_n^{1/4}\}}, \\ \xi''_k &= \xi_k \chi_{\{|\xi_k| \geq q_n^{1/4}\}} - \mathbf{E} \xi_1 \chi_{\{|\xi_1| \geq q_n^{1/4}\}}. \end{aligned}$$

From here it follows that

$$\begin{aligned} \zeta_n &= \zeta'_n + \zeta''_n = \sum_{k=1}^n u'_k + \sum_{k=1}^n u''_k, \\ u'_k &= \frac{1}{\sqrt{q_n}} Q\left(\frac{x_k - z_0}{h}\right) \xi'_k, \quad u''_k = \frac{1}{\sqrt{q_n}} Q\left(\frac{x_k - z_0}{h}\right) \xi''_k. \end{aligned}$$

Denoting  $\delta_n = q_n^{-1/4}$  and  $r_n = \mathbf{E}(\xi'_1)^2$  one gets

$$\begin{aligned} \tau_n &= \inf\{n : \sum_{i=1}^n \mathbf{E}((u'_i)^2 | \mathcal{F}_i) > r_n\} \\ &= \inf\{k \geq 1 : \sum_{j=1}^k Q\left(\frac{x_j - z_0}{h}\right) \geq q_n\}. \end{aligned}$$

Therefore we can write that  $\zeta'_n = \sum_{k=1}^{\tau_n} u'_k$  and  $|u'_k| \leq \delta_n = q_n^{-1/4}$ .

By the definition of  $\mathcal{P}_{\epsilon, M}$

$$\lim_{a \rightarrow \infty} \sup_{p \in \mathcal{P}_{\epsilon, M}} \int_{-\infty}^{+\infty} x^2 \chi_{\{|x| > a\}} p(x) dx = 0.$$

It means that  $r_n \rightarrow \sigma^2$  and  $\mathbf{E}(\zeta'_n)^2 = \mathbf{E}(\xi''_1)^2 \rightarrow 0$  as  $n \rightarrow \infty$  uniformly in  $\mathcal{P}_{\epsilon, M}$ . This completes the proof of lemma 6.2.  $\square$

## References

- [1] Barron, A., Birgé, L. and Massart, P.(1999) Risk bounds for model selection via penalization. *Probab. Theory Rel. Fields*, **113**, 301-413.
- [2] Belitser, E. (2000) Local minimax pointwise estimation of a multivariate density. *Statistica Neerlandica*, **54**(3) p. 351-365.
- [3] Dahlhaus, R. (1995) Efficient location and regression estimation for longrange dependent regression models. *Ann. Statist.*, **23** (3), p. 1029-1047
- [4] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage : asymptotia? *J. Roy. Statist. Soc. Ser. B*, **57**, p. 301-369
- [5] Efromovich, S.Yu. (1999) *Nonparametric curve estimation. Methods, theory and applications*. Springer-Verlag, Berlin, New York.
- [6] Freedman, D.(1971) *Brownian Motion and Diffusion*. Holden Day, San Francisco.
- [7] Galtchouk, L. and Pergamenshchikov, S. (2002) Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Bernoulli J.* (submitted).
- [8] Gunst, R.F. and Mason, R.L. (1980) *Regression analysis and its applications: a data oriented approach*. New-York, Marcel Dekker.
- [9] Golubev, G.K. (1992) Asymptotic minimax estimation of regression function in additive model.*Problems Inform. Trans.*, **28**, p. 3-15.
- [10] Golubev, G.K. and Nussbaum, M. (1993) Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.*, **37**, p. 521-529.
- [11] Ibragimov, I.A. and Hasminskii, R.Z. (1979) *Statistical Estimation: Asymptotic Theory*, Springer, Berlin, New York.
- [12] Kalifa, J. and Mallat,S. (2003) Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.*, **31** (1), p. 58-109.
- [13] Korostelev, A.P.. (1993) Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.* **38**, 737-743.



- [14] Nussbaum, M. (1985) Spline smothing in regression models and asymptotic efficiency in  $\mathbf{L}_2$ . *Ann. Statist.*, **13**, p. 984-997.
- [15] Nemirovski, A. (2000) Topics in non-parametric statistics. *Lectures on probability theory and statistics, in Ecole de Saint-Flour 1998, Lecture Notes in Math.*, **1738**, Springer, p. 85-277