Adaptive efficient robust estimation for nonparametric autoregressive models

Ouerdia Arkoun, Jean-Yves Brua, Serguei Pergamenshchikov

Adaptive efficient robust estimation for nonparametric autoregressive models *

Ouerdia Arkoun †, Jean-Yves Brua ‡and
Serguei Pergamenshchikov§

Abstract

In this paper for the first time the adaptive efficient estimation problem for nonparametric autoregressive models has been studied. First of all, through the Van Trees inequality the sharp bound for the robust quadratic risks, i.e. the Pinsker constant (see, for example, in [19]), in explicit form has been obtained. Then, through the sharp oracle inequalities method developed in [4] for non parametric autoregressions an adaptive efficient model selection procedure is proposed, i.e. such for which the upper bound of its robust quadratic risk coincides with the obtained Pinsker constant.

MSC: primary 62G08, secondary 62G05
Keywords: Non-parametric estimation; Non parametric autoregresion; Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequalities.

*This work was supported by the RSF grant 17-11-01049 (National Research Tomsk State University).

†Sup'Biotech, Laboratoire BIRL, 66 Rue Guy Moquet, 94800 Villejuif, France and Laboratoire de Mathématiques Raphael Salem, Normandie Université, UMR 6085 CNRS-Université de Rouen, France, e-mail: ouerdia.arkoun@gmail.com

‡Laboratoire de Mathématiques Raphael Salem, Normandie Université, UMR 6085 CNRS- Université de Rouen, France, e-mail : jean-yves.brua@univ-rouen.fr

§Laboratoire de Mathématiques Raphael Salem, UMR 6085 CNRS- Université de Rouen Normandie, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, National Research Tomsk State University, e-mail: Serge.Pergamenshchikov@univ-rouen.fr

1 Introduction

1.1 Model

In this paper we consider the nonparametric autoregression model defined as

$$y_k = S(x_k)y_{k-1} + \xi_k \quad \text{and} \quad x_k = a + \frac{k(b-a)}{n}\,, \tag{1.1}$$

where $S(\cdot) \in \mathbf{L}_2[a,b]$ is unknown function, $a < b$ are fixed known constants, $1 \le k \le n$, the initial value y_0 is a constant and the noise $(\xi_k)_{k\ge1}$ is i.i.d. sequence of unobservable random variables with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. In the sequel we denote by $\mathbf{p}$ the distribution density of the random variable ξ_1.

The problem is to estimate the function $S(\cdot)$ on the basis of the observations $(y_k)_{1\le k\le n}$ under the condition that the noise distribution $\mathbf{p}$ is unknown and belongs to some noise distributions class $\mathcal{P}$. There is a number of papers which consider these models such as [6], [7] and [5]. In all these papers, the authors propose some asymptotic (as $n \to \infty$) methods for different identification studies without considering optimal estimation issues. Firstly, minimax estimation problems for the model (1.1) has been treated in [2] and [18] in the nonadaptive case, i.e. for the known regularity of the function S. Then, in [1] and [3] it is proposed to use the sequential analysis method for the adaptive pointwise estimation problem in the case when the Hölder regularity is unknown.

1.2 Main contributions

In this paper we consider the adaptive estimation problem for the quadratic risk defined as

$$\mathcal{R}_{\mathbf{p}}(\widehat{S}_n, S) \;=\; \mathbf{E}_{\mathbf{p},S}\|\widehat{S}_n - S\|^2\,, \quad \|S\|^2 = \int_a^b S^2(x)\mathrm{d}x\,, \tag{1.2}$$

where $\widehat{S}_n$ is an estimator of S based on observations $(y_k)_{1\le k\le n}$ and $\mathbf{E}_{\mathbf{p},S}$ is the expectation with respect to the distribution law $\mathbf{P}_{\mathbf{p},S}$ of the process $(y_k)_{1\le k\le n}$ given the distribution density p and the coefficient S. Moreover, taking into account that the distribution p is unknown, we use the robust nonparametric estimation approach proposed in [8]. To this end we set the robust risk as

$$\mathcal{R}^*(\widehat{S}_n, S) \;=\; \sup_{\mathbf{p}\in\mathcal{P}} \mathcal{R}_{\mathbf{p}}(\widehat{S}_n, S)\,, \tag{1.3}$$

where $\mathcal{P}$ is a family of the distributions defined in Section 2.

To estimate the function S in model (1.1) we make use of the model selection procedures proposed in [4] based on the family of the optimal poinwise truncated sequential estimators from [3] for which using the model selection method developed in [9] a sharp oracle inequality is shown. In this paper, using this inequality we show that the model selection procedure is efficient in adaptive setting for the robust quadratic risks (1.3). To this end, first of all we have to study the sharp lower bound for the these risks, i.e. we have to study the best potential accuracy estimation for the model (1.1) which is called the Pinsker constant. For this we use the approach proposed in [10] - [11] which is based the Van-Trees inequality. It turns out that for the model (1.1) the Pinsker constant has the same form as for the filtration signal problem in the "signal - white noise" model studied in [19] but with new coefficient which equals to the optimal variance given by the Hajek - Le Cam inequality for the parametric model (1.1). This is the new result in the efficient non parametric estimation theory for the statistical models with dependent observations. Then, using the oracle inequality from [4] and the weight least square estimation method we show that for the model selection procedure with the Pinsker weight coefficients the upper bound asymptotically coincides with the obtained Pinsker constant without using the regularity properties of the unknown functions, i.e. it is efficient in adaptive setting with respect to the robust risks (1.3).

1.3 Plan of the paper

The paper is organized as follows. In Section 2 we give all conditions and construct the sequential point-wise estimation procedures. to pass from the auto-regression model to the corresponding regression model. In Section 3 we construct the model selection procedure based on the sequential estimators from Section 2. In Section 4 we announce the main results. In Section 5 we show the Van - Trees inequality for the model (1.1). In section 6 we obtain the lower bound for the robust risks. In Section 7 we obtain the upper bounds for the robust risks. In Appendix A we give the all auxiliary and technic tools.

2 Sequential procedures.

As in [3] we assume that in the model (1.1) the i.i.d. random variables $(\xi_k)_{k\geq 1}$ have a density p (with respect to the Lebesgue measure) from the functional

class $\mathcal{P}$ defined as

$$\mathcal{P} := \Big\{ p \geq 0 \, : \, \int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1\,, \quad \int_{-\infty}^{+\infty} x\,p(x)\,\mathrm{d}x = 0\,, \\ \int_{-\infty}^{+\infty} x^2\,p(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \sup_{k\geq 1} \frac{\int_{-\infty}^{+\infty} |x|^{2k}\,p(x)\,\mathrm{d}x}{\varsigma^k (2k-1)!!} \leq 1 \Big\}\,, \quad (2.1)$$

where $\varsigma \geq 1$ is some fixed parameter, which may be a function of the number observation n, i.e. $\varsigma = \varsigma(n)$, such that for any $\mathbf{b} > 0$

$$\lim_{n\to\infty} \frac{\varsigma(n)}{n^{\mathbf{b}}} = 0\,. \quad (2.2)$$

Note that the $(0,1)$-Gaussian density belongs to $\mathcal{P}$. In the sequel we denote this density by p_0. It is clear that for any $q > 0$

$$\mathbf{m}_q^* = \sup_{\mathbf{p}\in\mathcal{P}} \mathbf{E}_{\mathbf{p}}\,|\xi_1|^q < \infty\,, \quad (2.3)$$

where $\mathbf{E}_{\mathbf{p}}$ is the expectation with respect to the density p from $\mathcal{P}$. To obtain the stable (uniformly with respect to the function S) model (1.1), we assume that for some fixed $0 < \epsilon < 1$ and $L > 0$ the unknown function S belongs to the ε - *stability set* introduced in [3] as

$$\Theta_{\epsilon,L} = \Big\{ S \in \mathbf{C}_1([a,b],\mathbb{R}) : |S|_* \leq 1-\epsilon \quad \text{and} \quad |\dot{S}|_* \leq L \Big\}\,, \quad (2.4)$$

where $\mathbf{C}_1[a,b]$ is the Banach space of continuously differentiable $[a,b] \to \mathbb{R}$ functions and $|S|_* = \sup_{a\leq x\leq b} |S(x)|$.

We will use as a basic procedures the point wise procedure from [3] at the points $(z_l)_{1\leq l\leq d}$ defined as

$$z_l = a + \frac{l}{d}(b-a)\,, \quad (2.5)$$

where d is an integer value function of n, i.e. $d = d_n$, such that

$$\lim_{n\to\infty} \frac{d_n}{\sqrt{n}} = 1\,. \quad (2.6)$$

So we propose to use the first ι_l observations for the auxiliary estimation of $S(z_l)$. We set

$$\widehat{S}_l = \frac{1}{A_{\iota_l}} \sum_{j=1}^{\iota_l} Q_{l,j}\,y_{j-1}\,y_j\,, \quad A_{\iota_l} = \sum_{j=1}^{\iota_l} Q_{l,j}\,y_{j-1}^2\,, \quad (2.7)$$

where $Q_{l,j} = Q(u_{l,j})$ and the kernel $Q(\cdot)$ is the indicator function of the interval $[-1;1]$, i.e. $Q(u) = \mathbf{1}_{[-1,1]}(u)$. The points $(u_{l,j})$ are defined as

$$u_{l,j} = \frac{x_j - z_l}{h}\,. \tag{2.8}$$

Note that to estimate $S(z_l)$ on the basis of the kernel estimate with the kernel Q we use only the observations $(y_j)_{k_{1,l}\le j\le k_{2,l}}$ from the h - neighbor of the point z_l, i.e.

$$k_{1,l} = [n\widetilde{z}_l - n\widetilde{h}] + 1 \quad \text{and} \quad k_{2,l} = [n\widetilde{z}_l + n\widetilde{h}] \wedge n\,, \tag{2.9}$$

where $\widetilde{z}_l = (z_l - a)/(b-a)$ and $\widetilde{h} = h/(b-a)$. Note that, only for the last point $z_d = b$ the $k_{2,d} = n$. We chose ι_l in (2.7) as

$$\iota_l = k_{1,l} + \mathbf{q} \quad \text{and} \quad \mathbf{q} = \mathbf{q}_n = [(n\widetilde{h})^{\mu_0}] \tag{2.10}$$

for some $0 < \mu_0 < 1$. In the sequel for any $0 \le k < m \le n$ we set

$$A_{k,m} = \sum_{j=k+1}^{m} Q_{l,j}\, y_{j-1}^2 \quad \text{and} \quad A_m = A_{0,m}\,. \tag{2.11}$$

Next, similarly to [1], we use a some kernel sequential procedure based on the observations $(y_j)_{\iota_l\le j\le n}$. To transform the kernel estimator in the linear function of observations and we replace the number of observations n by the following stopping time

$$\tau_l = \inf\{\iota_l + 1 \le k \le k_{2,l} \,:\, A_{\iota_l,k} \ge H_l\}\,, \tag{2.12}$$

where $\inf\{\emptyset\} = k_{2,l}$ and the positive threshold H_l will be chosen as a positive random variable measurable with respect to the σ - field $\{y_1, \ldots, y_{\iota_l}\}$.
Now we define the sequential estimator as

$$S_l^* = \frac{1}{H_l}\left(\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j}\, y_{j-1}\, y_j \;+\; \varkappa_l\, Q(u_{l,\tau_l})\, y_{\tau_l-1}\, y_{\tau_l}\right)\mathbf{1}_{\Gamma_l}\,, \tag{2.13}$$

where $\Gamma_l = \{A_{\iota_l,k_{2,l}-1} \ge H_l\}$ and the correcting coefficient $0 < \varkappa_l \le 1$ on this set is defined as

$$A_{\iota_l,\tau_l-1} + \varkappa_l^2\, Q(u_{l,\tau_l}) y_{\tau_l-1}^2 = H_l\,. \tag{2.14}$$

Note that, to obtain the efficient kernel estimate of $S(z_l)$ we need to use the all $k_{2,l} - \iota_l - 1$ observations. Similarly to [15], one can show that $\tau_l \approx \gamma_l H_l$ as $H \to \infty$, where

$$\gamma_l = 1 - S^2(z_l)\,. \tag{2.15}$$

Therefore, one needs to chose H as $(k_{2,l} - \iota_l - 1)/\gamma_l$. Taking into account that the coefficients γ_l are unknown we define the threshold H_l as

$$H_l = \frac{1-\widetilde{\epsilon}}{\widetilde{\gamma}_l}(k_{2,l} - \iota_l - 1) \quad \text{and} \quad \widetilde{\epsilon} = \frac{1}{2 + \ln n}\,, \tag{2.16}$$

where $\widetilde{\gamma}_l = 1 - \widetilde{S}^2_{\iota_l}$ and $\widetilde{S}_{\iota_l}$ is the projection of the estimator $\widehat{S}_{\iota_l}$ in the interval $]-1+\widetilde{\epsilon}, 1-\widetilde{\epsilon}[$, i.e.

$$\widetilde{S}_{\iota_l} = \min(\max(\widehat{S}_{\iota_l}, -1+\widetilde{\epsilon}), 1-\widetilde{\epsilon})\,. \tag{2.17}$$

To obtain the uncorrelated stochastic terms in the kernel estimators for $S(z_l)$ we chose the bandwidth h as

$$h = \frac{b-a}{2d}\,. \tag{2.18}$$

As to the estimator $\widehat{S}_{\iota_l}$, we can show the following property.

Proposition 2.1. *The convergence rate in probability of the estimator* (2.17) *is more rapid than any power function, i.e. for any* $\mathbf{b} > 0$

$$\lim_{n\to\infty} n^{\mathbf{b}} \max_{1\le l\le d} \sup_{S\in\Theta_{\epsilon,L}} \sup_{\mathbf{p}\in\mathcal{P}} \mathbf{P}_{\mathbf{p},S}\left(|\widetilde{S}_{\iota_l} - S(z_l)| > \epsilon_0\right) = 0\,, \tag{2.19}$$

where $\epsilon_0 = \epsilon_0(n) \to 0$ *as* $n \to \infty$ *such that* $\lim_{n\to\infty} n^{\check{\delta}}\epsilon_0 = \infty$ *for any* $\check{\delta} > 0$.

Now we set

$$Y_l = S^*_{H,h}(z_l)\mathbf{1}_\Gamma \quad \text{and} \quad \Gamma = \cap_{l=1}^d \Gamma_l\,. \tag{2.20}$$

Using the convergence (2.19) we study the probability properties of the set Γ in the following proposition.

Proposition 2.2. *For any* $\mathbf{b} > 0$ *the probability of the set* Γ *satisfies the following asymptotic equality*

$$\lim_{n\to\infty} n^{\mathbf{b}} \sup_{S\in\Theta_{\epsilon,L}} \mathbf{P}_{\mathbf{p},S}\left(\Gamma^c\right) = 0\,. \tag{2.21}$$

In view of this proposition we can negligible the set Γ^c. So, using the estimators (2.20) on the set Γ we obtain the discrete time regression model

$$Y_l = S(z_l) + \zeta_l \quad \text{and} \quad \zeta_l = \xi^*_l + \varpi_l\,, \tag{2.22}$$

in which

$$\xi^*_l = \frac{\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j}\, y_{j-1}\, \xi_j + \varkappa_l\, Q(u_{l,\tau_l})\, y_{\tau_l-1}\, \xi_{\tau_l}}{H_l} \tag{2.23}$$

and $\varpi_l = \varpi_{1,l} + \varpi_{2,l}$, where

$$\varpi_{1,l} = \frac{\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j}\, y^2_{j-1}\Delta_{l,j} + \varkappa^2_l\, Q(u_{l,\tau_l})\, y^2_{\tau_l-1}\Delta_{l,\tau_l}}{H_l}\,, \quad \Delta_{l,j} = S(x_j) - S(z_l)$$

and

$$\varpi_{2,l} = \frac{(\varkappa_l - \varkappa^2_l)\, Q(u_{l,\tau_l})\, y^2_{\tau_l-1}\, S(x_{\tau_l})}{H_l}\,.$$

Note that in the model sec:In.1-11-1R the random variables $(\xi^*_j)_{1\le j\le d}$ are defined only on the set Γ. By the technical reasons we need the definitions for these variables on the set Γ^c was well. To this end for any $j \ge 1$ we set

$$\check{Q}_{l,j} = Q_{l,j}\, y_{j-1}\, \mathbf{1}_{\{j<k_{2,l}\}} + \sqrt{H_l}\, Q_{l,j}\, \mathbf{1}_{\{j=k_{2,l}\}} \tag{2.24}$$

and $\check{A}_{\iota_l,m} = \sum_{j=\iota_l+1}^{m} \check{Q}^2_{l,j}$. Note, that for any $j \ge 1$ and $l \ne m$

$$\check{Q}_{l,j}\, \check{Q}_{m,j} = 0\,. \tag{2.25}$$

and $\check{A}_{\iota_l,k_{2,l}} \ge H_l$. So we can modify now stopping time (2.12) as

$$\check{\tau}_l = \inf\{k \ge \iota_l + 1 : \check{A}_{\iota_l,k} \ge H_l\}\,. \tag{2.26}$$

Obviously, $\check{\tau}_l \le k_{2,l}$ and $\check{\tau}_l = \tau_l$ on the set Γ for any $1 \le l \le d$. Now similarly to (2.14) we define the correction coefficient as

$$\check{A}_{\iota_l,\check{\tau}_l-1} + \check{\varkappa}^2_l\, \check{Q}^2_{l,\check{\tau}_l} = H_l\,. \tag{2.27}$$

It is clear that $0 < \check{\varkappa}_l \le 1$ and $\check{\varkappa}_l = \varkappa_l$ on the set Γ for $1 \le l \le d$. Using this coefficient we set

$$\eta_l = \frac{\sum_{j=\iota_l+1}^{\check{\tau}_l-1} \check{Q}_{l,j}\, \xi_j + \check{\varkappa}_l\, \check{Q}_{l,\check{\tau}_l}\, \xi_{\check{\tau}_l}}{H_l}\,. \tag{2.28}$$

Note that on the set Γ for any $1 \leq l \leq d$ the random variables $\eta_l = \xi_l^*$. Moreover (see Lemma A.2 in [4]), for any $1 \leq l \leq d$ and $p \in \mathcal{P}$

$$\mathbf{E}_{\mathbf{p},S}\left(\eta_l \,|\mathcal{G}_l\right) = 0\,, \quad \mathbf{E}_{\mathbf{p},S}\left(\eta_l^2 \,|\mathcal{G}_l\right) = \sigma_l^2 \quad \text{and} \quad \mathbf{E}_{\mathbf{p},S}\left(\eta_l^4 \,|\mathcal{G}_l\right) \leq \breve{\mathbf{m}}\sigma_l^4\,, \tag{2.29}$$

where $\sigma_l = H_l^{-1/2}$, $\mathcal{G}_l = \sigma\{\eta_1, \ldots, \eta_{l-1}, \sigma_l\}$ and $\breve{\mathbf{m}} = 4(144/\sqrt{3})^4\,\mathbf{m}_4^*$. It is clear that

$$\sigma_{0,*} \leq \min_{1\leq l\leq d} \sigma_l^2 \leq \max_{1\leq l\leq d} \sigma_l^2 \leq \sigma_{1,*}\,, \tag{2.30}$$

where

$$\sigma_{0,*} = \frac{1-\epsilon^2}{2(1-\widetilde{\epsilon})nh} \quad \text{and} \quad \sigma_{1,*} = \frac{1}{(1-\widetilde{\epsilon})(2nh-\mathbf{q}-3)}\,.$$

Now, taking into account that $|\varpi_{1,l}| \leq Lh$, for any $S \in \Theta_{\epsilon,L}$ we obtain that

$$\sup_{S\in\Theta_{\epsilon,L}} \mathbf{E}_{\mathbf{p},S}\mathbf{1}_\Gamma \varpi_l^2 \leq \left(L^2h^2 + \frac{\breve{\upsilon}_n}{(nh)^2}\right)\,, \tag{2.31}$$

where $\breve{\upsilon}_n = \sup_{\mathbf{p}\in\mathcal{P}} \sup_{S\in\Theta_{\epsilon,L}} \mathbf{E}_{\mathbf{p},S} \max_{1\leq j\leq n} y_j^4$. The behavior of this coefficient is studied in the following Proposition.

Proposition 2.3. *For any* $\mathbf{b} > 0$ *the sequence* $(\breve{\upsilon}_n)_{n\geq 1}$ *satisfies the following limiting equality*

$$\lim_{n\to\infty} n^{-\mathbf{b}}\,\breve{\upsilon}_n = 0\,. \tag{2.32}$$

Remark 2.1. *It should be noted that the property* (2.32) *means that the asymptotic behavior of the upper bound* (2.31) *approximately almost as* h^{-2} *when* $n \to \infty$. *We will use this in the oracle inequalities below.*

Remark 2.2. *Note, that to estimate the function* S *in* (1.1) *we use the approach developed in [11] for the diffusion processes. To this end we use the efficient sequential kernel procedures developed in [1, 2, 3]. It should be emphasized that to obtain an efficient estimator, i.e. an estimator with the minimal asymptotic risk, one needs to take only indicator kernel as in* (2.13).

Remark 2.3. *Ii should be noted also that the sequential estimator* (2.13) *has the same form as in [3], but except the last term, in which the correction coefficient is replaced by the square root of the coefficient used in [14]. We modify this procedure to calculate the variance of the stochastic term* (2.23).

3 Model selection

In this section we consider the nonparametric estimation problem in the non asymptotic setting for the regression model sec:In.1-11-1R for some set $\Gamma \subseteq \Omega$. The design points $(z_l)_{1\le l\le d}$ are defined in (2.5). The function $S(\cdot)$ is unknown and has to be estimated from observations $Y_1, \ldots, Y_d$. Moreover, we assume that the unobserved random variables $(\eta_l)_{1\le l\le d}$ satisfy the properties (2.29) with some nonrandom constant $\breve{\mathbf{m}} > 1$ and the known random positive coefficients $(\sigma_l)_{1\le l\le d}$ satisfy the inequlity (2.30) for some nonrandom positive constants $\sigma_{0,*}$ and $\sigma_{1,*}$ Concerning the random sequence $\varpi = (\varpi_l)_{1\le l\le d}$ we suppose that

$$\mathbf{E}_{\mathbf{p},S}\mathbf{1}_\Gamma \|\varpi\|_d^2 < \infty\,. \tag{3.1}$$

The performance of any estimator $\widehat{S}$ will be measured by the empirical squared error

$$\|\widehat{S} - S\|_d^2 = (\widehat{S} - S, \widehat{S} - S)_d = \frac{b-a}{d}\sum_{l=1}^{d}(\widehat{S}(z_l) - S(z_l))^2\,. \tag{3.2}$$

Now we fix a basis $(\phi_j)_{1\le j\le d}$ which is orthonormal for the empirical inner product:

$$(\phi_i\,,\,\phi_j)_d = \frac{b-a}{d}\sum_{l=1}^{d}\phi_i(z_l)\phi_j(z_l) = \mathbf{1}_{\{i=j\}}\,. \tag{3.3}$$

For example, we can take the trigonometric basis $(\phi_j)_{j\ge 1}$ in $\mathbf{L}_2[a,b]$, i.e.

$$\phi_1 = \frac{1}{\sqrt{b-a}}\,,\quad \phi_j(x) = \sqrt{\frac{2}{b-a}}\,\mathrm{Tr}_j\left(2\pi[j/2]\mathbf{l}_0(x)\right)\,,\quad j \ge 2\,, \tag{3.4}$$

where the function $\mathrm{Tr}_j(x) = \cos(x)$ for even j and $\mathrm{Tr}_j(x) = \sin(x)$ for odd j, $[x]$ denotes integer part of x. and $\mathbf{l}_0(x) = (x-a)/(b-a)$. Note that, in this case to obtain the property (3.3) the numbers of points d must be odd. To obtain the property (2.6) we can choose, for example, $d = 2[\sqrt{n}/2] + 1$, where $[a]$ is the integer part of $a \in \mathbb{R}$.

Note that, using the orthonormality property (3.3) we can represent for any $1 \le l \le d$ the function S as

$$S(z_l) = \sum_{j=1}^{d}\theta_{j,d}\,\phi_j(z_l) \quad \text{and} \quad \theta_{j,d} = \left(S, \phi_j\right)_d\,. \tag{3.5}$$

So, to estimate the function S we have to estimate the Fourrier coefficients $(\theta_{j,d})_{1\le j\le d}$. To this end we reply the the function S by the observations, i.e.

$$\widehat{\theta}_{j,d} = \frac{b-a}{d}\sum_{l=1}^{d} Y_l \phi_j(z_l)\,. \tag{3.6}$$

From sec:In.1-11-1R we obtain immediately the following regression shceme

$$\widehat{\theta}_{j,d} \;=\; \theta_{j,d} \,+\, \zeta_{j,d} \quad \text{with} \quad \zeta_{j,d} = \sqrt{\frac{b-a}{d}}\eta_{j,d} + \varpi_{j,d}\,, \tag{3.7}$$

where

$$\eta_{j,d} = \sqrt{\frac{b-a}{d}}\sum_{l=1}^{d} \eta_l \phi_j(z_l) \quad \text{and} \quad \varpi_{j,d} = \frac{b-a}{d}\sum_{l=1}^{d} \varpi_l\, \phi_j(z_l)\,.$$

Note that the upper bound (2.30) and the Bounyakovskii-Cauchy-Schwarz inequality imply that

$$|\varpi_{j,d}| \le \|\varpi\|_d\, \|\phi_j\|_d = \|\varpi\|_d\,.$$

Now we set

$$\mathbf{B}_n = n\|\varpi\|_d^2\,. \tag{3.8}$$

Note here that Proposition 3.3 from [4] implies directly that for any $\mathbf{b} > 0$

$$\lim_{n\to\infty} \frac{1}{n^{\mathbf{b}}} \sup_{\mathbf{p}\in\mathcal{P}} \sup_{S\in\Theta_{\varepsilon,L}} \mathbf{E}_{\mathbf{p},S}\mathbf{B}_n\, \mathbf{1}_\Gamma \;=\; 0\,. \tag{3.9}$$

We estimate the function S on the sieve (2.5) by the weighted least squares estimator

$$\widehat{S}_\lambda(z_l) \;=\; \sum_{j=1}^{d} \lambda(j)\, \widehat{\theta}_{j,d}\, \phi_j(z_l)\, \mathbf{1}_\Gamma\,, \quad 1 \le l \le d\,, \tag{3.10}$$

where the weight vector $\lambda = (\lambda(1), \ldots, \lambda(d))'$ belongs to some finite set $\Lambda \subset [0,1]^d$, the prime denotes the transposition. We set for any $a \le t \le b$

$$\widehat{S}_\lambda(t) = \sum_{l=1}^{d} \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1}<t\le z_l\}}\,. \tag{3.11}$$

Denote by ν be the cardinal number of the set Λ and

$$\Lambda_* = \max_{\lambda\in\Lambda} \sum_{j=1}^{d} \lambda(j)\,.$$

$\mathbf{A}_1$) *For any* $\check{\delta} > 0$

$$\lim_{n\to\infty} \frac{\phi_n^* + \nu_n}{n^{\check{\delta}}} = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{\Lambda_*(n)}{n^{1/6+\check{\delta}}} = 0\,, \tag{3.12}$$

where $\phi_n^* = \max_{1\le j\le n} \max_{\mathbf{x}_0\le x\le \mathbf{x}_1} |\phi_j(x)|$.
In order to obtain a good estimator, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (3.10). We define the empirical squared risk as

$$\mathrm{Err}_d(\lambda) = \|\widehat{S}_\lambda - S\|_d^2\,.$$

Using (3.5) and (3.10) we can rewire this risk as

$$\mathrm{Err}_d(\lambda) = \sum_{j=1}^d \lambda^2(j)\widehat{\theta}_{j,d}^2 - 2\sum_{j=1}^d \lambda(j)\widehat{\theta}_{j,d}\,\theta_{j,d} + \sum_{j=1}^d \theta_{j,d}^2\,. \tag{3.13}$$

Since the coefficient $\theta_{j,d}$ is unknown, we need to replace the term $\widehat{\theta}_{j,d}\,\theta_{j,d}$ by some its estimator which we choose as

$$\widetilde{\theta}_{j,d} = \widehat{\theta}_{j,d}^2 - \frac{b-a}{d}\,\mathbf{s}_{j,d} \quad \text{with} \quad \mathbf{s}_{j,d} = \frac{b-a}{d}\sum_{l=1}^d \sigma_l^2\,\phi_j^2(z_l)\,. \tag{3.14}$$

Note that from (2.30) - (3.3) it follows that

$$\mathbf{s}_{j,d} \le \sigma_{1,*}\,. \tag{3.15}$$

Finally, we define the cost function of the form

$$J_d(\lambda) \;=\; \sum_{j=1}^d \lambda^2(j)\widehat{\theta}_{j,d}^2 - 2\sum_{j=1}^d \lambda(j)\,\widetilde{\theta}_{j,d} \;+\; \delta P_d(\lambda)\,, \tag{3.16}$$

where the penalty term is defined as

$$P_d(\lambda) = \frac{b-a}{d}\sum_{j=1}^d \lambda^2(j)\mathbf{s}_{j,d} \tag{3.17}$$

and $0 < \delta < 1$ is some positive constant which will be chosen later. We set

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda}\, J_d(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,. \tag{3.18}$$

To study the efficiency property we specify the weight coefficients $(\lambda(j))_{1\le j\le n}$ as it is proposed, for example, in [10]. First, for some $0 < \varepsilon < 1$ introduce

the two dimensional grid to adapt to the unknown parameters (regularity and size) of the Sobolev bull, i.e. we set

$$\mathcal{A} = \{1, \dots, k^*\} \times \{\varepsilon, \dots, m\varepsilon\}\,, \tag{3.19}$$

where $m = [1/\varepsilon^2]$. We assume that both parameters $k^* \geq 1$ and ε are functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0\,, \\ \\ \lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\check{\delta}}\varepsilon(n) = +\infty \end{cases} \tag{3.20}$$

for any $\check{\delta} > 0$. One can take, for example, for $n \geq 2$

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k_0^* + \sqrt{\ln n}\,, \tag{3.21}$$

where $k_0^* \geq 0$ is some fixed constant. For each $\alpha = (\beta, \mathbf{l}) \in \mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha = (\lambda_\alpha(j))_{1\leq j\leq p}$$

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\leq j<j_*\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{j_*\leq j\leq\omega_\alpha\}}, \tag{3.22}$$

where $j_* = 1 + [\ln n]$, $\omega_\alpha = (\varpi_\beta \mathbf{l}\, n)^{1/(2\beta+1)}$,

$$\varpi_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} = \frac{2\beta}{\pi^{2\beta}\iota_\beta} \quad \text{and} \quad \iota_\beta = \frac{2\beta^2}{(\beta+1)(2\beta+1)}\,.$$

Now we define the set Λ as

$$\Lambda \;=\; \{\lambda_\alpha\,,\, \alpha \in \mathcal{A}\}\,. \tag{3.23}$$

Note, that these weight coefficients are used in [16, 17] for continuous time regression models to show the asymptotic efficiency. It will be noted that in this case the cardinal of the set Λ is $\nu = k^*m$. It is clear that the properties (3.20) imply the condition (3.12).
In [4] we shown the following result.

Theorem 3.1. *Assume that the conditions (2.2) and (3.12) hold. Then for any $n \geq 3$, any $S \in \Theta_{\epsilon,L}$ and any $0 < \delta \leq 1/12$, the procedure (3.18) with the coefficients (3.23) satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{(1+4\delta)(1+\delta)^2}{1-6\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{D}_n^*}{\delta n}\,, \tag{3.24}$$

where the term $\mathbf{D}_n^*$ *is such that for any* $\mathbf{b} > 0$

$$\lim_{n\to\infty} \frac{\mathbf{D}_n^*}{n^{\mathbf{b}}} = 0\,.$$

Remark 3.1. *It sjhould be noted that the weight least square estimators* (3.10) *with the weight coefficients* (3.23) *is efficient for the Sobolev ball (see, for example, [***?** *19]). So, Theorem 3.1 means that this model selection procedure is best among all the effective procedures in the sharp oracle inequality sense* (3.24) *and below we will use this property to show the efficiency property in adaptive setting, i.e. in the case, when the regularity property of the function* S (1.1) *is unknown.*

4 Main results

For any fixed $r > 0$ and $k \geq 1$ we define the Sobolev ellipse as

$$W_{k,r} = \{f \in \Theta_{\epsilon,L} : \sum_{j=0}^{k} \mathbf{a}_j\, \theta_j^2 \leq r\}\,, \tag{4.1}$$

where $\mathbf{a}_j = \sum_{l=0}^{k} (2\pi[j/2])^{2l}$, $(\theta_j)_{j\geq 1}$ are the trigonometric Fourier coefficients, i.e.

$$\theta_j = \int_a^b f(x)\phi_j(x)\mathrm{d}x$$

and $(\phi_j)_{j\geq 1}$ is the trigonometric basis defined in (3.4). It is clear we can represent this functional class as

$$W_{k,r} = \{f \in \Theta_{\epsilon,L} : \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq r\}\,, \tag{4.2}$$

In order to formulate the asymptotic results we define the following normalizing coefficients. First, for any $\mathbf{r} > 0$ we set

$$l_*(\mathbf{r}) = ((1+2k)r)^{1/(2k+1)} \left(\frac{k}{\pi(k+1)}\right)^{2k/(2k+1)} \tag{4.3}$$

and

$$\varsigma(S) = \int_a^b (1 - S^2(u))\mathrm{d}u\,. \tag{4.4}$$

It is well known that for any $S \in W_{k,r}$ the optimal rate of convergence is $n^{-2k/(2k+1)}$. First we study the lower bound for the asymptotic risks in the class of all estimators $\mathcal{E}_n$, i.e. any measurable function with respect to the observations $\sigma\{y_1, \dots, y_n\}$.

Theorem 4.1. *For the model (1.1) with the noise distribution from the class $\mathcal{P}$ defined in (2.1)*

$$\liminf_{n\to\infty} \inf_{\widehat{S}_n \in \mathcal{E}_n} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} \upsilon(S)\mathcal{R}^*(\widehat{S}_n, S) \geq l_*(\mathbf{r})\,, \tag{4.5}$$

where $\upsilon(S) = (\varsigma(S))^{-2k/(2k+1)}$.

Now we stady the asymptotic upper bound for the quadratic risk of the estimator $\widehat{S}_*$. To this end we assume the following condition for the penalty coefficient δ in the objective function (3.16).

$\mathbf{A}_2$) *Assume that the parameter δ is a function of n, i.e. $\delta = \delta_n$ such that for any $\mathbf{b} > 0$*

$$\lim_{n\to\infty} \frac{\delta_n}{n^{\mathbf{b}}} = 0\,. \tag{4.6}$$

Theorem 4.2. *Assume that the conditions $\mathbf{A}_1$) – $\mathbf{A}_2$) hold. The model selection procedure $\widehat{S}_*$ defined in (3.18) with the penalty coefficient given in $\mathbf{A}_2$) admits the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} \upsilon(S)\,\mathcal{R}(\widehat{S}_*, S) \leq l_*(\mathbf{r})\,. \tag{4.7}$$

Corollary 4.3. *Assume that the conditions $\mathbf{A}_1$) – $\mathbf{A}_2$) hold. The model selection procedure $\widehat{S}_*$ defined in (3.18) with the penalty coefficient given in $\mathbf{A}_2$) is efficient, i.e.*

$$\lim_{n\to\infty} \frac{\inf_{\widehat{S}_n\in\mathcal{E}_n} \sup_{S\in W_{k,r}} \upsilon(S)\,\mathcal{R}^*(\widehat{S}_n, S)}{\sup_{S\in W_{k,r}} \upsilon(S)\mathcal{R}(\widehat{S}_*, S)} = 1\,. \tag{4.8}$$

Moreover,

$$\lim_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} \upsilon(S)\,\mathcal{R}(\widehat{S}_*, S) = l_*(\mathbf{r})\,. \tag{4.9}$$

Remark 4.1. *Note that the limit equalties (4.8) and (4.9) imply that the function $l_*(\mathbf{r})/\upsilon(S)$ is the minimal value of the normalised asymptotic quadratic*

robust risk, i.e. Pinsker constant in this case. We remind that the coefficient $l_*(\mathbf{r})$ *is the well known Pinsker constant for the "signal+standard white noise" model obtained in [19]. Therefore, the Pinsker constant for the model* (1.1) *is represented by the Pinsker constant for the "signal+white noise" model in which the noise intensity is given by the function* (4.4).

5 The van Trees inequality

In this section we consider the following continuous time parametric model (1.1) with the $(0,1)$ gaussian i.i.d. random variable $(\xi_j)_{1\le j\le n}$ and the parametric linear function S, i.e.

$$S_\theta(x) = \sum_{l=1}^{d} \theta_l\, \psi_l(x)\,, \quad \theta = (\theta_1, \ldots, \Xi_d)' \in \mathbb{R}^n\,. \tag{5.1}$$

The functions $(\psi_i)_{1\le i\le d}$ are orthogonal with respect to the scalar product (3.3).

Let now $\mathbf{P}_\theta^n$ be the distribution in $\mathbb{R}^n$ of the observations $y = (y_1, \ldots, y_n)$ in the model (1.1) with the function (5.1) and ν_ξ^n be the distribution in $\mathbb{R}^n$ of the gaussian vector $(\xi_1, \ldots, \xi_n)$. In this case the Radon - Nykodim density is given as

$$f_n(y,\theta) = \frac{\mathrm{d}\mathbf{P}_\theta^{(n)}}{\mathrm{d}\nu_\xi^n} = \exp\left\{ \sum_{j=1}^{n} S_\theta(x_j) y_{j-1} y_j - \frac{1}{2}\sum_{j=1}^{n} S_\theta^2(x_j) y_{j-1}^2 \right\}. \tag{5.2}$$

Let $\mathbf{u}$ be a prior distribution density on $\mathbb{R}^d$ for the parameter θ of the following form:

$$\mathbf{u}(\theta) = \prod_{j=1}^{d} \mathbf{u}_j(\theta_j)\,,$$

where $\mathbf{u}_j$ is some continuously differentiable probability density in $\mathbb{R}$ with the support $]-L_j\,,\, L_j[$, i.e. $\mathbf{u}_j(z) > 0$ for any $-L_j < z < L_j$ and $\mathbf{u}_j(z) = 0$ for all $|z| \ge L_j$, such that the Fisher information is finite, i.e.

$$I_j = \int_{-L_j}^{L_j} \frac{\dot{\mathbf{u}}_l^2(z)}{\mathbf{u}_j(z)}\,\mathrm{d}z < \infty\,. \tag{5.3}$$

Now, we set

$$\Xi =]-L_1\,,\, L_1[\times \ldots \times]-L_d\,,\, L_d[\quad \subseteq \quad \mathbb{R}^d\,. \tag{5.4}$$

Let $g(\theta)$ be a continuously differentiable $\Xi \to \mathbb{R}$ function such that, for each $1 \le j \le d$,

$$\lim_{|\theta_j| \to L_j} g(\theta)\, \mathbf{u}_j(\theta_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |g'_j(\theta)|\, \mathbf{u}(\theta)\, \mathrm{d}\theta < \infty\,, \tag{5.5}$$

where $g'_j(\theta) = \partial g(\theta)/\partial\theta_j$.
For any $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathbb{R}^d)$-measurable integrable function $H = H(y,\theta)$ we denote

$$\begin{aligned}\widetilde{\mathbf{E}}\, H &= \int_{\Xi} \left(\int_{\mathbb{R}^n} H(y,\theta)\, \mathrm{d}\mathbf{P}_\theta^n \right) \mathbf{u}(\theta)\mathrm{d}\theta \\ &= \int_{\Xi} \left(\int_{\mathbb{R}^n} H(y,\theta)\, f_n(y,\theta)\, \mathbf{u}(\theta)\mathrm{d}\nu_\xi^{(n)} \right) \mathrm{d}\theta\,.\end{aligned}$$

Now we obtain an lower bound for the corresponding bayesian risks in the case when the model (1.1) is gaussian with the function (5.1).

Lemma 5.1. *For any $\mathcal{F}_n^y$-measurable square integrable function $\widehat{g}_n$ and for any $1 \le j \le d$, the following inequality holds*

$$\widetilde{\mathbf{E}}(\widehat{g}_n - g(\theta))^2 \ge \frac{\overline{g}_j^2}{\widetilde{\mathbf{E}}\, \overline{\Psi}_{n,j} + I_j}\,, \tag{5.6}$$

where $\overline{\Psi}_{n,j} = \sum_{j=1}^n \psi_j^2(x_l)\, y_{l-1}^2$ and $\overline{g}_j = \int_\Xi g'_j(\theta)\, \mathbf{u}(\theta)\, \mathrm{d}\theta$.

Proof. First, for any $\theta \in \Xi$ we set

$$\widetilde{U}_j = \widetilde{U}_j(y,\theta) = \frac{1}{f(y,\theta)\mathbf{u}(\theta)}\, \frac{\partial\left(f(y,\theta)\mathbf{u}(\theta)\right)}{\partial\theta_j}\,.$$

Taking into account the condition (5.5) and integrating by parts we get

$$\begin{aligned}\widetilde{\mathbf{E}}\left((\widehat{g}_n - g(\theta))\widetilde{U}_j\right) &= \int_{\mathbb{R}^n \times \Xi} (\widehat{g}_n(y) - g(\theta)) \frac{\partial}{\partial\theta_j} \left(f(y,\theta)\mathbf{u}(\theta)\right) \mathrm{d}\theta\, \mathrm{d}\nu_\xi^{(n)} \\ &= \int_{\mathbb{R}^n \times \check{\Xi}_j} \left(\int_{-L_j}^{L_j} g'_j(\theta)\, f(y,\theta)\mathbf{u}(\theta)\mathrm{d}\theta_j \right) \left(\prod_{i \ne j} \mathrm{d}\theta_i \right) \mathrm{d}\nu_\xi^{(n)} = \overline{g}_j\,,\end{aligned}$$

where

$$\check{\Xi}_j = \prod_{i \ne j} \,]-L_i\,,\, L_i[\,.$$

Now by the Bouniakovskii-Cauchy-Schwarz inequality we obtain the following lower bound for the quadratic risk

$$\widetilde{\mathbf{E}}(\widehat{g}_n - g(\theta))^2 \geq \frac{\overline{g}_j^2}{\widetilde{\mathbf{E}}\widetilde{U}_j^2}\,.$$

To study the denominator in the left hand of this inequality note that in view of the representation (5.2)

$$\frac{1}{f_n(y,\theta)}\frac{\partial\, f_n(y,\theta)}{\partial\theta_j} = \sum_{l=1}^{n} \psi_j(x_l)\, y_{l-1}(y_l - S_\theta(x_l)y_{l-1})\,.$$

Therefore, for each $\theta \in \Xi$,

$$\mathbf{E}_\theta^{(n)}\, \frac{1}{f_n(y,\theta)}\frac{\partial\, f_n(y,\theta)}{\partial\theta_j} = 0$$

and

$$\mathbf{E}_\theta^{(n)} \left(\frac{1}{f_n(y,\theta)}\frac{\partial\, f_n(y,\theta)}{\partial\theta_j}\right)^2 = \mathbf{E}_\theta^{(n)} \sum_{l=1}^{n} \psi_j^2(x_l)\, y_{l-1}^2 = \mathbf{E}_\theta^{(n)}\, \overline{\Psi}_{n,l}\,.$$

Using the equality

$$\widetilde{U}_j = \frac{1}{f_n(y,\theta)}\frac{\partial\, f_n(y,\theta)}{\partial\theta_j} + \frac{1}{\mathbf{u}(\theta)}\frac{\partial\, \mathbf{u}(\theta)}{\partial\theta_j}\,,$$

we get

$$\widetilde{\mathbf{E}}\, \widetilde{U}_j^2 = \widetilde{\mathbf{E}}\, \overline{\Psi}_{n,j} + I_j\,,$$

where the Fisher information I_j is defined in (5.3). Hence Lemma 5.1. □

Remark 5.1. *It should be noted that in the definition of the prior distribution the bound L_j may be equal to infinity either for some $1 \leq j \leq d$ or for all $1 \leq j \leq d$.*

6 Low bound

First, note that

$$\mathcal{R}^*(\widehat{S}_n, S) \;\geq\; \mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S)\,, \tag{6.1}$$

where $\mathbf{p}_0$ is the $(0,1)$ gaussian density. Now for any fixed $0 < \varepsilon < 1$ we set

$$d = d_n = \left[\frac{k+1}{k} (\varsigma n)^{1/(2k+1)} l_*(\mathbf{r}_\varepsilon) \right] \tag{6.2}$$

where $\varsigma = 1/(b-a)$, $\mathbf{r}_\varepsilon = (1-\varepsilon)\mathbf{r}$ and

$$l_*(\mathbf{r}_\varepsilon) = ((1+2k)\mathbf{r}_\varepsilon)^{1/(2k+1)} \left(\frac{k}{\pi(k+1)} \right)^{2k/(2k+1)} = (1-\varepsilon)^{1/(2k+1)} l_*(\mathbf{r}) \,.$$

For any vector $\theta = (\theta_j)_{1 \le j \le d} \in \mathbb{R}^d$, we set

$$S_\theta(x) = \sum_{j=1}^{d_n} \theta_j \, \phi_j(x) \,, \tag{6.3}$$

where $(\phi_j)_{1 \le j \le d_n}$ is the trigonometric basis defined in (3.4). As ir is shown in [13] there exist continuously differentiable density $\check{\mathbf{p}}_L$ with the support on $[-L, L]]$ such that $\int_{-L}^{L} x \check{\mathbf{p}}_L(x) \mathrm{d}x = 0$, $\int_{-L}^{L} x^2 \check{\mathbf{p}}_L(x) \mathrm{d}x = 1$ and

$$\check{I}_L = \int_{-L}^{L} \frac{(\check{\mathbf{p}}'(x))^2}{\check{\mathbf{p}}(x)} \, \check{\mathbf{p}}(x) \mathrm{d}x = 1 + \check{\epsilon}_L \,,$$

where $\check{\epsilon}_L \to 0$ as $L \to \infty$. To define the bayesian risk we choose a prior distribution on $\mathbb{R}^d$ as

$$\theta = (\theta_j)_{1 \le j \le d_n} \quad \text{and} \quad \kappa_j = s_j \, \check{\eta}_j \,, \tag{6.4}$$

where $\check{\eta}_j$ are i.i.d. random variables with the density $\check{\mathbf{p}}_L$,

$$s_j = \sqrt{\frac{s_j^*}{\varsigma n}} \quad \text{and} \quad s_j^* = \left(\frac{d_n}{j} \right)^k - 1 \,.$$

Furthermore, for any function f, we denote by $\mathbf{h}(f)$ its projection in $\mathbf{L}_2[0,1]$ onto $W_{k,r}$, i.e.

$$\mathbf{h}(f) = \mathrm{Pr}_{W_{k,r}}(f) \,.$$

Since $W_{k,r}$ is a convex set, we obtain, that for any function $S \in W_{k,r}$

$$\|\widehat{S} - S\|^2 \ge \|\widehat{\mathbf{h}} - S\|^2 \quad \text{with} \quad \widehat{\mathbf{h}} = \mathbf{h}(\widehat{S}) \,.$$

From the definition of the prio distribution (6.4) we obtain that a.s.

$$\max_{a \le x \le b} \left(|S_\theta(x)| + |\dot{S}_\theta(x)| \right) \le \sqrt{\frac{2}{b-a}} \, \frac{b-a+1}{b-a} \epsilon_n := \epsilon_n^* \,, \tag{6.5}$$

where

$$\epsilon_n = \frac{L}{\sqrt{n}} \sum_{j=1}^{d_n} \left(\frac{d_n}{j}\right)^{k/2} j \to \quad 0 \quad \text{as} \quad n \to \infty$$

for any $k \geq 2$. Therefore, for sufficiently large n the function (6.3) belongs to the class (2.4) and the last property yileds

$$\sup_{S \in W_{k,r}} \upsilon(S)\, \mathcal{R}_{\mathbf{p}_0}(\widehat{S}, S) \;\geq\; \int_{\{z \in \mathbb{R}^d \,:\, S_z \in W_{k,r}\}} \upsilon(S_z) \mathbf{E}_{\mathbf{p}_0, S_z} \|\widehat{\mathbf{h}} - S_z\|^2 \, \mu_\kappa(\mathrm{d}z)$$

$$\geq\; \upsilon_* \int_{\{z \in \mathbb{R}^d \,:\, S_z \in W_{k,r}\}} \mathbf{E}_{\mathbf{p}_0, S_z} \|\widehat{\mathbf{h}} - S_z\|^2 \, \mu_\kappa(\mathrm{d}z)\,,$$

where

$$\upsilon_* = \inf_{|S|_* \leq \epsilon_n^*} \upsilon(S) \quad \to \quad \varsigma^{2k/(2k+1)} \quad \text{as} \quad n \to \infty\,.$$

Using the distribution μ_κ we introduce the following Bayes risk

$$\widetilde{\mathcal{R}}_0(\widehat{S}) = \int_{\mathbb{R}^d} \mathcal{R}_{\mathbf{p}_0}(\widehat{S}, S_z) \, \mu_\kappa(\mathrm{d}z)\,.$$

Taking into account now that $\|\widehat{\mathbf{h}}\|^2 \leq r$ we obtain

$$\sup_{S \in W_{k,r}} \upsilon(S)\, \mathcal{R}_{\mathbf{p}_0}(\widehat{S}, S) \;\geq\; \upsilon_* \, \widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) - 2\, \upsilon_* \, \mathbf{R}_{0,n} \tag{6.6}$$

with

$$\mathbf{R}_{0,n} = \int_{\{z \in \mathbb{R}^d \,:\, S_z \notin W_{k,r}\}} (r + \|S_z\|^2) \, \mu_\kappa(\mathrm{d}z)\,.$$

In Lemma A.2 we studied the last term in this inequality. Now it is easy to see that

$$\|\widehat{\mathbf{h}} - S_z\|^2 \geq \sum_{j=1}^{d_n} (\widehat{z}_j - z_j)^2\,,$$

where $\widehat{z}_j = \int_0^1 \widehat{\mathbf{h}}(t)\, \phi_j(t) \mathrm{d}t$. So, in view of Lemma 5.1, we obtain

$$\widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) \;\geq\; \frac{1}{\varsigma n} \sum_{j=1}^{d_n} \frac{1}{1 + \check{I}_L(s_j^*)^{-1}} \geq \frac{1}{\varsigma n \max(1, \check{I}_L)} \sum_{j=1}^{d_n} \left(1 - \frac{j^k}{d_n^k}\right)\,.$$

Therefore, using now the definition (6.2), Lemma A.2 and the inequality (6.1) we obtain that

$$\liminf_{n \to \infty} \inf_{\widehat{S} \in \Pi_n} n^{\frac{2k}{2k+1}} \sup_{S \in W_{k,r}} \upsilon(S)\, \mathcal{R}^*(\widehat{S}_n, S) \geq (1 - \varepsilon)^{\frac{1}{2k+1}} \frac{1}{\max(1, \check{I}_L)}\, l_*\,.$$

Taking here limit as $\varepsilon \to 0$ and $L \to \infty$ we come to the Theorem 4.1. $\square$

7 Upper bound

7.1 Known regularity

We start with the estimation problem for the functions S from $W_{k,r}$ with known parameters k, r and $\varsigma(S)$ defined in (4.4). In this case we use the estimator from family (3.23)

$$\widetilde{S} = \widehat{S}_{\widetilde{\alpha}}\,, \tag{7.1}$$

where $\widetilde{\alpha} = (k, \widetilde{t}_n)$, $\widetilde{\mathrm{l}}_n = [\overline{r}(S)/\varepsilon]\,\varepsilon$ and $\overline{r}(S) = r/\varsigma(S)$. We remind, that $\varepsilon = 1/\ln n$. Note that for sufficiently large n, the parameter $\widetilde{\alpha}$ belongs to the set (3.19). In this section we obtain the upper bound for the empiric risk (3.2).

Theorem 7.1. *The estimator $\widetilde{S}$ constructed on the trigonometric basis satisfies the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} \upsilon(S)\,\mathbf{E}_{\mathbf{p},S}\|\widetilde{S} - S\|_d^2\,\mathbf{1}_\Gamma \le l_*(\mathbf{r})\,. \tag{7.2}$$

Proof. We denote $\widetilde{\lambda} = \lambda_{\widetilde{\alpha}}$ and $\widetilde{\omega} = \omega_{\widetilde{\alpha}}$. Now we recall that the Fourier coefficients on the set Γ

$$\widehat{\theta}_{j,d} = \theta_{j,d} + \zeta_{j,d} \quad \text{with} \quad \zeta_{j,d} = \sqrt{\frac{b-a}{d}}\eta_{j,d} + \varpi_{j,d}\,,$$

Therefore, on the set Γ we can represent the empiric squared error as

$$\begin{aligned}\|\widetilde{S} - S\|_d^2 = &\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\,\theta_{j,d}^2 - 2M_n \\ &- 2\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))\,\widetilde{\lambda}(j)\theta_{j,d}\,\varpi_{j,d} + \sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\zeta_{j,d}^2\,,\end{aligned}$$

where

$$M_n = \sqrt{\frac{b-a}{d}}\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))\,\widetilde{\lambda}(j)\theta_{j,d}\,\eta_{j,d}\,.$$

Now for any $0 < \check{\varepsilon} < 1$

$$2\left|\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))\widetilde{\lambda}(j)\theta_{j,d}\,\varpi_{j,d}\right| \le \check{\varepsilon}\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 + \check{\varepsilon}^{-1}\sum_{j=1}^{d}\varpi_{j,d}^2\,.$$

Taking into account here the definition (3.8), we can rewrite this inequality as

$$2\left|\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))\widetilde{\lambda}(j)\theta_{j,d}\,\varpi_{j,d}\right| \le \check{\varepsilon}\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 + \frac{\mathbf{B}_n}{\check{\varepsilon}n}\,.$$

Therefore,

$$\|\widetilde{S}-S\|_d^2 \le (1+\check{\varepsilon})\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 - 2M_n + \frac{\mathbf{B}_n}{\check{\varepsilon}n} + \sum_{j=1}^{d}\widetilde{\lambda}^2(j)\zeta_{j,d}^2\,.$$

By the same way we estimate the last term on the right-hand side of this inequality as

$$\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\zeta_{j,d}^2 \le \frac{(1+\check{\varepsilon})(b-a)}{d}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\eta_{j,d}^2 + (1+\check{\varepsilon}^{-1})\frac{\mathbf{B}_n}{n}\,.$$

Thus, on the set Γ we find that for any $0<\check{\varepsilon}<1$

$$\|\widetilde{S}_n-S\|_d^2 \le (1+\check{\varepsilon})\Upsilon_n(S) - 2M_n + (1+\check{\varepsilon})\mathbf{U}_n + \frac{3\mathbf{B}_n}{\check{\varepsilon}n}\,, \tag{7.3}$$

where

$$\Upsilon_n(S) = \sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 + \frac{\varsigma(S)}{d^2}\sum_{j=1}^{d}\widetilde{\lambda}^2(j) \tag{7.4}$$

and

$$\mathbf{U}_n = \frac{1}{d^2}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\left(d(b-a)\eta_{j,d}^2 - \varsigma(S)\right)\,.$$

First, note that in view of Lemma A.7

$$\mathbf{E}_{\mathbf{p},S}\,M_n^2 \le \frac{\sigma_{1,*}(b-a)}{d}\sum_{j=1}^{d}\theta_{j,d}^2 = \frac{\sigma_{1,*}(b-a)}{d}\|S\|_d^2 \le \frac{\sigma_{1,*}(b-a)^2}{d}\,,$$

where the constant $\sigma_{1,*}$ is defined in (2.30). Moreover, taking into account here that $\mathbf{E}_{\mathbf{p},S}\,M_n=0$, we get

$$|\mathbf{E}_{\mathbf{p},S}\,M_n\,\mathbf{1}_\Gamma| \;=\; |\mathbf{E}_{\mathbf{p},S}\,M_n\,\mathbf{1}_{\Gamma^c}| \le (b-a)\sqrt{\frac{\sigma_{1,*}\mathbf{P}_{\mathbf{p},S}(\Gamma^c)}{d}}\,.$$

Therefore, Proposition 2.2 yields

$$\lim_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in\Theta_{\epsilon,L}} |\mathbf{E}_{\mathbf{p},S}\,M_n\,\mathbf{1}_\Gamma| = 0\,. \tag{7.5}$$

Now, the property (3.9) Lemma A.6 and Lemma A.8 imply the inequality (7.2). Hence Theorem 7.1. $\square$

7.2 Unknown smoothness

Theorem 3.1 and Theorem 7.1 imply Theorem 4.2. □

Acknowledgements. The last author was partially supported by by the Russian Federal Professor program (project no. 1.472.2016/1.4, the Ministry of Education and Science of the Russian Federation), the research project no. 2.3208.2017/4.6 (the Ministry of Education and Science of the Russian Federation), by RFBR Grant 16-01-00121 A and by "The Tomsk State University competitiveness improvement program" grant 8.1.18.2018.

A Appendix

A.1 Properties of the prior distribution (6.4)

Lemma A.1. *For any* $1 \le j \le d$

$$\lim_{n\to\infty} \frac{(b-a)}{n}\, \widetilde{\mathbf{E}}\, \overline{\Psi}_{n,j} = 1\,. \tag{A.1}$$

Proof. First, note that for any $k \ge 1$

$$y_k = y_0 \prod_{j=1}^{k} S_\theta(x_j) + \sum_{l=1}^{k} \prod_{j=l+1}^{k} S_\theta(x_j)\xi_l\,.$$

Using the distribution (6.4), we obtain that

$$\widetilde{\mathbf{E}}\, y_m^2 = y_0^2\, \widetilde{\mathbf{E}} \prod_{j=1}^{m} S_\theta^2(x_j) + \sum_{l=1}^{m} \widetilde{\mathbf{E}} \prod_{j=l+1}^{m} S_\theta^2(x_j)\,.$$

Therefore, due to the property (6.5) we obtain that for any $m \ge 1$ and for any $n \ge 1$ for which $\epsilon_n^* < 1$ we get, that

$$|\widetilde{\mathbf{E}}\, y_m^2 - 1| \le y_0^2\, (\epsilon_n^*)^m + \sum_{l=1}^{m-1} (\epsilon_n^*)^{m-l} \le \epsilon_n^* \frac{\left(y_0^2+1\right)}{1-\epsilon_n^*}\,.$$

Taking into account that for any $n \ge 1$

$$\frac{(b-a)}{n} \sum_{l=1}^{n} \phi^2(x_l) = 1\,,$$

we obtain that

$$\lim_{n\to\infty} \sup_{m\geq 1} |\widetilde{\mathbf{E}}\, y_m^2 - 1| = 0\,.$$

Hence Lemma A.1. $\square$

Lemma A.2. *For any* $\mathbf{b} > 0$ *the term* $\mathbf{R}_{0,n}$ *introduced in* (6.6) *satisfies the following property*

$$\lim_{n\to\infty} n^{\mathbf{b}}\, \mathbf{R}_{0,n} = 0\,. \tag{A.2}$$

Proof. First note, that the bound (6.5) implies that for sufficiently large n the function (6.3) with the random coefficients defined through the prior distribution (6.4) almost sure belongs to the class $\Theta_{\varepsilon,L}$. Therefore, the definition (4.1) we obtain that

$$\{S_\theta \notin W_{k,r}\} = \{\zeta_n > \mathbf{r}\}\ ,$$

where $\zeta_n = \sum_{j=1}^{d_n} \kappa_j^2\, \mathbf{a}_j$. So, it suffices to show that for any $\mathbf{b} > 0$

$$\lim_{n\to\infty} n^{\mathbf{b}}\, \mathbf{P}(\zeta_n > \mathbf{r}) = 0\,. \tag{A.3}$$

Note now, that

$$\lim_{n\to\infty} \mathbf{E}\,\zeta_n = \lim_{n\to\infty} \frac{1}{(b-a)n} \sum_{j=1}^{d_n} \mathbf{a}_j\, s_j^* = \mathbf{r}_\varepsilon = (1-\varepsilon)\mathbf{r}\,.$$

So, for sufficiently large n we obtain that

$$\{\zeta_n > \mathbf{r}\} \subset \left\{\widetilde{\zeta}_n > \mathbf{r}_1\right\}\ ,$$

where $\mathbf{r}_1 = \mathbf{r}\varepsilon/2$ and

$$\widetilde{\zeta}_n = \zeta_n - \mathbf{E}\,\zeta_n = \frac{1}{n} \sum_{j=1}^{d_n} s_j^*\, \mathbf{a}_j \widetilde{\eta}_j\,.$$

Using again the correlation inequality from [12] we get that for any $p \geq 2$ there exists some constant $C_p > 0$ for which

$$\mathbf{E}\,\widetilde{\zeta}_n^p \leq C_p \frac{1}{v_n^p} \left(\sum_{j=1}^{d} (s_j^*)^2\, \mathbf{a}_j^2 \right)^{p/2} \leq C_p\, n^{-\frac{p}{4k+2}}\,,$$

i.e. the expectation $\mathbf{E}\,\widetilde{\zeta}_n^p \to 0$ as $n \to \infty$. Therefore, using the Chebychev inequality we obtain that for any $\mathbf{b} > 0$

$$n^{\mathbf{b}} \mathbf{P}(\widetilde{\zeta}_n > \mathbf{r}_1) \to 0 \quad \text{as} \quad n \to \infty\,.$$

This implies (A.3), and therefore, Lemma A.2. $\square$

A.2 Relations between the norms $\|\cdot\|$ and $\|\cdot\|_d$.

Lemma A.3. *Let f be an absolutely continuous $[0,1] \to \mathbb{R}$ function with $\|\dot{f}\| < \infty$ and g be a simple $[0,1] \to \mathbb{R}$ function of the form*

$$g(t) = \sum_{j=1}^{d} c_j \, \chi_{(t_{j-1}, t_j]}(t),$$

where c_j are some constants and $t_j = j/d$. Then for any $\widetilde{\varepsilon} > 0$

$$\|f - g\|^2 \le (1 + \widetilde{\varepsilon})\|f - g\|_d^2 + (1 + \widetilde{\varepsilon}^{-1})\frac{\|\dot{f}\|^2}{d^2}$$

and

$$\|f - g\|_d^2 \le (1 + \widetilde{\varepsilon})\|f - g\|^2 + (1 + \widetilde{\varepsilon}^{-1})\frac{\|\dot{f}\|^2}{d^2}\,.$$

Proof. Setting $\Delta(t) = f(t) - g(t)$, we obtain, that for any $\widetilde{\varepsilon} > 0$

$$\|\Delta\|^2 = \|\Delta\|_d^2 + \sum_{l=1}^{d} \int_{t_{l-1}}^{t_l} \left(2\Delta(t_l)\,(\Delta(t) - \Delta(t_l)) + (\Delta(t) - \Delta(t_l))^2\right) \mathrm{d}t$$

$$\le (1 + \widetilde{\varepsilon})\|\Delta\|_d^2 + (1 + \widetilde{\varepsilon}^{-1}) \sum_{l=1}^{d} \int_{t_{l-1}}^{t_l} \left[\Delta(t_l) - (\Delta(t))\right]^2 \mathrm{d}t$$

$$= (1 + \widetilde{\varepsilon})\|\Delta\|_d^2 + (1 + \widetilde{\varepsilon}^{-1}) \sum_{l=1}^{d} \int_{t_{l-1}}^{t_l} |f(t_l) - f(t)|^2 \, \mathrm{d}t\,.$$

Noting that, for $t_{l-1} < t \le t_l$, one has the estimate

$$|f(t_l) - f(t)|^2 \le \left(\int_{t_{l-1}}^{t_l} |\dot{f}(u)| \mathrm{d}u\right)^2 \le \frac{1}{p} \int_{t_{l-1}}^{t_l} |\dot{f}(u)|^2 \mathrm{d}u\,,$$

one comes to the first inequality. Similarly, one can verify the second inequality. Hence Lemma A.3. □

A.3 Properties of the trigonometric basis.

Lemma A.4. *For any $1 \le j \le d$ the trigonometric Fourier coefficients $(\theta_{j,d})_{1 \le j \le p}$ for the functions S from the class $W_{k,r}$ satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality*

$$\theta_{j,d}^2 \;\le\; (1 + \widetilde{\varepsilon})\,\theta_j^2 \;+ (1 + \widetilde{\varepsilon}^{-1})\,\frac{2r}{d^{2k}}\,. \tag{A.4}$$

Proof. First we represent the function S as

$$S(x) = \sum_{l=1}^{d} \theta_l \, \phi_l(x) + \Delta_d(x) \,,$$

where

$$\Delta_d(x) = \sum_{l>d} \theta_l \, \phi_l(x) \,.$$

Therefore,

$$\theta_{j,d} = (S, \phi_j)_d = \theta_j + (\Delta_d, \phi_j)_d$$

and for any $0 < \widetilde{\varepsilon} < 1$

$$\theta_{j,d}^2 \le (1 + \widetilde{\varepsilon})\theta_j^2 + (1 + \widetilde{\varepsilon}^{-1})\|\Delta_d\|_d^2 \,.$$

By applying Lemma A.3 with $g = 0$, we obtain that

$$\|\Delta_d\|_d^2 \le 2\|\Delta_d\|^2 + 2\frac{\|\dot{\Delta}_d\|^2}{d^2} \,.$$

Note here, that for any $N \ge 1$

$$\|\dot{\Delta}_N\|^2 = (2\pi)^2 \sum_{l>N} \theta_l^2 \, [l/2]^2 \,.$$

Taking into account here that

$$2\pi[l/2] \ge l \quad \text{for} \quad l \ge 2 \,,$$

we obtain that

$$\|\dot{\Delta}_N\|^2 \le \sum_{l>N} \frac{\mathbf{a}_l}{l^{2(k-1)}} \theta_l^2 \le \frac{r}{N^{2(k-1)}} \,. \tag{A.5}$$

Hence Lemma A.4 □

Lemma A.5. *For any $d \ge 2$, $1 \le N \le d$ and $r > 0$, the coefficients $(\theta_{j,d})_{1\le j\le d}$ of functions S from the class W_r^1 satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality*

$$\sum_{j=N}^{d} \theta_{j,d}^2 \le (1 + \widetilde{\varepsilon}) \sum_{j\ge N} \theta_j^2 + (1 + \widetilde{\varepsilon}^{-1}) \frac{r}{d^2 N^{2(k-1)}} \,. \tag{A.6}$$

Proof. First we note that

$$\sum_{j=N}^{d} \theta_{j,d}^2 = \min_{x_1,\dots,x_{N-1}} \|S - \sum_{j=1}^{N-1} x_j \phi_j\|_d^2 \le \|\Delta_N\|_d^2 \,,$$

where $\Delta_N(t) = \sum_{j\ge N} \theta_j \phi_j(t)$. By applying Lemma A.3 and taking into account the inequality (A.5), we obtain the bound (A.6). Hence Lemma A.5 □

A.4 Technical lemmas

Lemma A.6. *The sequence* $\Upsilon_n(S)$ *satisfies the following upper bound*

$$\limsup_{n\to\infty} \sup_{S\in W_{k,r}} n^{2k/(2k+1)}\, \upsilon(S)\, \Upsilon_n(S) \,\leq\, l_*(\mathbf{r})\,. \tag{A.7}$$

Proof. First of all, note that

$$0 < \varepsilon^2(b-a) \leq \inf_{S\in\Theta_{\varepsilon,L}} \varsigma(S) \leq \sup_{S\in\Theta_{\varepsilon,L}} \varsigma(S) \leq b-a\,. \tag{A.8}$$

This implies directly that

$$\lim_{n\to\infty} \sup_{S\in\Theta_{\varepsilon,L}} \left| \frac{\widetilde{l}_n}{\overline{r}(S)} - 1 \right| = 0\,, \tag{A.9}$$

where $\overline{r}(S) = r/\varsigma(S)$. Moreover, note that

$$n^{2k/(2k+1)}\upsilon(S)\,\Upsilon_n(S) \leq n^{2k/(2k+1)}\,\upsilon(S)\,\Xi_d + \frac{(\varsigma(S))^{1/(2k+1)}}{n^{1/(2k+1)}} \sum_{j=1}^{d} \widetilde{\lambda}^2(j)$$

and

$$\Xi_d = \sum_{j=1}^{d} (1-\widetilde{\lambda}(j))^2\,\theta_{j,d}^2 = \Xi_{1,d} + \Xi_{2,d}\,,$$

where

$$\Xi_{1,d} = \sum_{j=j_*}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\,\theta_{j,d}^2 \quad \text{and} \quad \Xi_{2,d} = \sum_{j=[\widetilde{\omega}]+1}^{d} \theta_{j,d}^2\,.$$

We recall that

$$\widetilde{\omega} = \omega_{\widetilde{\alpha}} = \left(n\widetilde{\mathbf{l}}_n\varpi_k\right)^{1/(2k+1)}\,.$$

Lemma A.4 and Lemma A.5 yield

$$\Xi_{1,d} \leq (1+\widetilde{\varepsilon}) \sum_{j=j_*}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\,\theta_j^2 + 2\mathbf{r}(1+\widetilde{\varepsilon}^{-1})\,\frac{\widetilde{\omega}}{d^{2k}}\,.$$

and

$$\Xi_{2,d} \leq (1+\widetilde{\varepsilon}) \sum_{j\geq[\widetilde{\omega}]+1} \theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,\frac{\mathbf{r}}{d^2\,\widetilde{\omega}^{2(k-1)}}\,,$$

i.e.

$$\Xi_d \leq (1+\widetilde{\varepsilon}) \sum_{j\geq1} (1-\widetilde{\lambda}(j))^2\,\theta_j^2 + 2\mathbf{r}(1+\widetilde{\varepsilon}^{-1})\,\widetilde{\Xi}_n\,, \tag{A.10}$$

where

$$\Xi_n^* = \sum_{j\geq 1}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 = \sum_{j\leq\widetilde{\omega}}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 + \sum_{j>\widetilde{\omega}}\theta_j^2 := \Xi_{1,n}^* + \Xi_{2,n}^*$$

and

$$\widetilde{\Xi}_n = \frac{\widetilde{\omega}}{d^{2k}} + \frac{1}{d^2\widetilde{\omega}^{2(k-1)}}\,.$$

Note, that

$$n^{2k/(2k+1)}\upsilon(S)\Xi_{1,n}^* = \sum_{j=j_*}^{[\widetilde{\omega}]}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 = \frac{\upsilon(S)}{(\varpi_k\widetilde{\mathbf{l}}_n)^{2k/(2k+1)}}\sum_{j=j_*}^{[\widetilde{\omega}]} j^2\,\theta_j^2$$

$$\leq \frac{\upsilon(S)}{(\varpi_k\widetilde{\mathbf{l}}_n)^{2k/(2k+1)}}\,\mathbf{a}_{j_*}^*\sum_{j=j_*}^{[\widetilde{\omega}]}\mathbf{a}_j\,\theta_j^2\,,$$

where $\mathbf{a}_n^* = \sup_{j\geq n} j^{2k}/\mathbf{a}_j$. It is clear that

$$\lim_{n\to\infty}\mathbf{a}_n^* = \frac{1}{\pi^{2k}}\,.$$

Therefore, from (A.9) we obtain that

$$\limsup_{n\to\infty}\sup_{S\in\Theta_{\varepsilon,L}}\frac{n^{2k/(2k+1)}\upsilon(S)\,\Xi_{1,n}^*}{\sum_{j=j_*}^{[\widetilde{\omega}]}\mathbf{a}_j\,\theta_j^2}$$

$$\leq \frac{1}{\pi^{2k}(\varpi_k r)^{2k/(2k+1)}} = \frac{\iota_k^{2k/(2k+1)}}{(2k\pi\,r)^{2k/(2k+1)}}\,.$$

Next note, that for any $0<\widetilde{\varepsilon}<1$ and for sufficiently large n

$$\Xi_{2,n}^* \leq \frac{1}{\pi^{2k}\,\widetilde{\omega}^{2k}}\sum_{j\geq[\widetilde{\omega}]+1}\mathbf{a}_j\,\theta_j^2$$

$$\leq (1+\widetilde{\varepsilon})\frac{(\varsigma(S)\iota_k)^{2k/(2k+1)}}{\pi^{2k/(2k+1)}\,(2krn)^{2k/(2k+1)}}\sum_{j\geq[\widetilde{\omega}]+1}\mathbf{a}_j\,\theta_j^2,$$

i.e.

$$\limsup_{n\to\infty}\sup_{S\in\Theta_{\varepsilon,L}}\frac{n^{2k/(2k+1)}\upsilon(S)\,\sum_{j\geq[\widetilde{\omega}]+1}\theta_j^2}{\sum_{j=j_*}^{[\widetilde{\omega}]}\mathbf{a}_j\,\theta_j^2} \leq \frac{\iota_k^{2k/(2k+1)}}{(2k\pi\,r)^{2k/(2k+1)}}\,.$$

Moreover, we get directly that

$$\lim_{T\to\infty} \sup_{S\in\Theta_{\varepsilon,L}} \left| \frac{\sum_{j=1}^{d} \widetilde{\lambda}^2(j)}{n^{1/(2k+1)}(\overline{r}(S))^{1/(2k+1)}\mathbf{g}_k} - 1 \right| = 0\,, \tag{A.11}$$

where

$$\mathbf{g}_k = \frac{2k^2}{(\varpi_k)^{2k/(2k+1)}(2k+1)(k+1)}\,.$$

So, taking into account that in (A.10)

$$\lim_{n\to\infty} \sup_{S\in W_{k,r}} n^{2k/(2k+1)}\widetilde{\Xi}_n = 0\,,$$

we obtain the limit equality (A.7) and, hence Lemma A.6.
□

Lemma A.7. *For any non random coefficients* $(\mathbf{u}_j)_{1\le j\le d}$

$$\mathbf{E}\left(\sum_{j=1}^{d}\mathbf{u}_j\eta_{j,d}\right)^2 \le \sigma_{1,*}\sum_{j=1}^{d}\mathbf{u}_j^2\,.$$

Proof. Using the definition of $\eta_{j,d}$ in (3.7), we obtain that

$$\begin{aligned}\mathbf{E}\left(\sum_{j=1}^{d}\mathbf{u}_j\eta_{j,d}\right)^2 &= \frac{b-a}{d}\mathbf{E}\sum_{l=1}^{d}\sigma_l^2\left(\sum_{j=1}^{d}\mathbf{u}_j\phi_j(z_l)\right)^2 \\ &\le \sigma_{1,*}\frac{b-a}{d}\sum_{l=1}^{d}\left(\sum_{j=1}^{d}\mathbf{u}_j\phi_j(z_l)\right)^2\,.\end{aligned}$$

Now, the orthonormality property of the basis functions $(\phi_j(\cdot))_{1\le j\le d}$ implies this lemma. □

Lemma A.8. *Now we show that*

$$\lim_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_{k,r}} |\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n\mathbf{1}_\Gamma| = 0\,. \tag{A.12}$$

Proof. First of all, note that, using the definition of $\mathbf{s}_{j,d}$ in (3.14), we obtain

$$\mathbf{E}_{\mathbf{p},S}\,\eta_{j,d}^2 = \mathbf{E}_{\mathbf{p},S}\,\mathbf{s}_{j,d} = \frac{1}{d}\sum_{l=1}^{d}\mathbf{E}_{\mathbf{p},S}\,\frac{1}{H_l} + \frac{1}{d}\mathbf{E}_{\mathbf{p},S}\overline{\mathbf{s}}_{j,d}\,,$$

where

$$\overline{\mathbf{s}}_{j,d} = \sum_{l=1}^{d} \sigma_l^2 \overline{\phi}_j(x_l) \quad \text{and} \quad \overline{\phi}_j(z) = (b-a)\phi_j^2(z) - 1\,.$$

Therefore, we can represent the expectation of $\mathbf{U}_n$ as

$$\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n = \frac{\|\widetilde{\lambda}\|^2}{d^2}\mathbf{E}_{\mathbf{p},S}\mathbf{U}_{1,n} + \frac{b-a}{d^2}\mathbf{E}_{\mathbf{p},S}\mathbf{U}_{2,n}\,,$$

where $\|\widetilde{\lambda}\|^2 = \sum_{j=1}^{d} \widetilde{\lambda}^2(j)$,

$$\mathbf{U}_{1,n} = \frac{b-a}{d}\sum_{l=1}^{d}\frac{d}{H_l} - \varsigma(S) \quad \text{and} \quad \mathbf{U}_{2,n} = \sum_{j=1}^{d}\widetilde{\lambda}^2(j)\overline{\mathbf{s}}_{j,d}\,.$$

Note now, that using Proposition 2.19 and the dominated convergence theorem in the definition (2.16) we obtain that

$$\lim_{n\to\infty}\max_{1\le l\le d}\sup_{S\in\Theta_{\epsilon,L}}\sup_{\mathbf{p}\in\mathcal{P}}\mathbf{E}_{\mathbf{p},S}\left|\frac{d}{H_l} - (1-S^2(z_l))\right| = 0\,.$$

Taking into account that for the functions from the class (2.4) their derivatives are bounded in modulus by the fixed constant $L > 0$, we can easy deduce that

$$\lim_{n\to\infty}\sup_{S\in\Theta_{\epsilon,L}}\left|\frac{b-a}{d}\sum_{l=1}^{d}(1-S^2(z_l)) - \varsigma(S)\right| = 0\,,$$

i.e.

$$\lim_{n\to\infty}\sup_{S\in\Theta_{\epsilon,L}}\sup_{\mathbf{p}\in\mathcal{P}}|\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_{1,n}| = 0\,.$$

Therefore, taking into account that

$$\limsup_{n\to\infty} n^{2k/(2k+1)}\sup_{S\in\Theta_{\epsilon,L}}\frac{\|\widetilde{\lambda}\|^2}{d^2} < \infty\,,$$

we obtain that

$$\overline{\lim}_{n\to\infty}\, n^{2k/(2k+1)}\sup_{S\in\Theta_{\epsilon,L}}\sup_{\mathbf{p}\in\mathcal{P}}|\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n| \le (b-a)\overline{\lim}_{n\to\infty}\,\mathbf{U}^*_{2,n}\,, \tag{A.13}$$

where

$$\mathbf{U}^*_{2,n} = \frac{n^{2k/(2k+1)}}{d^2}\sup_{S\in\Theta_{\epsilon,L}}\sup_{\mathbf{p}\in\mathcal{P}}|\mathbf{E}_{\mathbf{p},S}\mathbf{U}_{2,n}|\,.$$

Now, using Lemma A.2 from [9] we obtain that

$$\begin{aligned}\big|\,\mathbf{E}_{\mathbf{p},S}\mathbf{U}_{2,n}\big| \;&=\; \left|\sum_{l=1}^{d}\mathbf{E}_{\mathbf{p},S}\,\sigma_l^2\,\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\overline{\phi}_j(z_l)\right| \\ &\le\; d\,\sigma_{1,*}\,(2^{2k+1}+2^{k+2}+1)\le\; 5d\,\sigma_{1,*}\,2^{2k}\,.\end{aligned}$$

Note that from the definition of $\sigma_{1,*}$ in (2.30) it follows that

$$\limsup_{n\to\infty}\, d\,\sigma_{1,*}<\infty\,,$$

i.e.

$$\limsup_{n\to\infty}\;\sup_{S\in\Theta_{\epsilon,L}}\;\sup_{\mathbf{p}\in\mathcal{P}}\big|\,\mathbf{E}_{\mathbf{p},S}\mathbf{U}_{2,n}\big|\;<\infty\,.$$

Therefore, the using this bound in (A.13) implies

$$\overline{\lim}_{n\to\infty}\,n^{2k/(2k+1)}\sup_{S\in\Theta_{\epsilon,L}}\;\sup_{\mathbf{p}\in\mathcal{P}}|\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n|=0\,.$$

Moreover, according to the inequality (A.4) from [4] we have that

$$\mathbf{E}_{\mathbf{p},S}\eta_{j,q}^4\le 64\breve{\mathbf{m}}\sigma_{1,*}^2\,,$$

where the coefficient $\breve{\mathbf{m}}$ is given in (2.29). From this we obtain, that

$$\begin{aligned}\mathbf{E}_{\mathbf{p},S}|\mathbf{U}_n|\mathbf{1}_{\Gamma^c}&\le\frac{(b-a)}{d}\sum_{j=1}^{d}\mathbf{E}_{\mathbf{p},S}\eta_{j,d}^2\mathbf{1}_{\Gamma^c}+\varsigma(S)\mathbf{P}_{\mathbf{p},S}(\Gamma^c)\\ &\le\frac{8\sigma_{1,*}(b-a)\sqrt{\breve{\mathbf{m}}}}{d}\sqrt{\mathbf{P}_{\mathbf{p},S}(\Gamma^c)}+\varsigma(S)\mathbf{P}_{\mathbf{p},S}(\Gamma^c)\,.\end{aligned}$$

So, Proposition 2.2 implies

$$\lim_{n\to\infty}\;n^{2k/(2k+1)}\sup_{S\in W_{k,r}}\mathbf{E}_{\mathbf{p},S}|\mathbf{U}_n|\mathbf{1}_{\Gamma^c}=0\,.$$

This implies the quality (A.12), and we obtain Lemma A.8. □

References

[1] Arkoun, O (2011): Sequential Adaptive Estimators in Nonparametric Autoregressive Models, *Sequential Analysis* 30: 228-246.

[2] Arkoun, O. and Pergamenchtchikov, S. (2008) : Nonparametric Estimation for an Autoregressive Model, *Tomsk State University Journal of Mathematics and Mechanics*. 2: 20-30.

[3] Arkoun, O. and Pergamenchtchikov, S. (2016) : Sequential Robust Estimation for Nonparametric Autoregressive Models. - *Sequential Analysis*, 2016, **35** (4), 489 – 515

[4] Arkoun, O. , Brua, J.-Y. and Pergamenchtchikov, S. (2019) Sequential Model Selection Method for Nonparametric Autoregression. - *Sequential Analysis*, accepted.

[5] Belitser, E. (2000) : Recursive Estimation of a Drifted Autoregressive Parameter, *The Annals of Statistics* 26: 860-870.

[6] Dahlhaus, R. (1996) : Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes, *Journal of Nonparametric Statistics* 6: 171-191.

[7] Dahlhaus, R. (1996) : On the Kullback-Leibler Information Divergence of Locally Stationary Processes, *Stochastic Processes and their Applications* 62: 139-168.

[8] Galtchouk, L. and Pergamenshchikov, S. (2006) : Asymptotically Efficient Estimates for Nonparametric Regression Models, *Statistics and Probability Letters* 76 : 852-860.

[9] Galtchouk, L.I. and Pergamenshchikov, S.M. (2009) Sharp nonasymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21**, 1-16.

[10] Galtchouk, L.I. and Pergamenshchikov, S.M. (2009) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *Journal of the Korean Statistical Society*, **38** (4), 305 - 322.

[11] Galtchouk, L. and Pergamenshchikov, S. (2011) : Adaptive Sequential Estimation for Ergodic Diffusion Processes in Quadratic Metric, *Journal of Nonparametric Statistics* 23: 255-285.

[12] Galtchouk, L. and Pergamenshchikov, S. (2013) Uniform concentration inequality for ergodic diffusion processes observe at discrete times, *Stochastic processes and their applications*, **123**, 91 - 109.

[13] Golubev, G.K. and Levit, B.Ya. (1996) On the second order minimax estimation of distribution functions. - *Mathematical Methods of Statistics*, **5**, 1 – 31.

[14] Konev V.V. On One Property of Martingales with Conditionally Gaussian Increments and Its Application in the Theory of Nonasymptotic Inference. *Doklady Mathematics*, 2016. Vol. 94, No 3. P. 1-5.

[15] Konev, V. and Pergamenshchikov, S. (1984): Estimate of the Number of Observations in Sequential Identification of the Parameters of Dynamical Systems, *Avtomat. i Telemekh* 12: 56-63.

[16] V.V. Konev and S.M. Pergamenshchikov. Efficient robust nonparametric estimation in a semimartingale regression model. *Ann. Inst. Henri Poincaré Probab. Stat.*, **48** (4), 2012, 1217–1244.

[17] V.V. Konev and S.M. Pergamenshchikov. Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic processes and their applications*, **125**, 2015, 294 – 326.

[18] Moulines, E., Priouret, P. and Roueff, F. (2005): On Recursive Estimation for Time Varying Autoregressive Processes, *The Annals of Statistics* 33: 2610-2654.

[19] M.S.Pinsker, Optimal filtration of square integrable signals in gaussian noise, Problems Inform. Transmission 17 (1981) 120-133.