

# Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression.

By Leonid Galtchouk and Sergey Pergamenshchikov  
Louis Pasteur University of Strasbourg and University of Rouen

## Abstract

The paper deals with asymptotic properties of the adaptive procedure proposed in the author paper, 2007, for estimating an unknown nonparametric regression. We prove that this procedure is asymptotically efficient for a quadratic risk, i.e. the asymptotic quadratic risk for this procedure coincides with the Pinsker constant which gives a sharp lower bound for the quadratic risk over all possible estimators.

<sup>1</sup> <sup>2</sup>

---

<sup>1</sup> *AMS 2000 Subject Classification* : primary 62G08; secondary 62G05, 62G20

<sup>2</sup> *Key words*: asymptotic bounds, adaptive estimation, efficient estimation, heteroscedastic regression, nonparametric regression, Pinsker's constant.

# 1 Introduction

The paper deals with the estimation problem in the heteroscedastic non-parametric regression model

$$y_j = S(x_j) + \sigma_j(S) \xi_j, \quad (1.1)$$

where the design points  $x_j = j/n$ ,  $S(\cdot)$  is an unknown function to be estimated,  $(\xi_j)_{1 \leq j \leq n}$  is a sequence of centered independent random variables with unit variance and  $(\sigma_j(S))_{1 \leq j \leq n}$  are unknown scale functionals depending on the design points and the regression function  $S$ .

Typically, the notion of asymptotic optimality is associated with the optimal convergence rate of the minimax risk (see e.g., Ibragimov, Hasminskii, 1981; Stone, 1982). An important question in optimality results is to study the exact asymptotic behavior of the minimax risk. Such results have been obtained only in a limited number of investigations. As to the nonparametric estimation problem for heteroscedastic regression models we should mention the papers by Efroimovich, 2007, Efroimovich, Pinsker, 1996, and Galtchouk, Pergamenshchikov, 2005, concerning the exact asymptotic behavior of the  $\mathcal{L}_2$ -risk and the paper by Brua, 2007, devoted to the efficient pointwise estimation for heteroscedastic regressions.

Heteroscedastic regression models are extensively used in the financial mathematics, in particular, in problems of calibrating (see e.g., Belomestny, Reiss, 2006). Moreover, these models are popular in econometrics (see, for e.g., Goldfeld, Quandt, 1972, p. 83), which, for example, for consumer budget problems, makes use of some semiparametric version of model (1.1) with the scale coefficients of type

$$\sigma_j^2(S) = c_0 + c_1 x_j + c_2 S^2(x_j) + c_3 \int_0^1 S^2(t) dt, \quad (1.2)$$

where  $(c_i)_{0 \leq i \leq 3}$  are some unknown positive constants.

The goal of this paper is to study asymptotic properties of the adaptive estimation procedure proposed in Galtchouk, Pergamenshchikov, 2007, for which a non-asymptotic oracle inequality was proved for quadratic risks. More precisely, in this paper we show that this procedure is efficient under some conditions on the scale functions  $(\sigma_j(S))_{1 \leq j \leq n}$  which hold for the functions (1.2). Note that in Efroimovich, 2007, Efroimovich, Pinsker, 1996, an efficient adaptive procedure is constructed for heteroscedastic regression

when the scale coefficient is independent of  $S$ , i.e.  $\sigma_j(S) = \sigma_j$ . In Galtchouk, Pergamenshchikov, 2005, for the model (1.1), the asymptotic efficiency was proved under conditions which are not satisfied in the case (1.2). Moreover, in these papers the efficiency is proved only for the gaussian random variables  $(\xi_j)_{1 \leq j \leq n}$  in model (1.1). This is very restrictive condition for applications.

In this paper we consider the robust quadratic risk, i.e. in the definition of the risk we take the additional supremum over the family of unknown noise distributions likely to Galtchouk, Pergamenshchikov, 2006. This modification allows us to eliminate from the risk dependence on the noise distribution. Moreover, for this risk the efficient procedure is robust with respect to changing the noise distribution.

As is well known, to prove the asymptotic efficiency one has to show that the asymptotic quadratic risk coincides with the lower bound which is equal to the Pinsker constant. In the paper two problems are resolved: in the first one a upper bound for the risk is obtained by making use of the non-asymptotic oracle inequality from Galtchouk, Pergamenshchikov, 2007, in the second one we prove that this upper bound coincides with the Pinsker constant. Let us remember that the adaptive procedure proposed in Galtchouk, Pergamenshchikov, 2007, is based on weighted least squares estimators, where the weights are proper modifications of the Pinsker weights for the homogeneous case (when  $\sigma_1(S) = \dots = \sigma_n(S) = 1$ ) relative to a certain smoothness of the function  $S$  and this procedure chooses a best estimator for the quadratic risk among these estimators. To obtain the Pinsker constant for the model (1.1) one has to prove a sharp asymptotic lower bound for the quadratic risk in the case when the noise variance depends on the unknown regression function. In this case, as usually, we minorize the minimax risk by a bayesian one for a respective parametric family. Then for the bayesian risk we make use of a lower bound (see Theorem 6.1) which is a modification of the van Trees inequality (see, Gill, Levit, 1995).

The paper is organized as follows. In Section 2 we construct an adaptive estimation procedure. Section 3 contains the principal conditions. The statements of our major results (the upper and lower bounds for quadratic risks) are presented in Section 4. The upper bound is proved in Section 5. In Section 6 we give all main steps of proving the lower bound: in Subsection 6.1 we find the lower bound for the bayesian risk for a parametric regression model which minorizes the minimax risk; in Subsection 6.2 we study a special family of parametric functions used to define the bayesian risk; in Subsection 6.3

we choose a prior distribution for bayesian risk to maximize the lower bound. Section 7 is devoted to explain how to use the given procedure in the case, when the unknown regression function is non periodic. In Section 8 we discuss the main results and their practical importance. The proofs are given in Section 9.

## 2 Adaptive procedure

In this section we describe the adaptive procedure proposed in Galtchouk, Pergamenschikov, 2006. To evaluate the error of estimation in the model (1.1) we make use of the empiric quadratic norm in the Hilbert space  $\mathcal{L}_2[0, 1]$  generated by the design points  $(x_j)_{1 \leq j \leq n}$  of model (1.1). To this end, for any functions  $u$  and  $v$  from  $\mathcal{L}_2[0, 1]$ , we define the empiric inner product

$$(u, v)_n = \frac{1}{n} \sum_{l=1}^n u(x_l) v(x_l). \quad (2.1)$$

Therefore, the estimation error of an estimator  $\widehat{S}$  of  $S$  will be evaluated by the empiric quadratic loss function

$$\|\widehat{S} - S\|_n^2 = \frac{1}{n} \sum_{l=1}^n (\widehat{S}(x_l) - S(x_l))^2.$$

Moreover, we make use of this inner product for vectors in  $\mathbb{R}^n$  as well, i.e. if  $u = (u_1, \dots, u_n)'$  and  $v = (v_1, \dots, v_n)'$ , then

$$(u, v)_n = \frac{1}{n} u'v = \frac{1}{n} \sum_{l=1}^n u_l v_l.$$

The prime denotes the transposition.

Let now  $(\phi_j)_{j \geq 1}$  be the standard trigonometric basis in  $\mathcal{L}_2[0, 1]$ , i.e.

$$\phi_1(x) = 1, \quad \phi_j(x) = \sqrt{2} Tr_j(2\pi[j/2]x), \quad j \geq 2, \quad (2.2)$$

where the function  $Tr_j(x) = \cos(x)$  for even  $j$  and  $Tr_j(x) = \sin(x)$  for odd  $j$ ;  $[x]$  denotes the integer part of  $x$ .

Notice that if  $n$  is odd, then the functions  $(\phi_j)_{1 \leq j \leq n}$  are orthonormal with respect to the empiric inner product (2.1), i.e. for any  $1 \leq i, j \leq n$ ,

$$(\phi_i, \phi_j)_n = \frac{1}{n} \sum_{l=1}^n \phi_i(x_l) \phi_j(x_l) = \mathbf{K}r_{ij}, \quad (2.3)$$

where  $\mathbf{K}r_{ij}$  is Kronecker's symbol,  $\mathbf{K}r_{ij} = 1$  if  $i = j$  and  $\mathbf{K}r_{ij} = 0$  for  $i \neq j$ .

**Remark 2.1.** *Note that in the case of even  $n$ , the basis (2.2) is orthogonal and it is orthonormal except of the  $n$ th function for which the normalizing constant should be changed. The corresponding modifications, for even  $n$ , one can see in Galtchouk, Pergamenschikov, 2005. To avoid these complications, we suppose  $n$  to be odd.*

Thanks to this basis we pass to the discrete Fourier transformation of model (1.1):

$$\widehat{\theta}_{j,n} = \theta_{j,n} + \frac{1}{\sqrt{n}} \xi_{j,n}, \quad (2.4)$$

where  $\widehat{\theta}_{j,n} = (Y, \phi_j)_n$ ,  $Y = (y_1, \dots, y_n)'$ ,  $\theta_{j,n} = (S, \phi_j)_n$  and

$$\xi_{j,n} = \frac{1}{\sqrt{n}} \sum_{l=1}^n \sigma_l(S) \xi_l \phi_j(x_l).$$

We estimate the function  $S$  by the weighted least squares estimator

$$\widehat{S}_\lambda = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \phi_j, \quad (2.5)$$

where the weight vector  $\lambda = (\lambda(1), \dots, \lambda(n))'$  belongs to some finite set  $\Lambda$  from  $[0, 1]^n$  with  $n \geq 3$ .

Here we make use of the weight family  $\Lambda$  introduced in Galtchouk, Pergamenschikov, 2009, i.e.

$$\Lambda = \{\lambda_\alpha, \alpha \in \mathcal{A}\}, \quad \mathcal{A} = \{1, \dots, k^*\} \times \{t_1, \dots, t_m\}, \quad (2.6)$$

where  $t_i = i\varepsilon$  and  $m = [1/\varepsilon^2]$ . We suppose that the parameters  $k^* \geq 1$  and  $0 < \varepsilon \leq 1$  are functions of  $n$ , i.e.  $k^* = k_n^*$  and  $\varepsilon = \varepsilon_n$ , such that,

$$\begin{cases} \lim_{n \rightarrow \infty} k_n^* = +\infty, & \lim_{n \rightarrow \infty} \frac{k_n^*}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \varepsilon_n = 0 & \text{and} \quad \lim_{n \rightarrow \infty} n^\nu \varepsilon_n = +\infty, \end{cases} \quad (2.7)$$

for any  $\nu > 0$ . For example, one can take for  $n \geq 3$

$$\varepsilon_n = 1/\ln n \quad \text{and} \quad k_n^* = \bar{k} + \sqrt{\ln n},$$

where  $\bar{k}$  is any nonnegative constant.

For each  $\alpha = (\beta, t) \in \mathcal{A}$  we define the weight vector  $\lambda_\alpha = (\lambda_\alpha(1), \dots, \lambda_\alpha(n))'$  as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + (1 - (j/\omega(\alpha))^\beta) \mathbf{1}_{\{j_0 < j \leq \omega(\alpha)\}}. \quad (2.8)$$

Here  $j_0 = j_0(\alpha) = [\omega(\alpha) \varepsilon_n]$  with

$$\omega(\alpha) = \bar{\omega} + (A_\beta t)^{1/(2\beta+1)} n^{1/(2\beta+1)}, \quad (2.9)$$

where  $\bar{\omega}$  is any nonnegative constant and

$$A_\beta = \frac{(\beta + 1)(2\beta + 1)}{\beta \pi^{2\beta}}.$$

**Remark 2.2.** Note that, the weighted least squares estimators (2.5) have been introduced by Pinsker, 1981, for the optimal filtering of a continuous time signal in the gaussian noise. It turned out, that the asymptotic quadratic risk for estimators of type (2.5) – (2.8) is minimal over all possible estimators. This sharp value of the asymptotic quadratic risk is called the Pinsker constant. Nussbaum, 1985, makes use of the same method with proper modification for efficient estimation of the function  $S$  of known smoothness in the homogeneous gaussian model (1.1), i.e. when  $\sigma_1(S) = \dots = \sigma_n(S) = 1$  and  $(\xi_j)_{1 \leq j \leq n}$  is a sequence of i.i.d.  $\mathcal{N}(0, 1)$  random variables.

To choose weights from the set (2.6) we minimize the special cost function introduced by Galtchouk, Pergamenschikov, 2007. This cost function is as follows

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \tilde{\theta}_{j,n} + \rho \hat{P}_n(\lambda), \quad (2.10)$$

where

$$\tilde{\theta}_{j,n} = \hat{\theta}_{j,n}^2 - \frac{1}{n} \hat{\varsigma}_n \quad \text{with} \quad \hat{\varsigma}_n = \sum_{j=l_n+1}^n \hat{\theta}_{j,n}^2 \quad (2.11)$$

and  $l_n = [n^{1/3} + 1]$ . The penalty term we define as

$$\hat{P}_n(\lambda) = \frac{|\lambda|^2 \hat{\varsigma}_n}{n}, \quad |\lambda|^2 = \sum_{j=1}^n \lambda^2(j) \quad \text{and} \quad \rho = \frac{1}{3 + L_n},$$

where  $L_n \geq 0$  is any slowly increasing sequence, i.e.

$$\lim_{n \rightarrow \infty} L_n = +\infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{L_n}{n^\nu} = 0, \quad (2.12)$$

for any  $\nu > 0$ .

Finally, we set

$$\hat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda) \quad \text{and} \quad \hat{S}_* = \hat{S}_{\hat{\lambda}}. \quad (2.13)$$

The goal of this paper is to study asymptotic properties (as  $n \rightarrow \infty$ ) of this estimation procedure.

**Remark 2.3.** *Now we explain why does one choose the cost function in the form (2.10). Developing the empiric quadratic loss function for the estimator (2.5), one obtains*

$$\|\hat{S}_\lambda - S\|_n^2 = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \theta_{j,n} + \|S\|_n^2.$$

*It's natural to choose the weight vector  $\lambda$  for which this function reaches the minimum. Since the last term on the right-hand part is independent of  $\lambda$ , it can be dropped and one has to minimize with respect to  $\lambda$  the function equals to the difference of the two first terms on the right-hand part. It's clear that the minimization problem can not be solved directly because the Fourier coefficients  $(\theta_{j,n})$  are unknown. To overcome this difficulty, we replace the product  $\hat{\theta}_{j,n} \theta_{j,n}$  by its asymptotically unbiased estimator  $\tilde{\theta}_{j,n}$ . Moreover, to pay this substitution, we introduce into the cost function the penalty term  $\hat{P}_n$  with a small coefficient  $\rho > 0$ . The form of the penalty term is provided by the principal term of the quadratic risk for weighted least-squares estimator, see Galtchouk, Pergamenshchikov, 2007, 2009. The coefficient  $\rho > 0$  is small, because the estimator  $\tilde{\theta}_{j,n}$  is close in mean to the quantity  $\theta_{j,n} \theta_{j,n}$  asymptotically, as  $n \rightarrow \infty$ .*

*Note that the principal difference between the procedure (2.13) and the adaptive procedure proposed by Golubev, Nussbaum, 1993, for a homogeneous gaussian regression, consists in presence of the penalty term in the cost function (2.10).*

**Remark 2.4.** *As it was noted in Remark 2.2, Nussbaum, 1985, has shown that the weight coefficients of type (2.8) provide the asymptotic minimum of*

the quadratic risk in the estimation problem of the regression function for the homogeneous gaussian model (1.1), when the smoothness of the function  $S$  is known. In fact, to obtain an efficient estimator one needs to take a weighted least squares estimator (2.5) with the weight vector  $\lambda_\alpha$ , where the index  $\alpha$  depends on the smoothness of the function  $S$  (the parameters  $k$  and  $r$  in (3.1)) and on the coefficients  $(\sigma_j(S))_{1 \leq j \leq n}$ , (the parameter  $\varsigma(S)$  in (3.6)), which are unknown in our case. For this reason, Galtchouk, Pergamenshchikov, 2007, have proposed to make use of the family of coefficients (2.6), which contains the weight vector providing the minimum of the quadratic risk. Indeed, the set (2.6) is a two-dimensional grille giving all possible values for the parameters  $k$  and  $\bar{r}(S) = r/\varsigma(S)$  in the estimator (2.5) with the weight (2.8) which is efficient if the parameters  $k$  and  $\bar{r}(S)$  are known, i.e. the set  $\mathcal{A}$  gives the family of efficient estimators. Moreover, under some weak conditions on the coefficients  $(\sigma_j(S))_{1 \leq j \leq n}$ , Galtchouk, Pergamenshchikov, 2007, shown that the procedure (2.13) is best in the class of these estimators in the sens of the non-asymptotic oracle inequality (see, Theorem 4.1 below). It is important to note that, due to the conditions (2.7) for the set  $\mathcal{A}$ , the secondary term in the oracle inequality is slowly increasing (slower than any degree of  $n$ ).

### 3 Conditions

First we impose some conditions on the function  $S$  in the model (1.1).

Let  $\mathcal{C}_{per,1}^k(\mathbb{R})$  be the set of 1-periodic  $k$  times differentiable  $\mathbb{R} \rightarrow \mathbb{R}$  functions. We assume that  $S$  belongs to the following set:

$$W_r^k = \{f \in \mathcal{C}_{per,1}^k(\mathbb{R}) : \sum_{j=0}^k \|f^{(j)}\|^2 \leq r\}, \quad (3.1)$$

where  $\|\cdot\|$  denotes the norm in  $\mathcal{L}_2[0, 1]$ , i.e.

$$\|f\|^2 = \int_0^1 f^2(t) dt. \quad (3.2)$$

Moreover, we suppose that  $r > 0$  and  $k \geq 1$  are unknown parameters.

Note that the set  $W_r^k$  can be represented as an ellipse in  $\mathcal{L}_2[0, 1]$ , i.e.

$$W_r^k = \{f \in \mathcal{L}_2[0, 1] : \sum_{j=1}^{\infty} a_j \theta_j^2 \leq r\}, \quad (3.3)$$

where

$$\theta_j = (f, \phi_j) = \int_0^1 f(t)\phi_j(t)dt \quad (3.4)$$

and

$$a_j = \sum_{l=0}^k \|\phi_j^{(l)}\|^2 = \sum_{i=0}^k (2\pi[j/2])^{2i}. \quad (3.5)$$

Here  $(\phi_j)_{j \geq 1}$  is the trigonometric basis defined in (2.2).

Now we describe the conditions on the scale coefficients  $(\sigma_j(S))_{j \geq 1}$ .

**H<sub>1</sub>)**  $\sigma_j(S) = g(x_j, S)$  for some unknown function  $g : [0, 1] \times \mathcal{L}_1[0, 1] \rightarrow \mathbb{R}_+$ , which is square integrable with respect to  $x$  and such that

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} \left| \frac{1}{n} \sum_{j=1}^n g^2(x_j, S) - \varsigma(S) \right| = 0, \quad (3.6)$$

where  $\varsigma(S) := \int_0^1 g^2(x, S)dx$ . Moreover,

$$g_* = \inf_{0 \leq x \leq 1} \inf_{S \in W_r^k} g^2(x, S) > 0 \quad (3.7)$$

and

$$\sup_{S \in W_r^k} \varsigma(S) < \infty. \quad (3.8)$$

**H<sub>2</sub>)** For any  $x \in [0, 1]$ , the operator  $g^2(x, \cdot) : \mathcal{C}[0, 1] \rightarrow \mathbb{R}$  is differentiable in the Fréchet sense at any fixed function  $f_0$  from  $\mathcal{C}[0, 1]$ , i.e. for any  $f$  from some vicinity of  $f_0$  in  $\mathcal{C}[0, 1]$ ,

$$g^2(x, f) = g^2(x, f_0) + \mathbf{L}_{x, f_0}(f - f_0) + \Upsilon(x, f_0, f),$$

where the Fréchet derivative  $\mathbf{L}_{x, f_0} : \mathcal{C}[0, 1] \rightarrow \mathbb{R}$  is a bounded linear operator and the residual term  $\Upsilon(x, f_0, f)$ , for each  $x \in [0, 1]$ , satisfies the following property:

$$\lim_{\|f - f_0\|_\infty \rightarrow 0} \frac{|\Upsilon(x, f_0, f)|}{\|f - f_0\|_\infty} = 0,$$

where  $\|f\|_\infty = \sup_{0 \leq t \leq 1} |f(t)|$ .

**H<sub>3</sub>)** *There exists some positive constant  $C^*$  such that, for any function  $S$  from  $\mathcal{C}[0, 1]$ , the operator  $\mathbf{L}_{x,S}$  defined in the condition **H<sub>2</sub>)** satisfies the following inequality, for any function  $f$  from  $\mathcal{C}[0, 1]$ :*

$$|\mathbf{L}_{x,S}(f)| \leq C^* (|S(x)f(x)| + |f|_1 + \|S\| \|f\|) , \quad (3.9)$$

where  $|f|_1 = \int_0^1 |f(t)| dt$ .

**H<sub>4</sub>)** *The function  $g_0(\cdot) = g(\cdot, S_0)$  corresponding to  $S_0 \equiv 0$  is continuous on the interval  $[0, 1]$ . Moreover,*

$$\lim_{\delta \rightarrow 0} \sup_{0 \leq x \leq 1} \sup_{\|S\|_\infty \leq \delta} |g(x, S) - g(x, S_0)| = 0.$$

**Remark 3.1.** *Let us explain the conditions **H<sub>1</sub>)**-**H<sub>4</sub>)** which are the regularity conditions of the function  $g(x, S)$  generating the scale coefficients  $(\sigma_j(S))_{1 \leq j \leq n}$ .*

*Condition **H<sub>1</sub>)** means that the function  $g(\cdot, S)$  must be uniformly integrable with respect to the first argument in the sense of convergence (3.6). Moreover, this function must be separated from zero (see inequality (3.7)) and bounded on the class (3.1) (see inequality (3.8)). Boundedness away from zero provides that the distribution of observations  $(y_j)_{1 \leq j \leq n}$  isn't degenerate in  $\mathbb{R}^n$ , and the boundedness means that the intensity of the noise vector must be finite, otherwise the estimation problem has not any sense.*

*Conditions **H<sub>2</sub>)** and **H<sub>3</sub>)** mean that the function  $g(x, \cdot)$  is regular, at any fixed  $0 \leq x \leq 1$ , with respect to  $S$  in the sense, that it is differentiable in the Frèchet sense (see e.g., Kolmogorov, Fomin, 1989) and, moreover, the Frèchet derivative satisfies the growth condition given by the inequality (3.9) which permits to consider the example (1.2).*

*The last condition **H<sub>4</sub>)** is the usual uniform continuity condition of the function  $g(\cdot, \cdot)$  at the point  $S \equiv 0$ .*

One check directly that the function (1.2) satisfies the conditions **H<sub>1</sub>)**-**H<sub>4</sub>)**. Another functions satisfying these conditions are given in Galtchouk, Pergamenshchikov , 2008.

## 4 Main results

Denote by  $\mathcal{P}_n$  the family of distributions  $p$  in  $\mathbb{R}^n$  of the vectors  $(\xi_1, \dots, \xi_n)'$  in the model (1.1) such that the components  $\xi_j$  are jointly independent, centered with unit variance and

$$\max_{1 \leq k \leq n} \mathbf{E} \xi_k^4 \leq l_n^*, \quad (4.1)$$

where  $l_n^* \geq 3$  is slowly increasing sequence, i.e. it satisfies the condition (2.12). It is easy to see that, for any  $n \geq 1$ , the centered gaussian distribution in  $\mathbb{R}^n$  with unit covariation matrix belongs to the family  $\mathcal{P}_n$ . We will denote by  $q$  this gaussian distribution.

For any estimator  $\widehat{S}$ , we define the following quadratic risk

$$\mathcal{R}_n(\widehat{S}, S) = \sup_{p \in \mathcal{P}_n} \mathbf{E}_{S,p} \|\widehat{S} - S\|_n^2, \quad (4.2)$$

where  $\mathbf{E}_{S,p}$  is the expectation with respect to the distribution  $\mathbf{P}_{S,p}$  of the observations  $(y_1, \dots, y_n)$  with the fixed function  $S$  and the fixed distribution  $p \in \mathcal{P}_n$  of random variables  $(\xi_j)_{1 \leq j \leq n}$  in the model (1.1).

Moreover, to make the risk independent of the design, we will make use of the risk with respect to the usual norm in  $\mathcal{L}_2[0, 1]$  (3.2) too, i.e.

$$\mathcal{T}_n(\widehat{S}, S) = \sup_{p \in \mathcal{P}_n} \mathbf{E}_{S,p} \|\widehat{S} - S\|^2. \quad (4.3)$$

If an estimator  $\widehat{S}$  is defined only at the design points  $(x_j)_{1 \leq j \leq n}$ , then we extend it as step function onto the interval  $[0, 1]$  by setting  $\widehat{S}(x) = T(\widehat{S}(x))$ , for all  $0 \leq x \leq 1$ , where

$$T(f)(x) = f(x_1) \mathbf{1}_{[0, x_1]}(x) + \sum_{k=2}^n f(x_k) \mathbf{1}_{(x_{k-1}, x_k]}(x). \quad (4.4)$$

In Galtchouk, Pergamenshchikov, 2007, 2009 the following non-asymptotic oracle inequality has been proved for the procedure (2.13) .

**Theorem 4.1.** *Assume that  $S \in W_r^1$  for some  $r > 0$ . Let  $\widehat{S}_*$  be from (2.13). Then, for any odd  $n \geq 3$ , the following oracle inequality holds:*

$$\mathcal{R}_n(\widehat{S}_*, S) \leq \frac{1 + 3\rho - 2\rho^2}{1 - 3\rho} \min_{\lambda \in \Lambda} \mathcal{R}_n(\widehat{S}_\lambda, S) + \frac{1}{n} \mathcal{B}_n(\rho), \quad (4.5)$$

where  $\rho = 1/(3 + L_n)$ ,  $L_n$  is from (2.12), the function  $\mathcal{B}_n(\rho)$  is such that, for any  $\nu > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{\mathcal{B}_n(\rho)}{n^\nu} = 0. \quad (4.6)$$

**Remark 4.1.** Note that in Galtchouk, Pergamenshchikov, 2007, 2009, the oracle inequality is proved for the model (1.1), where the random variables  $(\xi_j)_{1 \leq j \leq n}$  are independent identically distributed. In fact, the result is true for independent random variables which are not identically distributed, i.e. for any distribution of the random vector  $(\xi_1, \dots, \xi_n)'$  from  $\mathcal{P}_n$ .

Now we formulate the main asymptotic results. To this end, for any function  $S \in W_r^k$ , we set

$$\gamma_k(S) = \Gamma_k^* r^{-(2k+1)} (\varsigma(S))^{2k/(2k+1)}, \quad (4.7)$$

where

$$\Gamma_k^* = (2k + 1)^{-(2k+1)} (k/(\pi(k + 1)))^{2k/(2k+1)}.$$

It is well known (see e.g., Nussbaum, 1985) that the optimal rate of convergence is  $n^{2k/(2k+1)}$  when the risk is taken uniformly over  $W_r^k$ .

**Theorem 4.2.** Assume that in the model (1.1) the sequence  $(\sigma_j(S))$  satisfies the condition  $\mathbf{H}_1$ ). Then, for any integer  $k \geq 1$  and  $r > 0$ , the estimator  $\widehat{S}_*$  from (2.13) satisfies the inequalities

$$\limsup_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \sup_{S \in W_r^k} \frac{\mathcal{R}_n(\widehat{S}_*, S)}{\gamma_k(S)} \leq 1 \quad (4.8)$$

and

$$\limsup_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \sup_{S \in W_r^k} \frac{\mathcal{T}_n(\widehat{S}_*, S)}{\gamma_k(S)} \leq 1. \quad (4.9)$$

The following result gives the sharp lower bound for the risk (4.2) and show that  $\gamma_k(S)$  is the Pinsker constant.

**Theorem 4.3.** Assume that in the model (1.1) the sequence  $(\sigma_j(S))$  satisfies the conditions  $\mathbf{H}_2$ )-  $\mathbf{H}_4$ ). Then, for any integer  $k \geq 1$  and  $r > 0$ , the risks (4.2) and (4.3) admit the following asymptotic lower bounds, respectively,

$$\liminf_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \inf_{\widehat{S}_n} \sup_{S \in W_r^k} \frac{\mathcal{R}_n(\widehat{S}_n, S)}{\gamma_k(S)} \geq 1 \quad (4.10)$$

and

$$\liminf_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \inf_{\widehat{S}_n} \sup_{S \in W_r^k} \frac{\mathcal{T}_n(\widehat{S}_n, S)}{\gamma_k(S)} \geq 1. \quad (4.11)$$

**Remark 4.2.** *To obtain the non-asymptotic oracle inequality (4.5), it is not necessary to make use of equidistant design points and the trigonometric basis. One may take any design points (deterministic or random) and any orthonormal basis satisfying (2.3). But to obtain the property (4.6) one needs to impose some technical conditions (see Galtchouk, Pergamenschikov, 2009).*

*Note that the results of Theorem 4.2 and Theorem 4.3 are based on equidistant design points and the trigonometric basis.*

## 5 Upper bound

In this section we prove Theorem 4.2. To this end we will make use of the oracle inequality (4.5). We have to find an estimator from the family (2.5)-(2.6) for which we can prove the upper bound (4.8). We start with the construction of such an estimator. First we put

$$\tilde{l}_n = \inf\{i \geq 1 : i\varepsilon \geq \bar{r}(S)\} \wedge m \quad \text{and} \quad \bar{r}(S) = r/\varsigma(S), \quad (5.1)$$

where  $a \wedge b = \min(a, b)$ . Then we choose an index from the set  $\mathcal{A}$  as

$$\tilde{\alpha} = (k, \tilde{t}_n), \quad (5.2)$$

where  $k$  is the parameter of the set  $W_r^k$  and  $\tilde{t}_n = \tilde{l}_n \varepsilon$ . Finally, we set

$$\tilde{S} = \widehat{S}_{\tilde{\lambda}} \quad \text{and} \quad \tilde{\lambda} = \lambda_{\tilde{\alpha}}. \quad (5.3)$$

Now we formulate the upper bound (4.8) for this estimator.

**Theorem 5.1.** *Assume that the condition  $\mathbf{H}_1$ ) holds. Then, for any integer  $k \geq 1$  and  $r > 0$ ,*

$$\limsup_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \sup_{S \in W_r^k} \frac{\mathcal{R}_n(\tilde{S}, S)}{\gamma_k(S)} \leq 1. \quad (5.4)$$

**Remark 5.1.** *Note that the estimator  $\tilde{S}$  belongs to the family (2.5)-(2.6), but we can not use directly this estimator because the parameters  $k$ ,  $r$  and  $\bar{r}(S)$  are unknown. We can use this upper bound only via the oracle inequality (4.5) proved for the procedure (2.13).*

Now Theorem 4.1 and Theorem 5.1 imply the upper bound (4.8). To obtain the upper bound (4.9) one needs the following auxiliary result.

**Lemma 5.2.** *For any  $0 < \delta < 1$  and any estimator  $\hat{S}_n$  of  $S \in W_r^k$ ,*

$$\|\hat{S}_n - S\|_n^2 \geq (1 - \delta)\|T_n(\hat{S}) - S\|^2 - (\delta^{-1} - 1)r/n^2,$$

where the function  $T_n(\hat{S})(\cdot)$  is defined in (4.4).

Proof of this lemma is given in Galtchouk, Pergamenshchikov, 2008 ( Appendix A.1).

Now the inequality (4.8) and this lemma imply the upper bound (4.9). Hence Theorem 4.2.

## 6 Lower bound

In this section we give the main steps of proving the lower bounds (4.10) and (4.11). We follow the commonly used scheme (see e.g. Nussbaum, 1985). We begin with minorizing the minimax risk by a bayesian one constructed on a parametric functional family introduced in Section 6.2 ( see (6.9)) and using the prior distribution (6.10). Further, a special modification of the van Trees inequality (see, Theorem 6.1) yields a lower bound for the bayesian risk depending on the chosen prior distribution, of course. Finally, in Section 6.3, we choose parameters of the prior distribution (see (6.10)) providing the maximal value of the lower bound for the bayesian risk. This value coincides with the Pinsker constant as it is shown in Section 9.4. We emphasize that, by making use of the bayesian risk, one passes from the nonparametric regression model to a parametric one, for which the van Trees inequality holds. Note that this inequality is an extension of the Cramer-Rao inequality and gives a lower bound for the bayesian risk.

## 6.1 Lower bound for parametric heteroscedastic regression models

Let  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P}_\vartheta, \vartheta \in \Theta \subseteq \mathbb{R}^l)$  be the statistical model relative to the observations  $(y_j)_{1 \leq j \leq n}$  governed by the regression equation

$$y_j = S_\vartheta(x_j) + \sigma_j(\vartheta) \xi_j, \quad (6.1)$$

where  $\xi_1, \dots, \xi_n$  are i.i.d.  $\mathcal{N}(0, 1)$  random variables,  $\vartheta = (\vartheta_1, \dots, \vartheta_l)'$  is a unknown parameter vector,  $S_\vartheta(x)$  is a unknown (or known) function and  $\sigma_j(\vartheta) = g(x_j, S_\vartheta)$ , with the function  $g(x, S)$  defined in the condition  $\mathbf{H}_1$ . Assume that a prior distribution  $\mu_\vartheta$  of the parameter  $\vartheta$  in  $\mathbb{R}^l$  is defined by the density  $\Phi(\cdot)$  of the following form

$$\Phi(z) = \Phi(z_1, \dots, z_l) = \prod_{i=1}^l \varphi_i(z_i),$$

where  $\varphi_i$  is a continuously differentiable bounded density on  $\mathbb{R}$  with

$$I_i = \int_{\mathbb{R}} \frac{\dot{\varphi}_i^2(u)}{\varphi_i(u)} du < \infty.$$

Let  $\tau(\cdot)$  be a continuously differentiable  $\mathbb{R}^l \rightarrow \mathbb{R}$  function such that, for any  $1 \leq i \leq l$ ,

$$\lim_{|z_i| \rightarrow \infty} \tau(z) \varphi_i(z_i) = 0 \quad \text{and} \quad \int_{\mathbb{R}^l} |\tau'_i(z)| \Phi(z) dz < \infty, \quad (6.2)$$

where

$$\tau'_i(z) = (\partial/\partial z_i) \tau(z).$$

Let  $\hat{\tau}_n$  be an estimator of  $\tau(\vartheta)$  based on observations  $(y_j)_{1 \leq j \leq n}$ . For any  $\mathcal{B}(\mathbb{R}^n \times \mathbb{R}^l)$  - measurable integrable function  $G(x, z)$ ,  $x \in \mathbb{R}^n$ ,  $z \in \mathbb{R}^l$ , we set

$$\tilde{\mathbf{E}} G(Y, \vartheta) = \int_{\mathbb{R}^l} \mathbf{E}_z G(Y, z) \Phi(z) dz,$$

where  $\mathbf{E}_\vartheta$  is the expectation with respect to the distribution  $\mathbf{P}_\vartheta$  of the vector  $Y = (y_1, \dots, y_n)$ . Note that in this case

$$\mathbf{E}_\vartheta G(Y, \vartheta) = \int_{\mathbb{R}^n} G(v, \vartheta) f(v, \vartheta) dv,$$

where

$$f(v, z) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j(z)} \exp \left\{ -\frac{(v_j - S_z(x_j))^2}{2\sigma_j^2(z)} \right\}. \quad (6.3)$$

We prove the following result.

**Theorem 6.1.** *Assume that the conditions  $\mathbf{H}_1) - \mathbf{H}_2)$  hold. Moreover, assume that the function  $S_z(\cdot)$  with  $z = (z_1, \dots, z_l)'$  is uniformly over  $0 \leq x \leq 1$  differentiable with respect to  $z_i$ ,  $1 \leq i \leq l$ , i.e. for any  $1 \leq i \leq l$ , there exists a function  $S'_{z,i} \in C[0, 1]$  such that*

$$\lim_{h \rightarrow 0} \max_{0 \leq x \leq 1} \left| \left( S_{z+h\mathbf{e}_i}(x) - S_z(x) - S'_{z,i}(x)h \right) / h \right| = 0, \quad (6.4)$$

where  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)'$ , all coordinates are 0, except the  $i$ -th equals to 1. Then for any square integrable estimator  $\hat{\tau}_n$  of  $\tau(\vartheta)$  and any  $1 \leq i \leq l$ ,

$$\tilde{\mathbf{E}}(\hat{\tau}_n - \tau(\vartheta))^2 \geq \frac{\bar{\tau}_i^2}{F_i + B_i + I_i}, \quad (6.5)$$

where  $\bar{\tau}_i = \int_{\mathbb{R}^l} \tau'_i(z) \Phi(z) dz$ ,  $F_i = \sum_{j=1}^n \int_{\mathbb{R}^l} (S'_{z,i}(x_j)/\sigma_j(z))^2 \Phi(z) dz$  and

$$B_i = \frac{1}{2} \sum_{j=1}^n \int_{\mathbb{R}^l} \frac{\tilde{\mathbf{L}}_i^2(x_j, S_z)}{\sigma_j^4(S_z)} \Phi(z) dz,$$

$\tilde{\mathbf{L}}_i(x, z) = \mathbf{L}_{x, S_z}(S'_{z,i})$ , the operator  $\mathbf{L}_{x, S}$  is defined in the condition  $\mathbf{H}_2)$ .

**Remark 6.1.** *Note that the inequality (6.5) is some modification of the van Trees inequality (see, Gill, Levit, 1995) adapted to the model (6.1).*

## 6.2 Parametric family of kernel functions

In this section we define and study some special parametric family of kernel function which will be used to prove the sharp lower bound (4.10).

Let us begin by kernel functions. We fix  $\eta > 0$  and we set

$$\chi_\eta(x) = \eta^{-1} \int_{\mathbb{R}} \mathbf{1}_{(|u| \leq 1-\eta)} V \left( \frac{u-x}{\eta} \right) du, \quad (6.6)$$

where  $\mathbf{1}_A$  is the indicator of a set  $A$ , the kernel  $V \in \mathcal{C}^\infty(\mathbb{R})$  is such that

$$V(u) = 0 \quad \text{for } |u| \geq 1 \quad \text{and} \quad \int_{-1}^1 V(u) du = 1.$$

It is easy to see that the function  $\chi_\eta(x)$  possesses the properties :

$$\begin{aligned} 0 \leq \chi_\eta \leq 1, \quad \chi_\eta(x) = 1 \quad \text{for } |x| \leq 1 - 2\eta \quad \text{and} \\ \chi_\eta(x) = 0 \quad \text{for } |x| \geq 1. \end{aligned}$$

Moreover, for any  $c > 0$  and  $\nu \geq 0$

$$\lim_{\eta \rightarrow 0} \sup_{f: \|f\|_\infty \leq c} \left| \int_{\mathbb{R}} f(x) \chi_\eta^\nu(x) dx - \int_{-1}^1 f(x) dx \right| = 0. \quad (6.7)$$

We divide the interval  $[0, 1]$  into  $M$  equal subintervals of length  $2h$  and on each of them we construct a kernel-type function which equals to zero at the boundary of the subinterval together with all derivatives. This provides that the Fourier partial sums with respect to the trigonometric basis in  $\mathcal{L}_2[-1, 1]$  give a natural parametric approximation to the function on each subinterval.

Let  $(e_j)_{j \geq 1}$  be the trigonometric basis in  $\mathcal{L}_2[-1, 1]$ , i.e.

$$e_1 = 1/\sqrt{2}, \quad e_j(x) = Tr_j(\pi[j/2]x), \quad j \geq 2, \quad (6.8)$$

where the functions  $(Tr_j)_{j \geq 2}$  are defined in (2.2).

Now, for any array  $z = \{(z_{m,j})_{1 \leq m \leq M_n, 1 \leq j \leq N_n}\}$  we define the following function

$$S_{z,n}(x) = \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} z_{m,j} D_{m,j}(x), \quad (6.9)$$

where  $D_{m,j}(x) = e_j(v_m(x)) \chi_\eta(v_m(x))$ ,

$$v_m(x) = \frac{x - \tilde{x}_m}{h_n}, \quad \tilde{x}_m = 2mh_n \quad \text{and} \quad M_n = [1/(2h_n)] - 1.$$

We assume that the sequences  $(N_n)_{n \geq 1}$  and  $(h_n)_{n \geq 1}$ , satisfy the following conditions.

**A<sub>1</sub>)** The sequence  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$  and, for any  $\nu > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{N_n^\nu}{n} = 0.$$

Moreover, there exist  $0 < \delta_1 < 1$  and  $\delta_2 > 0$  such that

$$h_n = O(n^{-\delta_1}) \quad \text{and} \quad h_n^{-1} = O(n^{\delta_2}) \quad \text{as} \quad n \rightarrow \infty.$$

To define a prior distribution on the family of arrays, we choose the following random array  $\vartheta = \{(\vartheta_{m,j})_{1 \leq m \leq M_n, 1 \leq j \leq N_n}\}$  with

$$\vartheta_{m,j} = t_{m,j} \zeta_{m,j}, \quad (6.10)$$

where  $(\zeta_{m,j})$  are i.i.d.  $\mathcal{N}(0, 1)$  random variables and  $(t_{m,j})$  are some non-random positive coefficients. We make use of gaussian variables since they possess the minimal Fisher information and therefore maximize the lower bound (6.5). We set

$$t_n^* = \max_{1 \leq m \leq M_n} \sum_{j=1}^{N_n} t_{m,j}. \quad (6.11)$$

We assume that the coefficients  $(t_{m,j})_{1 \leq m \leq M_n, 1 \leq j \leq N_n}$  satisfy the following conditions.

**A<sub>2</sub>)** *There exists a sequence of positive numbers  $(d_n)_{n \geq 1}$  such that*

$$\lim_{n \rightarrow \infty} \frac{d_n}{h_n^{2k-1}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 j^{2(k-1)} = 0, \quad \lim_{n \rightarrow \infty} \sqrt{d_n} t_n^* = 0, \quad (6.12)$$

moreover, for any  $\nu > 0$ ,

$$\lim_{n \rightarrow \infty} n^\nu \exp\{-d_n/2\} = 0.$$

**A<sub>3</sub>)** *For some  $0 < \epsilon < 1$*

$$\limsup_{n \rightarrow \infty} \frac{1}{h_n^{2k-1}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 j^{2k} \leq (1 - \epsilon)r \left(\frac{2}{\pi}\right)^{2k}.$$

**A<sub>4</sub>)** *There exists  $\epsilon_0 > 0$  such that*

$$\lim_{n \rightarrow \infty} \frac{1}{h_n^{4k-2+\epsilon_0}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^4 j^{4k} = 0.$$

**Proposition 6.2.** *Let the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_2)$ . Then, for any  $\nu > 0$  and for any  $\delta > 0$ ,*

$$\lim_{n \rightarrow \infty} n^\nu \max_{0 \leq l \leq k-1} \mathbf{P} \left( \|S_{\vartheta, n}^{(l)}\| > \delta \right) = 0.$$

**Proposition 6.3.** *Let the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$ . Then, for any  $\nu > 0$ ,*

$$\lim_{n \rightarrow \infty} n^\nu \mathbf{P}(S_{\vartheta, n} \notin W_r^k) = 0.$$

**Proposition 6.4.** *Let the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$ . Then, for any  $\nu > 0$ ,*

$$\lim_{n \rightarrow \infty} n^\nu \mathbf{E} \|S_{\vartheta, n}\|^2 \left( \mathbf{1}_{\{S_{\vartheta, n} \notin W_r^k\}} + \mathbf{1}_{\Xi_n^c} \right) = 0.$$

**Proposition 6.5.** *Let the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$ . Then, for any function  $g$  satisfying the conditions (3.7) and  $\mathbf{H}_4)$ ,*

$$\lim_{n \rightarrow \infty} \sup_{0 \leq x \leq 1} \mathbf{E} |g^{-2}(x, S_{\vartheta, n}) - g_0^{-2}(x)| = 0.$$

Proofs of Propositions 6.2–6.5 are given in Galtchouk, Pergamenshchikov, 2008 (Appendix A.2–A.6).

### 6.3 Bayes risk

Now we will obtain the lower bound for the bayesian risk that yields the lower bound (4.11) for the minimax risk.

We make use of the sequence of random functions  $(S_{\vartheta, n})_{n \geq 1}$  defined in (6.9)–(6.10) with the coefficients  $(t_{m, j})$  satisfying the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$  which will be chosen later.

For any estimator  $\widehat{S}_n$  we introduce now the corresponding Bayes risk

$$\mathcal{E}_n(\widehat{S}_n) = \int_{\mathbb{R}^l} \mathbf{E}_{S_{z, n}, q} \|\widehat{S}_n - S_{z, n}\|^2 \mu_\vartheta(dz), \quad (6.13)$$

where the kernel family  $(S_{z, n})$  is defined in (6.9),  $\mu_\vartheta$  denotes the distribution of the random array  $\vartheta$  defined by (6.10) in  $\mathbb{R}^l$  with  $l = M_n N_n$ .

We remember that  $q$  is a centered gaussian distribution in  $\mathbb{R}^n$  with unit covariation matrix.

First of all, we replace the functions  $\widehat{S}_n$  and  $S$  by their Fourier series with respect to the basis

$$\widetilde{e}_{m,i}(x) = (1/\sqrt{h}) e_i(v_m(x)) \mathbf{1}_{(|v_m(x)| \leq 1)}.$$

By making use of this basis we can estimate the norm  $\|\widehat{S}_n - S_{z,n}\|^2$  from below as

$$\|\widehat{S}_n - S_{z,n}\|^2 \geq \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} (\widehat{\tau}_{m,j} - \tau_{m,j}(z))^2,$$

where

$$\widehat{\tau}_{m,j} = \int_0^1 \widehat{S}_n(x) \widetilde{e}_{m,j}(x) dx \quad \text{and} \quad \tau_{m,j}(z) = \int_0^1 S_{z,n}(x) \widetilde{e}_{m,j}(x) dx.$$

Moreover, from the definition (6.9) one gets

$$\tau_{m,j}(z) = \sqrt{h} \sum_{i=1}^{N_n} z_{m,i} \int_{-1}^1 e_i(u) e_j(u) \chi_\eta(u) du.$$

It is easy to see that the functions  $\tau_{m,j}(\cdot)$  satisfy the condition (6.2) for gaussian prior densities. In this case (see the definition in (6.5)) we have

$$\bar{\tau}_{m,j} = (\partial/\partial z_{m,j}) \tau_{m,j}(z) = \sqrt{h} \bar{e}_j(\chi_\eta),$$

where

$$\bar{e}_j(f) = \int_{-1}^1 e_j^2(v) f(v) dv. \quad (6.14)$$

Now to obtain a lower bound for the Bayes risk  $\mathcal{E}_n(\widehat{S}_n)$  we make use of Theorem 6.1 which implies that

$$\mathcal{E}_n(\widehat{S}_n) \geq \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} \frac{h \bar{e}_j^2(\chi_\eta)}{F_{m,j} + B_{m,j} + t_{m,j}^{-2}}, \quad (6.15)$$

where  $F_{m,j} = \sum_{i=1}^n D_{m,j}^2(x_i) \mathbf{E} g^{-2}(x_i, S_{\vartheta,n})$  and

$$B_{m,j} = \frac{1}{2} \sum_{i=1}^n \mathbf{E} \frac{\widetilde{\mathbf{L}}_{m,j}^2(x_i, S_{\vartheta,n})}{g^4(x_i, S_{\vartheta,n})} \quad \text{with} \quad \widetilde{\mathbf{L}}_{m,j}(x, S) = \mathbf{L}_{x,S}(D_{m,j}).$$

By making use of Proposition 6.5 we can show that

$$\lim_{n \rightarrow \infty} \sup_{1 \leq m \leq M_n} \sup_{1 \leq j \leq N_n} \left| \frac{1}{nh} F_{m,j} - \bar{e}_j(\chi_\eta^2) g_0^{-2}(\tilde{x}_m) \right| = 0 \quad (6.16)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{nh} \sup_{1 \leq m \leq M_n} \sup_{1 \leq j \leq N_n} B_{m,j} = 0. \quad (6.17)$$

The detailed proof of these equalities is given in Galtchouk, Pergamenshchikov, 2008 (Appendix A.8).

This means that, for any  $\nu > 0$  and for sufficiently large  $n$ ,

$$\sup_{1 \leq m \leq M_n} \sup_{1 \leq j \leq N_n} \frac{F_{m,j} + B_{m,j} + t_{m,j}^{-2}}{nh \bar{e}_j(\chi_\eta^2) g_0^{-2}(\tilde{x}_m) + t_{m,j}^{-2}} \leq 1 + \nu,$$

where  $\tilde{x}_m$  is defined in (6.9). Therefore, if we denote in (6.15)

$$\kappa_{m,j}^2 = nh g_0^{-2}(\tilde{x}_m) t_{m,j}^2 \quad \text{and} \quad \psi_j(\eta, y) = \frac{\bar{e}_j^2(\chi_\eta) y}{\bar{e}_j^2(\chi_\eta) y + 1}$$

we obtain, for sufficiently large  $n$ ,

$$n^{\frac{2k}{2k+1}} \mathcal{E}_n(\widehat{S}_n) \geq \frac{n^{-\frac{1}{2k+1}}}{1 + \nu} \sum_{m=1}^{M_n} g_0^2(\tilde{x}_m) \sum_{j=1}^{N_n} \psi_j(\eta, \kappa_{m,j}^2).$$

Moreover, the property (6.7) implies

$$\lim_{\eta \rightarrow 0} \sup_{N \geq 1} \sup_{(y_1, \dots, y_N) \in \mathbb{R}_+^N} \left| \frac{\sum_{j=1}^N \psi_j(\eta, y_j)}{\Psi_N(y_1, \dots, y_N)} - 1 \right| = 0, \quad (6.18)$$

where

$$\Psi_N(y_1, \dots, y_N) = \sum_{j=1}^N \frac{y_j}{y_j + 1}.$$

Therefore we can write that, for sufficiently large  $n$ ,

$$n^{\frac{2k}{2k+1}} \mathcal{E}_n(\widehat{S}_n) \geq \frac{1 - \nu}{1 + \nu} n^{-\frac{1}{2k+1}} \sum_{m=1}^{M_n} g_0^2(\tilde{x}_m) \Psi_{N_n}(\kappa_{m,1}^2, \dots, \kappa_{m,N_n}^2). \quad (6.19)$$

Obviously, to obtain a "good" lower bound for the risk  $\mathcal{E}_n(\widehat{S}_n)$  one needs to maximize the right-hand side of the inequality (6.19). Hence we choose the coefficients  $(\kappa_{m,j}^2)$  by maximizing the function  $\Psi_N$ , i.e.

$$\max_{y_1, \dots, y_N} \Psi_N(y_1, \dots, y_N) \quad \text{subject to} \quad \sum_{j=1}^N y_j j^{2k} \leq R.$$

The parameter  $R > 0$  will be chosen later to satisfy the condition  $\mathbf{A}_3$ ). By the Lagrange multipliers method it is easy to find that the solution of this problem is given by

$$y_j^*(R) = a^*(R) j^{-k} - 1 \quad (6.20)$$

with

$$a^*(R) = \frac{1}{\sum_{i=1}^N i^k} \left( R + \sum_{i=1}^N i^{2k} \right) \quad \text{and} \quad 1 \leq j \leq N.$$

To obtain a positive solution in (6.20) we need to impose the following condition

$$R > N^k \sum_{i=1}^N i^k - \sum_{i=1}^N i^{2k}. \quad (6.21)$$

Moreover, from the condition  $\mathbf{A}_3$ ) we obtain that

$$R \leq \frac{2^{2k+1}(1-\varepsilon)r n h_n^{2k+1}}{\pi^{2k} \widehat{g}_0} := R_n^*, \quad (6.22)$$

where

$$\widehat{g}_0 = 2h_n \sum_{m=1}^{M_n} g_0^2(\tilde{x}_m).$$

Note that, by the condition  $\mathbf{H}_4$ ), the function  $g_0(\cdot) = g(\cdot, S_0)$  is continuous on the interval  $[0, 1]$ , therefore,

$$\lim_{n \rightarrow \infty} \widehat{g}_0 = \int_0^1 g^2(x, S_0) dx = \varsigma(S_0) \quad \text{with} \quad S_0 \equiv 0. \quad (6.23)$$

Now we have to choose the sequence  $(h_n)$ . Note that if we put in (6.10)

$$t_{m,j} = g_0(\tilde{x}_m) \sqrt{y_j^*(R)} / \sqrt{nh_n} \quad \text{i.e.} \quad \kappa_{m,j} = y_j^*(R), \quad (6.24)$$

we can rewrite the inequality (6.19) as

$$n^{\frac{2k}{2k+1}} \mathcal{E}_n(\widehat{S}_n) \geq \frac{(1-\nu)}{(1+\nu)} \frac{\widehat{g}_0 \Psi_{N_n}^*(R)}{2h_n} n^{-\frac{1}{2k+1}}, \quad (6.25)$$

where

$$\Psi_{N_n}^*(R) = N - \frac{\left(\sum_{j=1}^N j^k\right)^2}{R + \sum_{j=1}^N j^{2k}}.$$

It is clear that

$$k^2/(k+1)^2 \leq \liminf_{N \rightarrow \infty} \inf_{R > 0} \Psi_{N_n}^*(R)/N \leq \limsup_{N \rightarrow \infty} \sup_{R > 0} \Psi_{N_n}^*(R)/N \leq 1.$$

Therefore, to obtain a positive finite asymptotic lower bound in (6.25) we have to take the parameter  $h_n$  as

$$h_n = h_* n^{-1/(2k+1)} N_n \quad (6.26)$$

with some positive coefficient  $h_*$ . Moreover, the conditions (6.21)-(6.22) imply that, for sufficiently large  $n$ ,

$$(1-\varepsilon)r \frac{2^{2k+1}}{\pi^{2k}} \frac{1}{\widehat{g}_0} h_*^{2k+1} > \frac{1}{N_n^{k+1}} \sum_{j=1}^{N_n} j^k - \frac{1}{N_n^{2k+1}} \sum_{j=1}^{N_n} j^{2k}.$$

Now taking into account that, for sufficiently large  $n$ ,

$$\widehat{g}_0 \frac{1}{N_n} \sum_{j=1}^{N_n} \left( (j/N_n)^k - (j/N_n)^{2k} \right) < \frac{(1+\varepsilon)\varsigma(S_0)k}{(k+1)(2k+1)},$$

we obtain the following condition on  $h_*$ :

$$h_* \geq (v_\varepsilon^*)^{1/(2k+1)}, \quad (6.27)$$

where

$$v_\varepsilon^* = \frac{(1+\varepsilon)k}{c_\varepsilon^*(k+1)(2k+1)} \quad \text{and} \quad c_\varepsilon^* = \frac{2^{2k+1}(1-\varepsilon)r}{\pi^{2k}\varsigma(S_0)}.$$

To maximize the function  $\Psi_{N_n}^*(R)$  on the right-hand side of the inequality (6.25), we take  $R = R_n^*$  defined in (6.22). Therefore, we obtain that

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{S}_n} n^{2k/(2k+1)} \mathcal{E}_n(\widehat{S}_n) \geq \varsigma(S_0) F(h_*)/2, \quad (6.28)$$

where

$$F(x) = \frac{1}{x} - \frac{2k+1}{(k+1)^2(c_\varepsilon^*(2k+1)x^{2k+2} + x)}.$$

Furthermore, taking into account that

$$F'(x) = -\frac{(c_\varepsilon^*(2k+1)(k+1)x^{2k+1} - k)^2}{(k+1)^2(c_\varepsilon^*(2k+1)x^{2k+2} + x)^2} \leq 0,$$

we get

$$\max_{h_* \geq (v_\varepsilon^*)^{1/(2k+1)}} F(h_*) = F((v_\varepsilon^*)^{1/(2k+1)}) = \frac{(1+\varepsilon')k}{k+1} (v_\varepsilon^*)^{-1/(2k+1)},$$

where

$$\varepsilon' = \frac{\varepsilon}{2k + \varepsilon k + 1}. \quad (6.29)$$

This means that to obtain in (6.28) the maximal lower bound, one has to take in (6.26)

$$h_* = (v_\varepsilon^*)^{1/(2k+1)}. \quad (6.30)$$

It is important to note that if one defines the prior distribution  $\mu_\vartheta$  in the bayesian risk (6.13) by formulas (6.10), (6.24), (6.26) and (6.30), then the bayesian risk should depend on a parameter  $0 < \varepsilon < 1$ , i.e.  $\mathcal{E}_n = \mathcal{E}_{\varepsilon,n}$ . Therefore, the inequality (6.28) implies that, for any  $0 < \varepsilon < 1$ ,

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n} n^{2k/(2k+1)} \mathcal{E}_{\varepsilon,n}(\hat{S}_n) \geq \frac{(1+\varepsilon')(1-\varepsilon)^{1/(2k+1)}}{(1+\varepsilon)^{1/(2k+1)}} \gamma_k(S_0), \quad (6.31)$$

where the function  $\gamma_k(S_0)$  is defined in (4.7) for  $S_0 \equiv 0$ .

Now to end the definition of the sequence of the random functions  $(S_{\vartheta,n})$  defined by (6.9) and (6.10), one has to define the sequence  $(N_n)$ . Let us remember that we make use of the sequence  $(S_{\vartheta,n})$  with the coefficients  $(t_{m,j})$  constructed in (6.24) for  $R = R_n^*$  given in (6.22) and for the sequence  $h_n$  given by (6.26) and (6.30), for some fixed  $0 < \varepsilon < 1$ .

We will choose the sequence  $(N_n)$  to satisfy the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$ . One can take, for example,  $N_n = \lceil \ln^4 n \rceil + 1$ . Then the condition  $\mathbf{A}_1)$  is trivial. Moreover, taking into account that in this case

$$R_n^* = \frac{2^{2k+1}(1-\varepsilon)r}{\pi^{2k}\hat{g}_0} v_\varepsilon^* N_n^{2k+1} = \frac{\varsigma(S_0)}{\hat{g}_0} \frac{k}{(k+1)(2k+1)} N_n^{2k+1},$$

we find, thanks to the convergence (6.23),

$$\lim_{n \rightarrow \infty} \frac{R_n^* + \sum_{j=1}^{N_n} j^{2k}}{N_n^k \sum_{j=1}^{N_n} j^k} = 1.$$

Therefore, the solution (6.20), for sufficiently large  $n$ , satisfies the following inequality:

$$\max_{1 \leq j \leq N_n} y_j^*(R_n^*) j^k \leq 2N_n^k.$$

Now it is easy to see that the condition  $\mathbf{A}_2$ ) holds with  $d_n = \sqrt{N_n}$  and the condition  $\mathbf{A}_4$ ) holds for arbitrary  $0 < \epsilon_0 < 1$ . As to the condition  $\mathbf{A}_3$ ), note that, in view of the definition of  $t_{m,j}$  in (6.24), we get

$$\begin{aligned} \frac{1}{h_n^{2k-1}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 j^{2k} &= \frac{1}{2nh_n^{2k+1}} \hat{g}_0 \sum_{j=1}^{N_n} y_j^*(R_n^*) j^{2k} \\ &= \frac{R_n^* \hat{g}_0}{N_n^{2k+1} 2v_\epsilon^*} = (1 - \epsilon)r \left( \frac{2}{\pi} \right)^{2k}. \end{aligned}$$

Hence the condition  $\mathbf{A}_3$ ).

## 7 Estimation of non periodic function

Now we consider the estimation problem of a non periodic regression function  $S$  in the model (1.1). In this case we will estimate the function  $S$  on any interior interval  $[a, b]$  of  $[0, 1]$ , i.e. for  $0 < a < b < 1$ .

It should be pointed out that at the boundary points  $x = 0$  and  $x = 1$ , one must to make use of kernel estimators (see Brua, 2007).

Let now  $\chi$  be a infinitely differentiable  $[0, 1] \rightarrow \mathbb{R}_+$  function such that  $\chi(x) = 1$  for  $a \leq x \leq b$  and  $\chi^{(k)}(0) = \chi^{(k)}(1) = 0$  for all  $k \geq 0$ , for example,

$$\chi(x) = \frac{1}{\eta} \int_{-\infty}^{\infty} V \left( \frac{x-z}{\eta} \right) \mathbf{1}_{[a', b']}(z) dz,$$

where  $V$  is some kernel function introduced in (6.6),

$$a' = \frac{a}{2}, \quad b' = \frac{b}{2} + \frac{1}{2} \quad \text{and} \quad \eta = \frac{1}{4} \min(a, 1 - b).$$

Multiplying the equation (1.1) by the function  $\chi(\cdot)$  and simulating some i.i.d.  $\mathcal{N}(0, 1)$  sequence  $(\zeta_j)_{1 \leq j \leq n}$  one comes to the estimation problem of the periodic regression function  $\tilde{S}(x) = S(x)\chi(x)$  in the model

$$\tilde{y}_j = \tilde{S}(x_j) + \tilde{\sigma}_j(S) \tilde{\xi}_j,$$

where  $\tilde{\sigma}_j(S) = \sqrt{\sigma_j^2(S) + \epsilon^2}$ ,

$$\tilde{\xi}_j = \frac{\sigma_j(S)}{\tilde{\sigma}_j(S)} \xi_j + \frac{\epsilon}{\tilde{\sigma}_j(S)} \zeta_j.$$

and  $\epsilon > 0$  is some sufficiently small parameter.

It is easy to see that if the sequence  $(\sigma_j(S))_{1 \leq j \leq n}$  satisfies the conditions  $\mathbf{H}_1) - \mathbf{H}_4)$ , then the sequence  $(\tilde{\sigma}_j(S))_{1 \leq j \leq n}$  satisfies these conditions as well with

$$\tilde{\sigma}_j(S) = \tilde{g}(x_j, S) = \sqrt{g^2(x_j, S)\chi^2(x_j) + \epsilon^2}.$$

## 8 Conclusion

In conclusion, it should be noted that this paper completes the investigation of the estimation problem for the nonparametric regression function in the heteroscedastic regression model (1.1) in the case of quadratic risk. It is shown that the adaptive procedure (2.13) satisfies the non asymptotic oracle inequality and it is asymptotically efficient for estimating a periodic regression function as well. From practical point of view, the procedure (2.13) gives an acceptable accuracy even for small samples as it is shown via simulations by Galtchouk, Pergamenschikov, 2009.

## 9 Proofs

### 9.1 Properties of the trigonometric basis

**Lemma 9.1.** *For any function  $S \in W_r^k$ ,*

$$\sup_{n \geq 1} \sup_{1 \leq m \leq n-1} m^{2k} \left( \sum_{j=m+1}^n \theta_{j,n}^2 \right) \leq \frac{4r}{\pi^{2(k-1)}}. \quad (9.1)$$

**Lemma 9.2.** For any  $m \geq 0$ ,

$$\sup_{N \geq 2} \sup_{x \in [0,1]} N^{-m} \left| \sum_{l=2}^N l^m (\phi_l^2(x) - 1) \right| \leq 2^m. \quad (9.2)$$

Proofs of Lemma 9.1 and Lemma 9.2 are given in Galtchouk, Pergamenshchikov, 2007.

**Lemma 9.3.** Let  $\theta_{j,n}$  and  $\theta_j$  be the Fourier coefficients defined in (2.4) and (3.4), respectively. Then, for  $1 \leq j \leq n$  and  $n \geq 2$ ,

$$\sup_{S \in W_r^1} |\theta_{j,n} - \theta_j| \leq 2\pi \sqrt{r} j/n. \quad (9.3)$$

**Proof.** Indeed, we have

$$\begin{aligned} |\theta_{j,n} - \theta_j| &= \left| \sum_{l=1}^n \int_{x_{l-1}}^{x_l} (S(x_l)\phi_j(x_l) - S(x)\phi_j(x)) dx \right| \\ &\leq n^{-1} \sum_{l=1}^n \int_{x_{l-1}}^{x_l} (|\dot{S}(z)\phi_j(z)| + |S(z)\dot{\phi}_j(z)|) dz \\ &= n^{-1} \int_0^1 (|\dot{S}(z)| |\phi_j(z)| + |S(z)| |\dot{\phi}_j(z)|) dz. \end{aligned}$$

By making use of the Bounyakovskii-Cauchy-Schwarz inequality we get

$$\begin{aligned} |\theta_{j,n} - \theta_j| &\leq n^{-1} \left( \|\dot{S}\| \|\phi\| + \|\dot{\phi}\| \|S\| \right) \\ &\leq n^{-1} \left( \|\dot{S}\| + \pi j \|S\| \right). \end{aligned}$$

The definition of the class  $W_r^1$  implies (9.3). Hence Lemma 9.1.  $\square$

## 9.2 Proof of Theorem 5.1

To prove the theorem we will adapt to the heteroscedastic case the corresponding proof from Nussbaum, 1985.

First, from (2.5) we obtain that, for any  $p \in \mathcal{P}_n$ ,

$$\mathbf{E}_{S,p} \|\tilde{S} - S\|_n^2 = \sum_{j=1}^n (1 - \tilde{\lambda}_j)^2 \theta_{j,n}^2 + \frac{1}{n} \sum_{j=1}^n \tilde{\lambda}_j^2 \zeta_{j,n}(S), \quad (9.4)$$

where

$$\varsigma_{j,n}(S) = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \phi_j^2(x_l).$$

Setting now  $\tilde{\omega} = \omega(\tilde{\alpha})$  with the function  $\omega$  defined in (2.9), the index  $\tilde{\alpha}$  defined in (5.2),  $\tilde{j}_0 = [\tilde{\omega}\varepsilon_n]$ ,  $\tilde{j}_1 = [\tilde{\omega}/\varepsilon_n]$  and

$$\varsigma_n(S) = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S),$$

we rewrite (9.4) as follows

$$\begin{aligned} \mathbf{E}_{S,p} \|\tilde{S} - S\|_n^2 &= \sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (1 - \tilde{\lambda}_j)^2 \theta_{j,n}^2 \\ &\quad + \frac{\varsigma_n(S)}{n} \sum_{j=1}^n \tilde{\lambda}_j^2 + \tilde{\Delta}_{1,n} + \tilde{\Delta}_{2,n} \end{aligned} \quad (9.5)$$

with

$$\tilde{\Delta}_{1,n} = \sum_{j=\tilde{j}_1}^n \theta_{j,n}^2 \quad \text{and} \quad \tilde{\Delta}_{2,n} = \frac{1}{n} \sum_{j=1}^n \tilde{\lambda}_j^2 (\varsigma_{j,n}(S) - \varsigma_n(S)).$$

Note that we have decomposed the first term on the right-hand of (9.4) into the sum

$$\sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (1 - \tilde{\lambda}_j)^2 \theta_{j,n}^2 + \tilde{\Delta}_{1,n}.$$

This decomposition allows us to show that  $\tilde{\Delta}_{1,n}$  is negligible and further to approximate the first term by a similar term in which the coefficients  $\theta_{j,n}$  will be replaced by the Fourier coefficients  $\theta_j$  of the function  $S$ .

Taking into account the definition of  $\omega$  in (2.9), we can bound  $\tilde{\omega}$  as

$$\tilde{\omega} \geq (A_k)^{\frac{1}{2k+1}} (n\varepsilon_n)^{\frac{1}{2k+1}}.$$

Therefore, by Lemma 9.1 we obtain

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} n^{\frac{2k}{2k+1}} \tilde{\Delta}_{1,n} = 0.$$

Let us consider now the next term  $\tilde{\Delta}_{2,n}$ . We have

$$|\tilde{\Delta}_{2,n}| = \frac{1}{n^2} \left| \sum_{i=1}^n \sigma_i^2(S) \sum_{j=1}^n \tilde{\lambda}_j^2 (\phi_j^2(x_i) - 1) \right| \leq \frac{\sigma_*}{n} \sup_{0 \leq x \leq 1} \left| \sum_{j=1}^n \tilde{\lambda}_j^2 (\phi_j^2(x) - 1) \right|.$$

Now by Lemma 9.2 and the definition (2.8), we obtain directly the same property for  $\tilde{\Delta}_{2,n}$ , i.e.

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} n^{\frac{2k}{2k+1}} |\tilde{\Delta}_{2,n}| = 0.$$

Setting

$$\hat{\gamma}_{k,n}(S) = n^{\frac{2k}{2k+1}} \sum_{j=\tilde{j}_0}^{\tilde{j}_1-1} (1 - \tilde{\lambda}_j)^2 \theta_j^2 + \varsigma_n(S) n^{-\frac{1}{2k+1}} \sum_{j=1}^n \tilde{\lambda}_j^2$$

and applying the well-known inequality

$$(a + b)^2 \leq (1 + \delta)a^2 + (1 + 1/\delta)b^2$$

to the first term on the right-hand side of the inequality (9.5), we obtain that, for any  $\delta > 0$  and for any  $p \in \mathcal{P}_n$ ,

$$\begin{aligned} \mathbf{E}_{S,p} \|\tilde{S} - S\|_n^2 &\leq (1 + \delta) \hat{\gamma}_{k,n}(S) n^{-2k/(2k+1)} \\ &\quad + \tilde{\Delta}_{1,n} + \tilde{\Delta}_{2,n} + (1 + 1/\delta) \tilde{\Delta}_{3,n}, \end{aligned} \quad (9.6)$$

where

$$\tilde{\Delta}_{3,n} = \sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (\theta_{j,n} - \theta_j)^2.$$

Taking into account that  $k \geq 1$  and that

$$\tilde{j}_1 \leq \bar{\omega} \varepsilon_n^{-1} + (A_k)^{\frac{1}{2k+1}} n^{\frac{1}{2k+1}} (\varepsilon_n)^{-(2k+2)/(2k+1)},$$

we can show through Lemma 9.3 that

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} n^{\frac{2k}{2k+1}} \tilde{\Delta}_{3,n} = 0.$$

Therefore, the inequality (9.6) yields

$$\limsup_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \sup_{S \in W_r^k} \mathcal{R}_n(\tilde{S}, S) / \gamma_k(S) \leq \limsup_{n \rightarrow \infty} \sup_{S \in W_r^k} \hat{\gamma}_{k,n}(S) / \gamma_k(S)$$

and to prove (5.4) it suffices to show that

$$\limsup_{n \rightarrow \infty} \sup_{S \in W_r^k} \hat{\gamma}_{k,n}(S) / \gamma_k(S) \leq 1. \quad (9.7)$$

First, it should be noted that the definition (5.1) and the inequalities (3.7)-(3.8) imply directly

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} |\tilde{t}_n / \bar{r}(S) - 1| = 0.$$

Moreover, by the definition of  $(\tilde{\lambda}_j)_{1 \leq j \leq n}$  in (5.3), for sufficiently large  $n$  for which  $\tilde{t}_n \geq \bar{r}(S)$ , we find

$$\sup_{j \geq 1} n^{\frac{2k}{2k+1}} \frac{(1 - \tilde{\lambda}_j)^2}{(\pi j)^{2k}} = \pi^{-2k} (A_k \tilde{t}_n)^{-2k/(2k+1)} \leq \pi^{-2k} (A_k \bar{r}(S))^{-2k/(2k+1)}.$$

Therefore, by the definition of the coefficients  $(a_j)_{j \geq 1}$  in (3.5), one has

$$\limsup_{n \rightarrow \infty} n^{\frac{2k}{2k+1}} \sup_{S \in W_r^k} \sup_{j \geq \tilde{j}_0} \pi^{2k} (A_k \bar{r}(S))^{2k/(2k+1)} (1 - \tilde{\lambda}_j)^2 / a_j \leq 1.$$

Furthermore, in view of the definition (2.8) we calculate directly

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^k} \left| n^{-\frac{1}{2k+1}} \sum_{j=1}^n \tilde{\lambda}_j^2 - (A_k \bar{r}(S))^{\frac{1}{2k+1}} \int_0^1 (1 - z^k)^2 dz \right| = 0.$$

Now, the definition of  $W_r^k$  in (3.3) and the condition (3.6) imply the inequality (9.7). Hence Theorem 5.1.  $\square$

### 9.3 Proof of Theorem 6.1

For any  $z = (z_1, \dots, z_l)' \in \mathbb{R}^n$ , we set

$$\varrho_i(v, z) = \frac{1}{f(v, z)\Phi(z)} \frac{\partial}{\partial z_i} (f(v, z)\Phi(z)).$$

Note that, due to the condition (3.7), the density (6.3) is bounded, i.e.

$$f(v, z) \leq (2\pi g_*)^{-n/2}.$$

So through (6.2) we obtain that

$$\lim_{|z_i| \rightarrow \infty} \tau(z) f(v, z) \varphi_i(z_i) = 0.$$

Therefore, integrating by parts yields

$$\begin{aligned} \tilde{\mathbf{E}}(\hat{\tau}_n - \tau(\vartheta)) \varrho_i &= \int_{\mathbb{R}^{n+l}} (\hat{\tau}_n(v) - \tau(z)) \frac{\partial}{\partial z_i} (f(v, z) \Phi(z)) \, dz \, dv \\ &= \int_{\mathbb{R}^l} \left( \frac{\partial}{\partial z_i} \tau(z) \right) \Phi(z) \left( \int_{\mathbb{R}^n} f(v, z) \, dv \right) \, dz = \bar{\tau}_i. \end{aligned}$$

Now the Bounyakovskii-Cauchy-Schwarz inequality gives the following lower bound

$$\tilde{\mathbf{E}}(\hat{\tau}_n - \tau(\vartheta))^2 \geq \bar{\tau}_i^2 / \tilde{\mathbf{E}} \varrho_i^2.$$

To estimate the denominator in the last ratio, note that

$$\varrho_i(v, z) = \tilde{f}_i(v, z) + \frac{\dot{\varphi}_i(z_i)}{\varphi_i(z_i)} \quad \text{with} \quad \tilde{f}_i(v, z) = (\partial/\partial z_i) \ln f(v, z).$$

From (6.1) it follows that

$$\tilde{f}_i(v, z) = \sum_{j=1}^n (\xi_j^2 - 1) \frac{1}{2\sigma_j^2(z)} \frac{\partial}{\partial z_i} \sigma_j^2(z) + \sum_{j=1}^n \xi_j \frac{S'_i(x_j)}{\sigma_j(z)}.$$

Moreover, the conditions  $\mathbf{H}_2$ ) and (6.4) imply

$$(\partial/\partial z_i) \sigma_j^2(z) = (\partial/\partial z_i) g^2(x_j, S_z) = \tilde{\mathbf{L}}_i(x_j, z)$$

from which it follows

$$\tilde{\mathbf{E}} \left( \tilde{f}_i(Y, \vartheta) \right)^2 = F_i + B_i.$$

This implies inequality (6.5). Hence Theorem 6.1.  $\square$

## 9.4 Proof of Theorem 4.3

In this section we prove Theorem 4.3. Lemma 5.2 implies that to prove the lower bounds (4.10) and (4.11), it suffices to show

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{S}_n} n^{\frac{2k}{2k+1}} \mathcal{R}_0(\widehat{S}_n) \geq 1, \quad (9.8)$$

where

$$\mathcal{R}_0(\widehat{S}_n) = \sup_{S \in W_r^k} \mathbf{E}_{S,q} \|\widehat{S}_n - S\|^2 / \gamma_k(S).$$

For any estimator  $\widehat{S}_n$ , we denote by  $\widehat{S}_n^0$  its projection onto  $W_r^k$ , i.e.  $\widehat{S}_n^0 = \Pr_{W_r^k}(\widehat{S}_n)$ . Since  $W_r^k$  is a convex set, we get

$$\|\widehat{S}_n - S\|^2 \geq \|\widehat{S}_n^0 - S\|^2.$$

Now we introduce the following set

$$\Xi_n = \left\{ \max_{1 \leq m \leq M_n} \max_{1 \leq j \leq N} \zeta_{m,j}^2 \leq d_n \right\}, \quad (9.9)$$

where  $(\zeta_{m,j})$  are i.i.d.  $\mathcal{N}(0, 1)$  random variables from (6.10) and the sequence  $(d_n)_{n \geq 1}$  is given in the condition  $\mathbf{A}_2$ ). It is clear that the condition  $\mathbf{A}_1$ ) implies

$$\limsup_{n \rightarrow \infty} n^\nu \mathbf{P}(\Xi_n^c) = 0. \quad (9.10)$$

Therefore, we can write that

$$\mathcal{R}_0(\widehat{S}_n) \geq \int_{\{z: S_{z,n} \in W_r^k\} \cap \Xi_n} \frac{\mathbf{E}_{S_{z,n},q} \|\widehat{S}_n^0 - S_{z,n}\|^2}{\gamma_k(S_{z,n})} \mu_\vartheta(dz).$$

Here the kernel function family  $(S_{z,n})$  is given in (6.9) in which  $N_n = \lceil \ln^4 n \rceil + 1$  and the parameter  $h$  is defined in (6.26) and (6.30); the measure  $\mu_\vartheta$  is defined in (6.13). Moreover, note that on the set  $\Xi$  the random function  $S_{\vartheta,n}$  is uniformly bounded, i.e.

$$\|S_{\vartheta,n}\|_\infty = \sup_{0 \leq x \leq 1} |S_{\vartheta,n}(x)| \leq \sqrt{d_n} t_n^*, \quad (9.11)$$

where the coefficient  $t_n^*$  is defined in (6.11).

Thus, we estimate the risk  $\mathcal{R}_0(\widehat{S}_n)$  from below as

$$\mathcal{R}_0(\widehat{S}_n) \geq \frac{1}{\gamma_n^*} \int_{\{z: S_{z,n} \in W_r^k\} \cap \Xi_n} \mathbf{E}_{S_{z,n}, q} \|\widehat{S}_n^0 - S_{z,n}\|^2 \mu_\vartheta(dz)$$

with

$$\gamma_n^* = \sup_{\|S\|_\infty \leq \sqrt{d_n} t_n^*} \gamma_k(S). \quad (9.12)$$

By making use of the Bayes risk (6.13) with the prior distribution given by formulas (6.10), (6.24), (6.26) and (6.30), for any fixed parameter  $0 < \varepsilon < 1$ , we rewrite the lower bound for  $\mathcal{R}_0(\widehat{S}_n)$  as

$$\mathcal{R}_0(\widehat{S}_n) \geq \mathcal{E}_{\varepsilon, n}(\widehat{S}_n^0) / \gamma_n^* - 2 \Omega_n / \gamma_n^* \quad (9.13)$$

with

$$\Omega_n = \mathbf{E}(\mathbf{1}_{\{S_{\vartheta, n} \notin W_r^k\}} + \mathbf{1}_{\Xi_n^c})(r + \|S_{\vartheta, n}\|^2).$$

In Section 6.3 we proved that the parameters in chosen prior distribution satisfy the conditions  $\mathbf{A}_1)$ – $\mathbf{A}_4)$ . Therefore, Propositions 6.3–6.4 and the limit (9.10) imply that, for any  $\nu > 0$ ,

$$\lim_{n \rightarrow \infty} n^\nu \Omega_n = 0.$$

Moreover, by the condition  $\mathbf{H}_4)$  the sequence  $\gamma_n^*$  goes to  $\gamma_k(S_0)$  as  $n \rightarrow \infty$ . Therefore, from this, (6.31) and (9.13) we get, for any  $0 < \varepsilon < 1$ ,

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{S}_n} n^{\frac{2k}{2k+1}} \mathcal{R}_0(\widehat{S}_n) \geq \frac{(1 + \varepsilon')(1 - \varepsilon)^{\frac{1}{2k+1}}}{(1 + \varepsilon)^{\frac{1}{2k+1}}}.$$

where  $\varepsilon'$  is defined in (6.29). Limiting here  $\varepsilon \rightarrow 0$  implies the inequality (9.8). Hence Theorem 4.3.  $\square$

### Acknowledgements

We are grateful for the comments and stimulating questions by the associate Editor and two anonymous Referees which have led to considerable improvements.

## References

- [1] Barron, A., Birgé, L., & Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*. **113**, 301-413.
- [2] Belomestny, D., & Reiss M. (2006). Spectral calibration of exponential Lévy models. *Finance and Stochastics*, **10**(4), 449-474.
- [3] Brua, J.-Y. (2007). Asymptotic efficient estimators for non-parametric heteroscedastic model. *Statistical Metodologie*, accepted to publication, available at <http://hal.archives-ouvertes.fr/hal/-00178536/fr/>
- [4] Efroimovich, S. (2007). Sequential design and estimation in heteroscedastic nonparametric regression. *Sequential Analysis*, **26**, 3-25.
- [5] Efroimovich, S., & Pinsker, M. (1996). Sharp - optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica*, **6**, 925-942.
- [6] Fourdrinier, D., & Pergamenshchikov, S. (2007). Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*. **59**(3), 435-464.
- [7] Galtchouk, L., & Pergamenshchikov, S. (2006). Asymptotically efficient estimates for nonparametric regression models. *Statistics & Probability Letters*. **76**, 852-860.
- [8] Galtchouk, L., & Pergamenshchikov, S. (2005). Efficient adaptive non-parametric estimation in heteroscedastic regression models. Preprint of the Strasbourg Louis Pasteur University, IRMA, 2005/020 available at <http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchtchikov>
- [9] Galtchouk, L., & Pergamenshchikov, S. (2007). Adaptive nonparametric estimation in heteroscedastic regression models. Part 1. Sharp non-asymptotic oracle inequalities. Preprint of the Strasbourg Louis Pasteur University, IRMA, 2007/09, available at <http://hal.archives-ouvertes.fr/hal/-00179856/fr/>
- [10] Galtchouk, L., & Pergamenshchikov, S. (2009). Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21** (1), 1-16

- [11] Galtchouk, L., & Pergamenshchikov, S. (2008). "Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection." available at  
<http://hal.archives-ouvertes.fr/hal-00326910/fr/>
- [12] Gill, R.D., & Levit, B.Y. (1995). Application of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1**, 59-79.
- [13] Goldfeld, S.M., & Quandt, R.E. (1972). *Nonlinear Methods in Econometrics*. North-Holland, London.
- [14] Golubev, G.K., & Nussbaum, M. (1993). Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.* **37**, 521-529.
- [15] Ibragimov, I.A., & Hasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.
- [16] Kolmogorov, A. N., & Fomin, S. V. (1989). *Elements of the theory of functions and functional analysis*. Sixth edition. Nauka, Moscow.
- [17] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in  $L_2$ . *Ann. Statist.* **13**, 984-997.
- [18] Pinsker, M.S. (1981) Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission information*, **17**, 120-133.
- [19] Rohde, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statistics & Decisions* **22**, 235-243.
- [20] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

L.Galtchouk  
 Department of Mathematics  
 Strasbourg University  
 7, rue Rene Descartes  
 67084, Strasbourg, France  
 e-mail: galtchou@math.u-strasbg.fr

S. Pergamenshchikov  
 Laboratoire de Mathématiques Raphael Salem,  
 Avenue de l'Université, BP. 12,  
 Université de Rouen,  
 F76801, Saint Etienne du Rouvray, Cedex France.  
 Serge.Pergamenchtchikov@univ-rouen.fr