Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency.

Victor Konev, Serguei Pergamenchtchikov

Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *

Konev, V. † S. Pergamenshchikov‡

September 16, 2009

Abstract

In this paper we prove the asymptotic efficiency of the model selection procedure proposed by the authors in [10]. To this end we introduce the robust risk as the least upper bound of the quadratical risk over a broad class of observation distributions. Asymptotic upper and lower bounds for the robust risk have been derived. The asymptotic efficiency of the procedure is proved. The Pinsker constant is found.

Keywords: Non-parametric regression; Model selection; Sharp oracle inequality; Robust risk; Asymptotic efficiency; Pinsker constant; Semimartingale noise.

AMS 2000 Subject Classifications: Primary: 62G08; Secondary: 62G05

*The paper is supported by the RFFI-Grant 09-01-00172-a.

†Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia, e-mail: vvkonev@mail.tsu.ru

‡Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France, Department of Mathematics and Mechanics,Tomsk State University, Lenin str. 36, 634041 Tomsk, Russia, e-mail: Serge.Pergamenchtchikov@univ-rouen.fr

1 Introduction

In this paper we will investigate the asymptotic efficiency of the model selection procedure proposed in [10] for estimating a 1-periodic function $S : \mathbb{R} \to \mathbb{R}$, $S \in \mathcal{L}_2[0,1]$, in a continuous time regression model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le n\,, \tag{1.1}$$

with a semimartingale noise $\xi = (\xi_t)_{0\le t\le n}$. The quality of an estimate $\widetilde{S}$ (any real-valued function measurable with respect to $\sigma\{y_t\,,\, 0 \le t \le n\}$) for S is given by the mean integrated squared error, i.e.

$$\mathcal{R}_Q(\widetilde{S}, S) = \mathbf{E}_{Q,S}\,||\widetilde{S} - S||^2\,, \tag{1.2}$$

where $\mathbf{E}_{Q,S}$ is the expectation with respect to the noise distribution Q given a function S;

$$||S||^2 = \int_0^1 S^2(x)\mathrm{d}x\,.$$

The semimartingale noise $(\xi_t)_{0\le t\le n}$ is assumed to take values in the Skorohod space $\mathcal{D}[0,n]$ and has the distribution Q on $\mathcal{D}[0,n]$ such that for any function f from $\mathcal{L}_2[0,n]$ the stochastic integral

$$I_n(f) = \int_0^n f_s \mathrm{d}\xi_s \tag{1.3}$$

is well defined with

$$\mathbf{E}_Q I_n(f) = 0 \quad \text{and} \quad \mathbf{E}_Q I_n^2(f) \le \sigma^* \int_0^n f_s^2\,\mathrm{d}s \tag{1.4}$$

where σ^* is some positive constant which may, in general, depend on n, i.e. $\sigma^* = \sigma_n^*$, such that

$$0 < \liminf_{n\to\infty} \sigma_n^* \le \limsup_{n\to\infty} \sigma_n^* < \infty\,. \tag{1.5}$$

Now we define a robust risk function which is required to measure the quality of an estimate $\widetilde{S}$ provided that a true distribution of the noise $(\xi_t)_{0\le t\le n}$ is known to belong to some family of distributions $\mathcal{Q}_n^*$ which will be specified below. Just as in [6] we define the robust risk as

$$\mathcal{R}_n^*(\widetilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n^*} \mathcal{R}_Q(\widetilde{S}_n, S)\,. \tag{1.6}$$

The goal of this paper is to prove that the model selection procedure for estimating S in the model (1.1) constructed in [10] is asymptotically efficient with respect to this risk. When studying the asymptotic efficiency of this procedure, described in detail in Section 2, we suppose that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0,1]\,, \sum_{j=0}^{k} ||f^{(j)}||^2 \le r\}\,, \tag{1.7}$$

where $r > 0\,,\ k \ge 1$ are some parameters, $\mathcal{C}_{per}^k[0,1]$ is a set of k times continuously differentiable functions $f : [0,1] \to \mathbb{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \le i \le k$. The functional class W_r^k can be written as the ellipsoid in l_2, i.e.

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0,1] \,:\, \sum_{j=1}^{\infty} a_j\, \theta_j^2 \,\le r\} \tag{1.8}$$

where

$$a_j = \sum_{i=0}^{k} \left(2\pi[j/2]\right)^{2i}\ .$$

In [10] we established a sharp non-asymptotic oracle inequality for mean integrated squared error (1.2). The proof of the asymptotic efficiency of the model selection procedure below largely bases on the counterpart of this inequality for the robust risk (1.6) given in Theorem 2.1.

It will be observed that the notion "nonparametric robust risk" was initially introduced in [3] for estimating a regression curve at a fixed point. The greatest lower bound for such risks have been derived and a point estimate is found for which this bound is attained. The latter means that the point estimate turns out to be robust efficient. In [1] this approach was applied for pointwise estimation in a heteroscedastic regression model.

The optimal convergence rate of the robust quadratic risks has been obtained in [9] for the non-parametric estimation problem in a continuous time regression model with a coloured noise having unknown correlation properties under full and partial observations. The asymptotic efficiency with respect to the robust quadratic risks, has been studied in [6], [7] for the problem of non-parametric estimation in heteroscedastic regression models. In this paper we apply this approach for the model (1.1).

The rest of the paper is organized as follows. In Section 2 we construct the model selection procedure and formulate (Theorem 2.1) the oracle inequality for the robust risk. Section 3 gives the main results. In Section 4 we consider an example of the model (1.1) with the Levy type martingale noise. In Section 5 and 6 we obtain the upper and lower bounds for the robust risk. In Section 7 some technical results are established.

2 Oracle inequality for the robust risk

The model selection procedure is constructed on the basis of a weighted least squares estimate having the form

$$\widehat{S}_\gamma = \sum_{j=1}^{\infty} \gamma(j) \widehat{\theta}_{j,n} \phi_j \quad \text{with} \quad \widehat{\theta}_{j,n} = \frac{1}{n} \int_0^n \phi_j(t) \, \mathrm{d}y_t \,, \tag{2.1}$$

where $(\phi_j)_{j\geq 1}$ is the standard trigonometric basis in $\mathcal{L}_2[0,1]$ defined as

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

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. The sample functionals $\widehat{\theta}_{j,n}$ are estimates of the corresponding Fourier coefficients

$$\theta_j = (S, \phi_j) = \int_0^1 S(t) \, \phi_j(t) \, \mathrm{d}t \,. \tag{2.3}$$

Further we introduce the cost function as

$$J_n(\gamma) \, = \, \sum_{j=1}^{\infty} \gamma^2(j) \widehat{\theta}^2_{j,n} \, - 2 \sum_{j=1}^{\infty} \gamma(j) \, \widetilde{\theta}_{j,n} \, + \, \rho \, \widehat{P}_n(\gamma) \,.$$

Here

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{\widehat{\sigma}_n}{n} \quad \text{with} \quad \widehat{\sigma}_n = \sum_{j=l}^{n} \widehat{\theta}^2_{j,n} \,, \quad l = [\sqrt{n}] + 1 \,;$$

$\widehat{P}_n(\gamma)$ is the penalty term defined as

$$\widehat{P}_n(\gamma) = \frac{\widehat{\sigma}_n \, |\gamma|^2}{n} \,.$$

As to the parameter ρ, we assume that this parameter is a function of n, i.e. $\rho = \rho_n$ such that $0 < \rho < 1/3$ and

$$\lim_{n\to\infty} n^{\delta}\,\rho_n = 0 \quad \text{for all} \quad \delta > 0\,.$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\widehat{\gamma}} \tag{2.4}$$

where $\widehat{\gamma}$ is the minimizer of the cost function $J_n(\gamma)$ in some given class Γ of weight sequences $\gamma = (\gamma(j))_{j\geq 1} \in [0,1]^{\infty}$, i.e.

$$\widehat{\gamma} = \mathrm{argmin}_{\gamma\in\Gamma}\, J_n(\gamma)\,. \tag{2.5}$$

Now we specify the family of distributions $\mathcal{Q}_n^*$ in the robust risk (1.6). Let $\mathcal{P}_n$ denote the class of all distributions Q of the semimartingale (ξ_t) satisfying the condition (1.4). It is obvious that the distribution Q_0 of the process $\xi_t = \sqrt{\sigma^*}w_t$, where (w_t) is a standard Brownian motion, enters the class $\mathcal{P}_n$, i.e. $Q \in \mathcal{P}_n$. In addition, we need to impose some technical conditions on the distribution Q of the process $(\xi_t)_{0\leq t\leq n}$. Let denote

$$\sigma(Q) = \overline{\lim}_{n\to\infty} \max_{1\leq j\leq n} \mathbf{E}_Q\, \xi_{j,n}^2\,, \tag{2.6}$$

where

$$\xi_{j,n} = \frac{1}{\sqrt{n}}\, I_n(\phi_j)\,,$$

($I_n(\phi_j)$ is given in (1.3)) and introduce two $\mathcal{P}_n \to \mathbb{R}_+$ functionals

$$\mathbf{L}_{1,n}(Q) = \sup_{x\in\mathcal{H}\,,\,\#(x)\leq n} \left| \sum_{j=1}^{\infty} x_j \, \Bigl(\mathbf{E}_Q\, \xi_{j,n}^2 - \sigma(Q) \Bigr) \right|$$

and

$$\mathbf{L}_{2,n}(Q) = \sup_{|x|\leq 1\,,\,\#(x)\leq n} \mathbf{E}_Q \left(\sum_{j=1}^{\infty} x_j\, \widetilde{\xi}_{j,n} \right)^2$$

where $\mathcal{H} = [-1,1]^{\infty}$, $|x|^2 = \sum_{j=1}^{\infty} x_j^2$, $\#(x) = \sum_{j=1}^{\infty} \mathbf{1}_{\{|x_j|>0\}}$ and

$$\widetilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2\,.$$

Now we consider the family of all distributions Q from $\mathcal{P}_n$ with the growth restriction on $\mathbf{L}_{1,n}(Q) + \mathbf{L}_{2,n}(Q)$, i.e.

$$\mathcal{P}_n^* = \left\{ Q \in \mathcal{P}_n \, : \, \mathbf{L}_{1,n}(Q) + \mathbf{L}_{2,n}(Q) \leq \, l_n \right\} \, ,$$

where l_n is a slowly increasing positive function, i.e. $l_n \to +\infty$ as $n \to +\infty$ and for any $\delta > 0$

$$\lim_{n\to\infty} \frac{l_n}{n^\delta} = 0 \, .$$

It will be observed that any distribution Q from $\mathcal{P}_n^*$ satisfies conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) on the noise distribution from [10] with $c_{1,n}^* \leq l_n$ and $c_{2,n}^* \leq l_n$. We remind that these conditions are

$\mathbf{C}_1$)

$$c_{1,n}^* = \mathbf{L}_{1,n}(Q) < \infty \, ;$$

$\mathbf{C}_2$)

$$c_{2,n}^* = \mathbf{L}_{2,n}(Q) < \infty \, .$$

In the sequel we assume that the distribution of the noise (ξ_t) in (1.1) is known up to its belonging to some distribution family satisfying the following condition.

$\mathbf{C}^*$) *Let $\mathcal{Q}_n^*$ be a family of the distributions Q from $\mathcal{P}_n^*$ such that $Q_0 \in \mathcal{Q}_n^*$.*

An important example for such family is given in Section 4.

Now we specify the set Γ in the model selection procedure (2.4) and state the oracle inequality for the robust risk (1.6) which is a counterpart of that obtained in [10] for the mean integrated squared error (1.2). Consider the numerical grid

$$\mathcal{A}_n = \{1, \ldots, k^*\} \times \{t_1, \ldots, t_m\} \, , \tag{2.7}$$

where $t_i = i\varepsilon$ and $m = [1/\varepsilon^2]$; parameters $k^* \geq 1$ and $0 < \varepsilon \leq 1$ are functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that for any $\delta > 0$

$$\left\{ \begin{array}{l} \lim_{n\to\infty} k^*(n) = +\infty \, , \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0 \, , \\ \\ \lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^\delta \varepsilon(n) \, = +\infty \, . \end{array} \right. \tag{2.8}$$

For example, one can take

$$\varepsilon(n) = \frac{1}{\ln(n+1)} \quad \text{and} \quad k^*(n) = \sqrt{\ln(n+1)}$$

for $n \geq 1$.

Define the set Γ as

$$\Gamma \,=\, \{\gamma_\alpha\,,\, \alpha \in \mathcal{A}_n\}\,, \tag{2.9}$$

where γ_α is the weight sequence corresponding to an element $\alpha = (\beta, t) \in \mathcal{A}_n$, given by the formula

$$\gamma_\alpha(j) = \mathbf{1}_{\{1\leq j\leq j_0\}} + \left(1 - (j/\omega_\alpha)^\beta\right)\, \mathbf{1}_{\{j_0<j\leq\omega_\alpha\}} \tag{2.10}$$

where $j_0 = j_0(\alpha) = [\omega_\alpha/(1+\ln n)]$, $\omega_\alpha = (\tau_\beta\, t\, n)^{1/(2\beta+1)}$ and

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

Along the lines of the proof of Theorem 2.1 in [10] one can establish the following result.

Theorem 2.1. *Assume that the unknown function S is continuously differentiable and the distribution family $\mathcal{Q}_n^*$ in the robust risk* (1.6) *satisfies the condition* $\mathbf{C}^*$*). Then the estimator* (2.4)*, for any $n \geq 1$, satisfies the oracle inequality*

$$\mathcal{R}_n^*(\widehat{S}_*, S) \,\leq\, \frac{1+3\rho-2\rho^2}{1-3\rho}\, \min_{\gamma\in\Gamma} \mathcal{R}_n^*(\widehat{S}_\gamma, S) + \frac{1}{n}\, \mathcal{D}_n(\rho)\,, \tag{2.11}$$

where the term $\mathcal{D}_n(\rho)$ is defined in [10] such that

$$\lim_{n\to\infty} \frac{\mathcal{D}_n(\rho)}{n^\delta} = 0 \tag{2.12}$$

for each $\delta > 0$.

Remark 2.1. *The inequality* (2.11) *will be used to derive the upper bound for the robust risk* (1.6)*. It will be noted that the second summand in* (2.11) *when multiplied by the optimal rate $n^{2k/(2k+1)}$ tends to zero as $n \to \infty$ for each $k \geq 1$. Therefore, taking into account that $\rho \to 0$ as $n \to \infty$, the principal term in the upper bound is given by the minimal risk over the family of estimates $(\widehat{S}_\gamma)_{\gamma\in\Gamma}$. As is shown in [5], the efficient estimate enters this family. However one can not use this estimate because it depends on the unknown parameters $k \geq 1$ and $r > 0$ of the Sobolev ball. It is this fact that shows an adaptive role of the oracle inequality* (2.11) *which gives the asymptotic upper bound in the case when this information is not available.*

3 Main results

In this Section we will show, proceeding from (2.11), that the Pinsker constant for the robust risk (1.6) is given by the equation

$$R^*_{k,n} = ((2k+1)r)^{1/(2k+1)} \left(\frac{\sigma_n^* k}{(k+1)\pi}\right)^{2k/(2k+1)} . \tag{3.1}$$

It is well known that the optimal (minimax) rate for the Sobolev ball W_r^k is $n^{2k/(2k+1)}$ (see, for example, [13], [12]). We will see that asymptotically the robust risk of the model selection (2.4) normalized by this rate is bounded from above by $R^*_{k,n}$. Moreover, this bound can not be diminished if one considers the class of all admissible estimates for S.

Theorem 3.1. *Assume that, in model* (1.1)*, the distribution of* (ξ_t) *satisfies the condition* $\mathbf{C}^*$*). Then the robust risk* (1.6) *of the model selection estimator* $\widehat{S}_*$ *defined in* (2.4), (2.9)*, has the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \frac{1}{R^*_{k,n}} \sup_{S\in W_r^k} \mathcal{R}^*_n(\widehat{S}_*, S) \le 1 . \tag{3.2}$$

Now we obtain a lower bound for the robust risk (1.6). Let Π_n be the set of all estimators $\widetilde{S}_n$ measurable with respect to the sigma-algebra

$\sigma\{y_t\,,\, 0 \le t \le n\}$ generated by the process (1.1).

Theorem 3.2. *Under the conditions of Theorem 3.1*

$$\liminf_{n\to\infty} n^{2k/(2k+1)} \frac{1}{R^*_{k,n}} \inf_{\widetilde{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}^*_n(\widetilde{S}_n, S) \ge 1 . \tag{3.3}$$

Theorem 3.1 and Theorem 3.2 imply the following result

Corollary 3.3. *Under the conditions of Theorem 3.1*

$$\lim_{n\to\infty} n^{2k/(2k+1)} \frac{1}{R^*_{k,n}} \inf_{\widetilde{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}^*_n(\widetilde{S}_n, S) = 1 . \tag{3.4}$$

Remark 3.1. *The equation* (3.4) *means that the sequence* $R^*_{k,n}$ *defined by* (3.1) *is the Pinsker constant (see, for example, [13], [12]) for the model* (1.1).

4 Example

Let the process (ξ_t) be defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \,, \tag{4.1}$$

where $(w_t)_{t\geq 0}$ is a standard Brownian motion, $(z_t)_{t\geq 0}$ is a compound Poisson process defined as

$$z_t = \sum_{j=1}^{N_t} Y_j \,,$$

where $(N_t)_{t\geq 0}$ is a standard homogeneous Poisson process with unknown intensity $\lambda > 0$ and $(Y_j)_{j\geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E}\, Y_j = 0 \,, \quad \mathbf{E}\, Y_j^2 = 1 \quad \text{and} \quad \mathbf{E}\, Y_j^4 < \infty \,.$$

Substituting (4.1) in (1.3) yields

$$\mathbf{E}\, I_n(f) = (\varrho_1^2 + \varrho_2^2 \lambda)||f||^2 \,.$$

In order to meet the condition (1.4) the coefficients ϱ_1, ϱ_2 and the intensity $\lambda > 0$ must satisfy the inequality

$$\varrho_1^2 + \varrho_2^2 \lambda \leq \sigma^* \,. \tag{4.2}$$

Note that the coefficients ϱ_1, ϱ_2 and the intensity λ in (1.4) as well as σ^* may depend on n, i.e. $\varrho_i = \varrho_i(n)$ and $\lambda = \lambda(n)$.

As is stated in ([10], Theorem 2.2), the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) hold for the process (4.1) with $\sigma = \sigma(Q) = \varrho_1^2 + \varrho_2^2 \lambda$ defined in (2.6), $c_1^*(n) = 0$ and

$$c_2^*(n) \leq 4\sigma(\sigma + \varrho_2^2 \mathbf{E}\, Y_1^4) \,.$$

Let now $\mathcal{Q}_n^*$ be the family of distributions of the processes (4.1) with the coefficients satisfying the conditions (4.2) and

$$\varrho_2^2 \,\leq\, \sqrt{l_n} \,, \tag{4.3}$$

where the sequence l_n is taken from the definition of the set $\mathcal{P}_n^*$. Note that the distribution Q_0 belongs to $\mathcal{Q}_n^*$. One can obtain this distribution putting in (4.1) $\varrho_1 = \sqrt{\sigma^*}$ and $\varrho_2 = 0$. It will be noted that $\mathcal{Q}_n^* \subset \mathcal{P}_n^*$ if

$$4\sigma^*(\sigma^* + \sqrt{l}_n \mathbf{E}\, Y_1^4) \leq l_n \,.$$

5 Upper bound

5.1 Known smoothness

First we suppose that the parameters $k \geq 1$, $r > 0$ and σ^* in (1.4) are known. Let the family of admissible weighted least squares estimates $(\widehat{S}_\gamma)_{\gamma\in\Gamma}$ for the unknown function $S \in W_r^k$ be given (2.9), (2.10). Consider the pair

$$\alpha_0 = (k, t_0)$$

where $t_0 = [\overline{r}_n/\varepsilon]\varepsilon$, $\overline{r}_n = r/\sigma_n^*$ and ε satisfies the conditions in (2.8). Denote the corresponding weight sequence in Γ as

$$\gamma_0 = \gamma_{\alpha_0}\,. \tag{5.1}$$

Note that for sufficiently large n the parameter α_0 belongs to the set (2.9). In this section we obtain the upper bound for the empiric squared error of the estimator (1.6).

Theorem 5.1. *The estimator $\widehat{S}_{\gamma_0}$ satisfies the following asymptotic upper bound*

$$\limsup_{n\to\infty}\, n^{2k/(2k+1)}\,\frac{1}{R^*_{k,n}}\,\sup_{S\in W_r^k}\,\mathcal{R}_n^*\,(\widehat{S}_{\gamma_0}, S)\,\leq 1\,. \tag{5.2}$$

Proof. First by substituting the model (1.1) in the definition of $\widehat{\theta}_{j,n}$ in (2.1) we obtain

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,,$$

where the random variables $\xi_{j,n}$ are defined in (2.6). Therefore, by the definition of the estimators $\widehat{S}_\gamma$ in (2.1) we get

$$||\widehat{S}_{\gamma_0} - S||^2 = \sum_{j=1}^{n}(1-\gamma_0(j))^2\,\theta_j^2 - 2M_n + \sum_{j=1}^{n}\gamma_0^2(j)\,\xi_{j,n}^2$$

with

$$M_n \,=\, \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(1\,-\,\gamma_0(j))\,\gamma_0(j)\,\theta_j\,\xi_{j,n}\,.$$

It should be observed that

$$\mathbf{E}_{Q,S}\,M_n = 0$$

for any $Q \in \mathcal{Q}_n^*$. Further the condition (1.4) implies also the inequality $\mathbf{E}_Q\, \xi_{j,n}^2 \le \sigma_n^*$ for each distribution $Q \in \mathcal{Q}_n^*$. Thus,

$$\mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \,\le\, \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\, \theta_j^2 + \frac{\sigma_n^*}{n} \sum_{j=1}^{n} \gamma_0^2(j) \tag{5.3}$$

where $\iota_0 = j_0(\alpha_0)$. Denote

$$\upsilon_n = n^{2k/(2k+1)} \sup_{j \ge \iota_0} (1-\gamma_0(j))^2 / a_j\,,$$

where a_j is the sequence as defined in (1.8). Using this sequence we estimate the first summand in the right hand of (5.3) as

$$n^{2k/(2k+1)} \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\, \theta_j^2 \le \upsilon_n \sum_{j\ge 1} a_j\, \theta_j^2\,.$$

From here and (1.8) we obtain that for each $S \in W_r^k$

$$\Upsilon_{1,n}(S) = \, n^{2k/(2k+1)} \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\, \theta_j^2 \,\le\, \upsilon_n\, r\,.$$

Further we note that

$$\limsup_{n\to\infty} \, (\overline{r}_n)^{2k/(2k+1)} \, \upsilon_n \le \, \frac{1}{\pi^{2k}\, (\tau_k)^{2k/(2k+1)}}\,,$$

where the coefficient τ_k is given (2.10). Therefore, for any $\eta > 0$ and sufficiently large $n \ge 1$

$$\sup_{S\in W_r^k} \Upsilon_{1,n}(S) \,\le\, (1+\eta)\, (\sigma_n^*)^{2k/(2k+1)}\, \Upsilon_1^* \tag{5.4}$$

where

$$\Upsilon_1^* = \frac{r^{1/(2k+1)}}{\pi^{2k} (\tau_k)^{2k/(2k+1)}}\,.$$

To examine the second summand in the right hand of (5.2) we set

$$\Upsilon_{2,n} = \frac{1}{n^{1/(2k+1)}} \sum_{j=1}^{n} \gamma_0^2(j)\,.$$

Since by the condition (1.5)

$$\lim_{n\to\infty} \frac{t_0}{\overline{r}_n} = 1\,,$$

one gets

$$\lim_{n\to\infty} \frac{1}{(\overline{r}_n)^{1/(2k+1)}}\,\Upsilon_{2,n} = \Upsilon^*_2 \quad \text{with} \quad \Upsilon^*_2 = \frac{2(\tau_k)^{1/(2k+1)}\,k^2}{(k+1)(2k+1)}\,.$$

Note that by the definition (3.2)

$$(\sigma^*_n)^{2k/(2k+1)}\,\Upsilon^*_{1,n} + \sigma^*_n(\overline{r}_n)^{1/(2k+1)}\Upsilon^*_2 = R^*_{k,n}\,.$$

Therefore, for any $\eta > 0$ and sufficiently large $n \geq 1$

$$n^{2k/(2k+1)} \sup_{S\in W^k_r} \mathcal{R}^*_n(\widehat{S}_{\gamma_0}, S) \leq\ (1+\eta)R^*_{k,n}\,.$$

Hence Theorem 5.1. □

5.2 Unknown smoothness

Combining Theorem 5.1 and Theorem 2.1 yields Theorem 3.1.
□

6 Lower bound

First we obtain the lower bound for the risk (1.2) in the case of "white noise" model (1.1), when $\xi_t = \sqrt{\sigma^*}w_t$. As before let Q_0 denote the distribution of $(\xi_t)_{0\leq t\leq n}$ in $\mathcal{D}[0,n]$.

Theorem 6.1. *The risk* (1.2) *corresponding to the the distribution* Q_0 *in the model* (1.1) *has the following lower bound*

$$\liminf_{n\to\infty} n^{2k/(2k+1)} \inf_{\widetilde{S}_n\in\Pi_n} \frac{1}{R^*_{k,n}} \sup_{S\in W^k_r} \mathcal{R}_0(\widetilde{S}_n, S) \geq 1\,, \tag{6.1}$$

where $\mathcal{R}_0(\cdot,\cdot) = \mathcal{R}_{Q_0}(\cdot,\cdot)$.

Proof. The proof of this result proceeds along the lines of Theorem 4.2 from [6]. Let V be a function from $\mathbf{C}^\infty(\mathbb{R})$ such that $V(x) \geq 0$, $\int_{-1}^{1} V(x)\mathrm{d}x = 1$ and $V(x) = 0$ for $|x| \geq 1$. For each $0 < \eta < 1$ we introduce a smoother indicator of the interval $[-1+\eta, 1-\eta]$ by the formula

$$I_\eta(x) = \eta^{-1} \int_{\mathbb{R}} \mathbf{1}_{(|u| \leq 1-\eta)} G\left(\frac{u-x}{\eta}\right) \mathrm{d}u\,.$$

It will be noted that $I_\eta \in \mathcal{C}^\infty(\mathbb{R})$, $0 \leq I_\eta \leq 1$ and for any $m \geq 1$ and positive constant $c > 0$

$$\lim_{\eta \to 0} \sup_{\{f\,:\,|f|_* \leq c\}} \left| \int_{\mathbb{R}} f(x) I_\eta^m(x)\,\mathrm{d}x \,-\, \int_{-1}^{1} f(x)\,\mathrm{d}x \right| = 0 \tag{6.2}$$

where $|f|_* = \sup_{-1 \leq x \leq 1} |f(x)|$. Further, we need the trigonometric basis in $\mathcal{L}_2[-1,1]$, that is

$$e_1(x) = 1/\sqrt{2}\,,\ e_j(x) \,=\, Tr_j(\pi[j/2]x)\,,\ j \geq 2\,. \tag{6.3}$$

Now we will construct of a family of approximation functions for a given regression function S following [6]. For fixed $0 < \varepsilon < 1$ one chooses the bandwidth function as

$$h = h_n = (\upsilon_\varepsilon^*)^{\frac{1}{2k+1}}\, N_n\, n^{-\frac{1}{2k+1}} \tag{6.4}$$

with

$$\upsilon_\varepsilon^* = \frac{\sigma_n^* k \pi^{2k}}{(1-\varepsilon) r 2^{2k+1}(k+1)(2k+1)} \quad \text{and} \quad N_n = \ln^4 n$$

and considers the partition of the interval $[0,1]$ with the points $\widetilde{x}_m = 2hm$, $1 \leq m \leq M$, where

$$M = [1/(2h)] - 1\,.$$

For each interval $[\widetilde{x}_m - h, \widetilde{x}_m + h]$ we specify the smoothed indicator as $I_\eta(v_m(x))$, where $v_m(x) = (x - \widetilde{x}_m)/h$. The approximation function for $S(t)$ is given by

$$S_{z,n}(x) = \sum_{m=1}^{M} \sum_{j=1}^{N} z_{m,j}\, D_{m,j}(x)\,, \tag{6.5}$$

where $z = (z_{m,j})_{1 \leq m \leq M\,, 1 \leq j \leq N}$ is an array of real numbers;

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

are orthogonal functions on $[0,1]$.

Note that the set W_r^k is a subset of the ball

$$\mathbf{B}_r = \{f \in \mathcal{L}_2[0,1] \,:\, ||f||^2 \le r\}\,.$$

Now for a given estimate $\widetilde{S}_n$ we construct its projection in $\mathcal{L}_2[0,1]$ into $\mathbf{B}_r$

$$\widetilde{F}_n := Pr_{\mathbf{B}_r}(\widetilde{S}_n)\,.$$

In view of the convexity of the set $\mathbf{B}_r$ one has

$$||\widetilde{S}_n - S||^2 \ge ||\widetilde{F}_n - S||^2$$

for each $S \in W_r^k \subset \mathbf{B}_r$.

From here one gets the following inequalities for the the risk (1.2)

$$\sup_{S\in W_r^k} \mathcal{R}_0(\widetilde{S}_n, S) \ge \sup_{S\in W_r^k} \mathcal{R}_0(\widetilde{F}_n, S) \ge \sup_{\{z\in\mathbb{R}^d\,:\,S_{z,n}\in W_r^k\}} \mathcal{R}_0(\widetilde{F}_n, S)\,,$$

where $d = MN$.

In order to continue this chain of estimates we need to introduce a special prior distribution on $\mathbb{R}^d$. Let $\kappa = (\kappa_{m,j})_{1\le m\le M\,,1\le j\le N}$ be a random array with the elements

$$\kappa_{m,j} \;=\; t_{m,j}\,\kappa^*_{m,j}\,, \tag{6.6}$$

where $\kappa^*_{m,j}$ are i.i.d. gaussian $\mathcal{N}(0,1)$ random variables and the coefficients

$$t_{m,j} = \frac{\sqrt{\sigma^*_n y^*_j}}{\sqrt{nh}}\,.$$

We choose the sequence $(y^*_j)_{1\le j\le N}$ in the same way as in [6](see (8.11)) , i.e.

$$y^*_j \;=\; N_n^k\, j^{-k} - 1\,.$$

We denote the distribution of κ by μ_κ. We will consider it as a prior distribution of the random parametric regression $S_{\kappa,n}$ which is obtained from (6.5) by replacing z with κ.

Besides we introduce

$$\Xi_n = \left\{ z\in\mathbb{R}^d \,:\, \max_{1\le m\le M}\,\max_{1\le j\le N}\,\frac{|z_{m,j}|}{t_{m,j}} \le \ln n \right\}\,. \tag{6.7}$$

By making use of the distribution μ_κ, one obtains

$$\sup_{S\in W_r^k} \mathcal{R}_0(\widetilde{S}_n, S) \geq \int_{\{z\in\mathbb{R}^d\,:\,S_{z,n}\in W_r^k\}\cap\Xi_n} \mathbf{E}_{Q_0,S_{z,n}} ||\widetilde{F}_n - S_{z,n}||^2 \,\mu_\kappa(\mathrm{d}z)\,.$$

Further we introduce the Bayes risk as

$$\widetilde{\mathcal{R}}(\widetilde{F}_n) = \int_{\mathbb{R}^d} \mathcal{R}_0(\widetilde{F}_n, S_{z,n})\mu_\kappa(\mathrm{d}z)$$

and noting that $||\widetilde{F}_n||^2 \leq r$ we come to the inequality

$$\sup_{S\in W_r^k} \mathcal{R}_0(\widetilde{S}_n, S) \;\geq\; \widetilde{\mathcal{R}}(\widetilde{F}_n) \;-\; \varpi_n \tag{6.8}$$

where

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

By Proposition A.1 from Appendix A.1 one has, for any $p > 0$,

$$\lim_{n\to\infty} n^p\,\varpi_n \,= 0\,.$$

Now we consider the first term in the right-hand side of (6.8). To obtain a lower bound for this term we use the $\mathcal{L}_2[0,1]$-orthonormal function family $(G_{m,j})_{1\leq m\leq M, 1\leq j\leq N}$ which is defined as

$$G_{m,j}(x) = \frac{1}{\sqrt{h}} e_j\left(v_m(x)\right)\mathbf{1}_{(|v_m(x)|\leq 1)}\,.$$

We denote by $\widetilde{g}_{m,j}$ and $g_{m,j}(z)$ the Fourier coefficients for functions $\widetilde{F}_n$ and S_z, respectively, i.e.

$$\widetilde{g}_{m,j} = \int_0^1 \widetilde{F}_n(x) G_{m,j}(x)\mathrm{d}x \quad \text{and} \quad g_{m,j}(z) = \int_0^1 S_{z,n}(x)\, G_{m,j}(x)\mathrm{d}x\,.$$

Now it is easy to see that

$$||\widetilde{F}_n - S_{z,n}||^2 \geq \sum_{m=1}^{M}\sum_{j=1}^{N} \left(\widetilde{g}_{m,j} - g_{m,j}(z)\right)^2.$$

Let us introduce the functionals $K_j(\cdot) : \mathcal{L}_1[-1,1] \to \mathbb{R}$ as

$$K_j(f) = \int_{-1}^{1} e_j^2(v)\, f(v)\, \mathrm{d}v\,.$$

In view of (6.5) we obtain that

$$\frac{\partial}{\partial z_{m,j}} g_{m,j}(z) = \int_0^1 D_{m,j}(x)\, G_{m,j}(x)\, \mathrm{d}x = \sqrt{h}\, K_j(I_\eta)\,.$$

Now Proposition A.2 implies

$$\begin{aligned}\widetilde{\mathcal{R}}(\widetilde{F}_n) \;&\geq \sum_{m=1}^{M}\sum_{j=1}^{N} \int_{\mathbb{R}^d} \mathbf{E}_{S_{z,n}} (\widetilde{g}_{m,j} - g_{m,j}(z))^2 \mu_\kappa(\mathrm{d}z) \\ &\geq\; h \sum_{m=1}^{M}\sum_{j=1}^{N} \frac{\sigma^* K_j^2(I_\eta)}{K_j(I_\eta^2)\, nh \,+\, t_{m,j}^{-2}\sigma^*}\,.\end{aligned}$$

Therefore, taking into account the definition of the coefficients $(t_{m,j})$ in (6.6) we get

$$\widetilde{\mathcal{R}}(\widetilde{F}_n) \;\geq\; \frac{\sigma^*}{2nh} \sum_{j=1}^{N} \tau_j(\eta, y_j^*)$$

with

$$\tau_j(\eta, y) = \frac{K_j^2(I_\eta) y}{K_j(I_\eta^2) y + 1}\,.$$

Moreover, the limit equality (6.2) implies directly

$$\lim_{\eta\to 0} \sup_{j\geq 1} \sup_{y\geq 0} \left| \frac{(y+1)\tau_j(\eta, y)}{y} - 1 \right| = 0\,.$$

Therefore, we can write that for any $\nu > 0$

$$\widetilde{\mathcal{R}}(\widetilde{F}_n) \;\geq\; \frac{\sigma^*}{2nh(1+\nu)} \sum_{j=1}^{N} \frac{y_j^*}{y_j^* + 1}\,.$$

It is easy to check directly that

$$\lim_{n\to\infty} \frac{\sigma_n^*}{2nhR_{k,n}^*} \sum_{j=1}^{N} \frac{y_j^*}{y_j^* + 1} = (1-\varepsilon)^{\frac{1}{2k+1}}\,,$$

where the coefficient $R^*_{k,n}$ is defined in (3.1). Therefore, (6.8) implies for any $0 < \varepsilon < 1$

$$\liminf_{T\to\infty} \inf_{\widehat{S}_n} n^{\frac{2k}{2k+1}} \frac{1}{R^*_{k,n}} \sup_{S\in W^k_r} \mathcal{R}_0(\widehat{S}_n, S) \ \geq \ (1-\varepsilon)^{\frac{1}{2k+1}} .$$

Taking here limit as $\varepsilon \to 0$ implies Theorem 6.1. □

7 Appendix

A.1 Properties of the parametric family (6.5)

In this subsection we consider the sequence of the random functions $S_{\kappa,n}$ defined in (6.5) corresponding to the random array $\kappa = (\kappa_{m,j})_{1\leq m\leq M, 1\leq j\leq N}$ given in (6.6).

Proposition A.1. *For any $p > 0$*

$$\lim_{n\to\infty} n^p \lim_{n\to\infty} \mathbf{E}\, ||S_{\kappa,n}||^2 \left(\mathbf{1}_{\{S_{\kappa,n}\notin W^k_r\}} + \mathbf{1}_{\Xi^c_n} \right) = 0 .$$

This proposition follows directly from Proposition 6.4 in [7].

A.2 Lower bound for parametric "white noise" models.

In this subsection we prove some version of the van Trees inequality from [8] for the following model

$$\mathrm{d}y_t = S(t,z)\mathrm{d}t + \sqrt{\sigma^*}\,\mathrm{d}w_t\,, \quad 0 \leq t \leq n\,, \tag{A.1}$$

where $z = (z_1, \ldots, z_d)'$ is vector of unknown parameters, $w = (w_t)_{0\leq t\leq T}$ is a Winier process. We assume that the function $S(t,z)$ is a linear function with respect to the parameter z, i.e.

$$S(t,z) = \sum_{j=1}^{d} z_j\, S_j(t)\,. \tag{A.2}$$

Moreover, we assume that the functions $(S_j)_{1\leq j\leq d}$ are continuous.

Let Φ be a prior density in $\mathbb{R}^d$ having the following form:

$$\Phi(z) = \Phi(z_1, \ldots, z_d) = \prod_{j=1}^{d} \varphi_j(z_j)\,,$$

where φ_j is some continuously differentiable density in $\mathbb{R}$. Moreover, let $g(z)$ be a continuously differentiable $\mathbb{R}^d \to \mathbb{R}$ function such that for each $1 \le j \le d$

$$\lim_{|z_j|\to\infty} g(z)\,\varphi_j(z_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |g_j'(z)|\,\Phi(z)\,\mathrm{d}z < \infty\,, \tag{A.3}$$

where

$$g_j'(z) = \frac{\partial g(z)}{\partial z_j}\,.$$

Let now $\mathcal{X}_n = \mathbf{C}[0,T]$ and $\mathcal{B}(\mathcal{X}_n)$ be σ - field generated by cylindric sets in $\mathcal{X}_n$.

For any $\mathcal{B}(\mathcal{X}_n) \bigotimes \mathcal{B}(\mathbb{R}^d)$- measurable integrable function $\xi = \xi(x,\theta)$ we denote

$$\widetilde{\mathbf{E}}\xi = \int_{\mathbb{R}^d} \int_{\mathcal{X}} \xi(y,z)\,\mu_z(\mathrm{d}y)\,\Phi(z)\mathrm{d}z\,,$$

where μ_z is distribution of the process (A.1) in $\mathcal{X}_n$. Let now $\nu = \mu_0$ be the distribution of the process $(\sigma^* w_t)_{0\le t\le n}$ in $\mathcal{X}$. It is clear (see, for example [11]) that $\mu_z << \nu$ for any $z \in \mathbb{R}^d$. Therefore, we can use the measure ν as a dominated measure, i.e. for the observations (A.1) in $\mathcal{X}_n$ we use the following likelihood function

$$f(y,z) = \frac{\mathrm{d}\mu_z}{\mathrm{d}\nu} = \exp\left\{ \int_0^n \frac{S(t,z)}{\sqrt{\sigma^*}}\,\mathrm{d}y_t - \int_0^n \frac{S^2(t,z)}{2\sigma^*}\,\mathrm{d}t \right\}. \tag{A.4}$$

Proposition A.2. *For any square integrable function $\widehat{g}_n$ measurable with respect to $\sigma\{y_t\,,\, 0 \le t \le n\}$ and for any $1 \le j \le d$ the following inequality holds*

$$\tilde{\mathbf{E}}(\widehat{g}_n - g(z))^2 \ge \frac{\sigma^* B_j^2}{\int_0^n S_j^2(t)\,\mathrm{d}t + \sigma^* I_j}\,, \tag{A.5}$$

where

$$B_j = \int_{\mathbb{R}^d} g_j'(z)\,\Phi(z)\,\mathrm{d}z \quad \textit{and} \quad I_j = \int_{\mathbb{R}} \frac{\dot{\varphi}_j^2(z)}{\varphi_j(z)}\,\mathrm{d}z\,.$$

Proof. First of all note that the density (A.3) is bounded with respect to $\theta_j \in \mathbb{R}$ for any $1 \le j \le d$, i.e. for any $y = (y_t)_{0\le t\le n} \in \mathcal{X}$

$$\limsup_{|z_j|\to\infty} f(y,z) \,<\, \infty\,.$$

Therefore, putting

$$\Psi_j = \Psi_j(y,z) = \frac{\partial}{\partial\theta_j}\,\ln(f(y,z)\Phi(z))$$

and taking into account condition (A.3) by integration by parts one gets

$$\begin{aligned}\widetilde{\mathbf{E}}\,((\widehat{g}_T - g(z))\Psi_j) &= \int_{\mathbb{R}^N\times\mathbb{R}^d} (\widehat{g}_T(y) - g(z))\frac{\partial}{\partial z_j}\,(f(y,z)\Phi(z))\,\mathrm{d}z\,\mathrm{d}\nu(y)\\ &= \int_{\mathbb{R}^N\times\mathbb{R}^d} g_j'(z)\,f(y,z)\Phi(z)\,\mathrm{d}z\,\mathrm{d}\nu(y) = B_j\,.\end{aligned}$$

Now by the Bounyakovskii-Cauchy-Schwarz inequality we obtain the following lower bound for the quiadratic risk

$$\widetilde{\mathbf{E}}(\widehat{g}_T - g(z))^2 \ge \frac{B_j^2}{\widetilde{\mathbf{E}}\Psi_j^2}\,.$$

Note that from (A.4) it is easy to deduce that under the distribution μ_z

$$\begin{aligned}\frac{\partial}{\partial z_j}\,\ln f(y,z) &= \int_0^n \frac{S_j(t)}{\sqrt{\sigma^*}}\,\mathrm{d}y_t - \int_0^n \frac{S(t,z)S_j(t)}{\sigma^*}\,\mathrm{d}t\\ &= \int_0^n \frac{S_j(t)}{\sqrt{\sigma^*}}\,\mathrm{d}w_t\,.\end{aligned}$$

This implies directly

$$\mathbf{E}_z\,\frac{\partial}{\partial z_j}\,\ln f(y,z) \,= 0$$

and

$$\mathbf{E}_z\,\left(\frac{\partial}{\partial z_j}\,\ln f(y,z)\right)^2 = \frac{1}{\sigma^*}\int_0^n S_j^2(t)\,\mathrm{d}t\,.$$

Therefore,

$$\widetilde{\mathbf{E}}\Psi_j^2 = \frac{1}{\sigma^*}\int_0^n S_j^2(t)\,\mathrm{d}t + I_j\,.$$

Hence Proposition A.2. □

References

[1] Brua, J. (2009) Asymptotically efficient estimators for nonparametric heteroscedastic regression models. *Stat. Methodol.* 6(1), p. 47-60.

[2] Fourdrinier, D. and Pergamenshchikov, S. (2007) Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*, **59** (3), p. 435-464.

[3] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotically efficient estimates for non parametric regression models.*Statistics and Probability Letters*, v. 76 , 8, p. 852-860

[4] Galtchouk, L. and Pergamenshchikov, S. (2004) Nonparametric sequential estimation of the drift in diffusion processes. *Mathematical Methods of Statistics*, **13**, 1, 25-49.

[5] Galtchouk, L. and Pergamenshchikov, S. (2009) Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, 2009, 21, 1, p. 1-16

[6] Galtchouk, L. and Pergamenshchikov, S. (2009) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *Journal of Korean Statistical Society*,

http://ees.elsivier.com/jkss

[7] Galtchouk, L. and Pergamenshchikov, S. (2009) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection.

http://hal.archives-ouvertes.fr/hal-00326910/fr/

[8] Gill, R.D. and Levit, B.Y. Application of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1** 59-79 (1995)

[9] Konev, V.V. and Pergamenshchikov, S.M. (2008) General model selection estimation of a periodic regression with a Gaussian noise. - *Annals of the Institute of Statistical Mathematics*, 2008, Available online at *http://dx.doi.org/10.1007/s10463-008-0193-1*

[10] Konev, V.V. and Pergamenshchikov, S.M. (2009) Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities. *Vestnik Tomskogo Universiteta, Mathematics and Mechanics.*

[11] Liptser, R. Sh. and Shiryaev, A.N. (1977) *Statistics of Random Processes. I. General theory.* New York : Springer.

[12] Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in $\mathbf{L}_2$. *Ann. Statist.*, **13**, p. 984-997.

[13] Pinsker, M.S. (1981) Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission information*, **17** 120–133