



HAL
open science

Efficient robust sequential analysis for autoregressive big data models

Ouerdia Arkoun, Jean-Yves Brua, Serguei Pergamenshchikov

► **To cite this version:**

Ouerdia Arkoun, Jean-Yves Brua, Serguei Pergamenshchikov. Efficient robust sequential analysis for autoregressive big data models. 2021. hal-03154295

HAL Id: hal-03154295

<https://hal.science/hal-03154295>

Preprint submitted on 28 Feb 2021

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

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

Efficient robust sequential analysis for autoregressive big data models *

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

Abstract

In this paper we study high dimension models based on dependent observations defined through autoregressive processes. For such models we study the efficient robust estimation problem in adaptive settings, i.e. in the case when the nonparametric regularity is unknown. To this end we use the sequential model selection procedures proposed in [4]. First, through the Van Trees inequality, we obtain the sharp lower bound for robust risks in explicit form, i.e. the famous Pinsker's constant (see [28] for example). Then, through sharp non asymptotic oracle inequalities for robust risks, we show that the upper bound for the robust risk of the proposed model selection sequential procedure coincides with the obtained Pinsker constant, i.e. this means that this procedure is efficient in the minimax sens.

MSC: primary 62G08, secondary 62G05

Keywords: Non-parametric estimation; Non parametric autoregression; Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequalities.

*This research was supported by RSF, project no 20-61-47043 (National Research Tomsk State University, Russia).

[†]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 Problem and motivations

In this paper we consider the following autoregressive model

$$y_j = \sum_{i=1}^q \theta_i \psi_i(x_j) y_{j-1} + \xi_j, \quad x_j = a + \frac{(b-a)j}{n}, \quad (1.1)$$

where the functions $(\psi_i)_{1 \leq i \leq q}$ are known linear independent functions, $a < b$ are fixed known constants, $1 \leq j \leq n$, the initial value y_0 is a constant and the noise $(\xi_j)_{j \geq 1}$ 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 unknown parameters $(\theta_j)_{1 \leq j \leq q}$ in the high dimension setting, i.e. when the number of parameters $q > n$. Usually, in these cases for statistical models with independent observations one uses one of two methods: Lasso algorithm (see, for example, [27]) or the Dantzig selector method proposed in [8]. For dependent observations such models were studied, for example, in [11] and [12]. But in all these papers the number of parameters q is known and, therefore, unfortunately, these methods can't be used to estimate, for example, the number of parameters q . In this paper similar to [18, 19] we study this problem in nonparametric setting, i.e. we consider the observations (1.1) defined as

$$y_j = S(x_j) y_{j-1} + \xi_j, \quad (1.2)$$

where $S(\cdot) \in \mathbf{L}_2[a, b]$ is unknown function. The nonparametric setting allows to consider the models (1.1) with unknown q or with $q = +\infty$. Note that the case when the number of parameters q is unknown is one of challenging problems in the signal and image processing theory (see, for example, [7, 5]). Therefore, the problem is now to estimate the function $S(\cdot)$ in the model on the basis of the observations (1.2) 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 [9], [10] and [6]. In all these papers, the authors propose some asymptotic (as $n \rightarrow \infty$) methods for different identification studies without considering optimal estimation issues. Firstly, minimax estimation problems for the model (1.2) has been treated in [2] and [25] 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.

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

$$\mathcal{R}_{\mathbf{p}}(\widehat{S}_n, S) = \mathbf{E}_{\mathbf{p}, S} \|\widehat{S}_n - S\|^2, \quad \|S\|^2 = \int_a^b S^2(x) dx, \quad (1.3)$$

where \widehat{S}_n is an estimator of S based on observations $(y_j)_{1 \leq j \leq n}$ and $\mathbf{E}_{\mathbf{p}, S}$ is the expectation with respect to the distribution law $\mathbf{P}_{\mathbf{p}, S}$ of the process $(y_j)_{1 \leq j \leq n}$ given the distribution density p and the function S . Moreover, taking into account that the distribution p is unknown, we use the robust nonparametric estimation method developed in [13]. To this end we set the robust risk as

$$\mathcal{R}^*(\widehat{S}_n, S) = \sup_{\mathbf{p} \in \mathcal{P}} \mathcal{R}_{\mathbf{p}}(\widehat{S}_n, S), \quad (1.4)$$

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

1.2 Main tools

To estimate the function S in model (1.2) we make use of the model selection procedures proposed in [4]. These procedures are based on the family of the optimal pointwise truncated sequential estimators from [3] for which a sharp oracle inequality is shown, using the model selection method developed in [14]. In this paper, using this inequality, we show that the model selection procedure is efficient in adaptive setting for the robust quadratic risk (1.4). To this end, first of all, we have to study the sharp lower bound for the these risks, i.e. we have to provide the best potential accuracy estimation for the model (1.2) which is called the Pinsker constant. For this we use the approach proposed in [15, 16] which is based on the Van-Trees inequality. It turns out that for the model (1.2) the Pinsker constant has the same form as for the filtration signal problem in the "signal - white noise" model studied in [28] but with new coefficient which equals to the optimal variance given by the Hajek-Le Cam inequality for the parametric model (1.2). 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 risk (1.4).

1.3 Plan of the paper

The paper is organized as follows. In Section 2 we give all conditions and construct the sequential pointwise estimation procedures which allows us to pass from the autoregression model to the corresponding regression model. We construct in Section 3 the model selection procedure based on the sequential estimators from Section 2. Then we announce the main results in Section 4. In Section 5 we show the Van-Trees inequality for the model (1.2). We obtain the lower bound for the robust risk in section 6 and we get, in Section 7, the upper bound for the robust risk of the constructed sequential estimator. In Appendix A we give the all auxiliary and technic tools.

2 Sequential procedure

As in [3] we assume that in the model (1.2) the i.i.d. random variables $(\xi_j)_{j \geq 1}$ have a density p (with respect to the Lebesgue measure) from the functional class \mathcal{P} defined as

$$\mathcal{P} := \left\{ \mathbf{p} \geq 0 : \int_{-\infty}^{+\infty} \mathbf{p}(x) dx = 1, \quad \int_{-\infty}^{+\infty} x \mathbf{p}(x) dx = 0, \right. \\ \left. \int_{-\infty}^{+\infty} x^2 \mathbf{p}(x) dx = 1 \quad \text{and} \quad \sup_{l \geq 1} \frac{\int_{-\infty}^{+\infty} |x|^{2l} \mathbf{p}(x) dx}{\zeta^k (2l-1)!!} \leq 1 \right\}, \quad (2.1)$$

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

$$\lim_{n \rightarrow \infty} \frac{\zeta(n)}{n^{\mathbf{b}}} = 0. \quad (2.2)$$

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 $\mathbf{b} > 0$

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

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 (1.2), 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

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

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

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

$$z_l = a + \frac{l}{d+1}(b-a), \quad (2.5)$$

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

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

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

$$u_{l,j} = \frac{x_j - z_l}{h}, \quad (2.8)$$

where the bandwidth h will be chosen later such that $a < z_1 - h < z_d + h < b$, i.e. $h < (b-a)/d$.

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

$$k_{1,l} = \lfloor n\tilde{z}_l - n\tilde{h} \rfloor \quad \text{and} \quad k_{2,l} = \lfloor n\tilde{z}_l + n\tilde{h} \rfloor, \quad (2.9)$$

where $\lfloor x \rfloor$ is the smallest integer greater than $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the integer part of x , $\tilde{z}_l = (z_l - a)/(b-a)$ and $\tilde{h} = h/(b-a)$.

Moreover, we assume here that

$$n > \tilde{h}^{-1}. \quad (2.10)$$

Now we chose ι_l in (2.7) as

$$\iota_l = k_{1,l} + \mathbf{q} \quad \text{and} \quad \mathbf{q} = \mathbf{q}_n = \lfloor (n\tilde{h})^{\mu_0} \rfloor \quad (2.11)$$

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

$$A_{k,m} = \sum_{j=k+1}^m Q_{l,j} y_{j-1}^2 \quad \text{and} \quad A_m = A_{0,m}. \quad (2.12)$$

Next, similarly to [1], we use some kernel sequential procedure based on the observations $(y_j)_{\iota_l \leq j \leq n}$

To transform the kernel estimator in the linear function of observations and we replace the number of observations n by the following stopping time

$$\tau_l = \inf\{\iota_l + 1 \leq k \leq k_{2,l} : A_{\iota_l,k} \geq H_l\}, \quad (2.13)$$

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 $\{y_1, \dots, y_{\iota_l}\}$.

Now we define the sequential estimator as

$$S_l^* = \frac{1}{H_l} \left(\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j} y_{j-1} y_j + \varkappa_l Q(u_{l,\tau_l}) y_{\tau_l-1} y_{\tau_l} \right) \mathbf{1}_{\Gamma_l}, \quad (2.14)$$

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

$$A_{\iota_l, \tau_l-1} + \varkappa_l^2 Q(u_{l,\tau_l}) y_{\tau_l-1}^2 = H_l. \quad (2.15)$$

We set here $A_{\iota_l, \tau_l-1} = 0$ if $\tau_l = \iota_l + 1$. Note that, to obtain the efficient kernel estimate of $S(z_l)$, we need to use the all $k_{2,l} - \iota_l$ observations.

Similarly to [21], one can show that

$$\tau_l \approx (1 - S^2(z_l)) H_l \quad \text{as} \quad H_l \rightarrow \infty. \quad (2.16)$$

Therefore, it will be natural to choose H_l as $(k_{2,l} - \iota_l)/(1 - S^2(z_l))$.

Since the value $S(z_l)$ is unknown, we define the threshold H_l as

$$H_l = \frac{1 - \tilde{\epsilon}}{1 - \tilde{S}_l^2} (k_{2,l} - \iota_l) \quad \text{and} \quad \tilde{\epsilon} = \frac{1}{2 + \ln n}, \quad (2.17)$$

where \tilde{S}_l is the projection of the estimator \hat{S}_l in $] -1 + \tilde{\epsilon}, 1 - \tilde{\epsilon}[$, i.e.

$$\tilde{S}_l = \min(\max(\hat{S}_l, -1 + \tilde{\epsilon}), 1 - \tilde{\epsilon}). \quad (2.18)$$

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

$$h = \frac{b - a}{2d}. \quad (2.19)$$

As to the estimator \hat{S}_l , we can show the following property.

Proposition 2.1. *The convergence rate in probability of the estimator (2.18) is more rapid than any power function, i.e. for any $\mathbf{b} > 0$*

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \max_{1 \leq l \leq d} \sup_{S \in \Theta_{\epsilon, L}} \sup_{\mathbf{p} \in \mathcal{P}} \mathbf{P}_{\mathbf{p}, S} \left(|\tilde{S}_l - S(z_l)| > \epsilon_0 \right) = 0, \quad (2.20)$$

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

Now we set

$$Y_l = S_l^* \mathbf{1}_{\Gamma} \quad \text{and} \quad \Gamma = \bigcap_{l=1}^d \Gamma_l. \quad (2.21)$$

Using the convergence (2.20) 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*

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \sup_{S \in \Theta_{\epsilon, L}} \mathbf{P}_{\mathbf{p}, S} (\Gamma^c) = 0. \quad (2.22)$$

In view of this proposition we can negligible the set Γ^c . So, using the estimators (2.21) on the set Γ we obtain the discrete time regression model

$$Y_l = S(z_l) + \zeta_l \quad \text{and} \quad \zeta_l = \xi_l^* + \varpi_l, \quad (2.23)$$

in which

$$\xi_l^* = \frac{\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j} y_{j-1} \xi_j + \varkappa_l Q(u_{l,\tau_l}) y_{\tau_l-1} \xi_{\tau_l}}{H_l} \quad (2.24)$$

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

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

and

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

Note that in the model (2.23) the random variables $(\xi_j^*)_{1 \leq j \leq d}$ 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 $j \geq 1$ we set

$$\check{Q}_{l,j} = Q_{l,j} y_{j-1} \mathbf{1}_{\{j < k_{2,l}\}} + \sqrt{H_l} Q_{l,j} \mathbf{1}_{\{j = k_{2,l}\}} \quad (2.25)$$

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

$$\check{Q}_{l,j} \check{Q}_{m,j} = 0, \quad (2.26)$$

and $\check{A}_{\iota_l, k_{2,l}} \geq H_l$. So we can modify now the stopping time (2.13) as

$$\check{\tau}_l = \inf\{k \geq \iota_l + 1 : \check{A}_{\iota_l, k} \geq H_l\}. \quad (2.27)$$

Obviously, $\check{\tau}_l \leq k_{2,l}$ and $\check{\tau}_l = \tau_l$ on the set Γ for any $1 \leq l \leq d$. Now similarly to (2.15) we define the correction coefficient $\check{\varkappa}_l$ as

$$\check{A}_{\iota_l, \check{\tau}_l - 1} + \check{\varkappa}_l^2 \check{Q}_{l, \check{\tau}_l}^2 = H_l. \quad (2.28)$$

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

$$\eta_l = \frac{\sum_{j=\iota_l+1}^{\check{\tau}_l-1} \check{Q}_{l,j} \xi_j + \check{\varkappa}_l \check{Q}_{l, \check{\tau}_l} \xi_{\check{\tau}_l}}{H_l}. \quad (2.29)$$

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 $p \in \mathcal{P}$

$$\mathbf{E}_{\mathbf{p}, S}(\eta_l | \mathcal{G}_l) = 0, \quad \mathbf{E}_{\mathbf{p}, S}(\eta_l^2 | \mathcal{G}_l) = \sigma_l^2 \quad \text{and} \quad \mathbf{E}_{\mathbf{p}, S}(\eta_l^4 | \mathcal{G}_l) \leq \check{\mathbf{m}} \sigma_l^4, \quad (2.30)$$

where $\sigma_l = H_l^{-1/2}$, $\mathcal{G}_l = \sigma\{\eta_1, \dots, \eta_{l-1}, \sigma_l\}$ and $\check{\mathbf{m}} = 4(144)^4 \mathbf{m}_4^*$. Note that

$$\sigma_{0,*} \leq \min_{1 \leq l \leq d} \sigma_l^2 \leq \max_{1 \leq l \leq d} \sigma_l^2 \leq \sigma_{1,*}, \quad (2.31)$$

where

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

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

$$\sup_{S \in \Theta_{\epsilon, L}} \mathbf{E}_{\mathbf{p}, S} \mathbf{1}_{\Gamma} \varpi_l^2 \leq \left(L^2 h^2 + \frac{\check{v}_n}{(nh)^2} \right), \quad (2.32)$$

where $\check{v}_n = \sup_{\mathbf{p} \in \mathcal{P}} \sup_{S \in \Theta_{\epsilon, L}} \mathbf{E}_{\mathbf{p}, S} \max_{1 \leq j \leq n} y_j^4$. As it is shown in [4] (Theorem 3.3) for any $\mathbf{b} > 0$

$$\lim_{n \rightarrow \infty} n^{-\mathbf{b}} \check{v}_n = 0. \quad (2.33)$$

Remark 2.1. *It should be noted that the property (2.33) means that the asymptotic behavior of the upper bound (2.32) approximately almost as h^{-2} when $n \rightarrow \infty$.*

Remark 2.2. *Note that to estimate the function S in (1.2), we use the approach developed in [16] 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 (2.14).*

Remark 2.3. *It should also be noted that the sequential estimator (2.14) 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 [20]. We modify this procedure to calculate the variance of the stochastic term (2.24).*

3 Model selection

Now to estimate the function S we use the sequential model selection procedure from [4] for the regression (2.23). To this end, first we chose the trigonometric basis $(\phi_j)_{j \geq 1}$ in $\mathbf{l}_2[a, b]$, i.e.

$$\phi_1 = \frac{1}{\sqrt{b-a}}, \quad \phi_j(x) = \sqrt{\frac{2}{b-a}} \operatorname{Tr}_j(2\pi[j/2]\mathbf{l}_0(x)), \quad j \geq 2, \quad (3.1)$$

where the function $\operatorname{Tr}_j(x) = \cos(x)$ for even j and $\operatorname{Tr}_j(x) = \sin(x)$ for odd j , and $\mathbf{l}_0(x) = (x-a)/(b-a)$. Moreover, we chose the odd number d of regression points (2.5), for example, $d = 2[\sqrt{n}/2] + 1$. Then the functions $(\phi_j)_{1 \leq j \leq d}$ are orthonormal for the empirical inner product, i.e.

$$(\phi_i, \phi_j)_d = \frac{b-a}{d} \sum_{l=1}^d \phi_i(z_l) \phi_j(z_l) = \mathbf{1}_{\{i=j\}}. \quad (3.2)$$

It is clear that, the function S can be represented as

$$S(z_l) = \sum_{j=1}^d \theta_{j,d} \phi_j(z_l) \quad \text{and} \quad \theta_{j,d} = (S, \phi_j)_d. \quad (3.3)$$

We define the estimators for the coefficients $(\theta_{j,d})_{1 \leq j \leq d}$ as

$$\hat{\theta}_{j,d} = \frac{b-a}{d} \sum_{l=1}^d Y_l \phi_j(z_l). \quad (3.4)$$

From (2.23) we obtain immediately the following regression on the set Γ

$$\hat{\theta}_{j,d} = \theta_{j,d} + \zeta_{j,d} \quad \text{with} \quad \zeta_{j,d} = \sqrt{\frac{b-a}{d}} \eta_{j,d} + \varpi_{j,d}, \quad (3.5)$$

where

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

Through the Bounyakovskii-Cauchy-Schwarz we get that

$$|\varpi_{j,d}| \leq \|\varpi\|_d \|\phi_j\|_d = \|\varpi\|_d := \frac{\varpi_n^*}{n}. \quad (3.6)$$

Moreover, from (2.32) and (2.33) it follows that for any $\mathbf{b} > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n^{\mathbf{b}}} \sup_{\mathbf{p} \in \mathcal{P}} \sup_{S \in \Theta_{\epsilon,L}} \mathbf{E}_{\mathbf{p},S} \varpi_n^* \mathbf{1}_\Gamma = 0. \quad (3.7)$$

To construct the model selection procedure we use weighted least squares estimators which for the points (2.5) defined as

$$\widehat{S}_\lambda(z_l) = \sum_{j=1}^d \lambda(j) \widehat{\theta}_{j,d} \phi_j(z_l) \mathbf{1}_\Gamma, \quad 1 \leq l \leq d, \quad (3.8)$$

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

$$\widehat{S}_\lambda(t) = \sum_{l=1}^d \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < t \leq z_l\}}. \quad (3.9)$$

Denote by ν the cardinal number of the set Λ , for which we impose the following condition.

A₁) Assume that the number of the weight vectors ν is a function of n , i.e. $\nu = \nu_n$, such that for any $\mathbf{b} > 0$

$$\lim_{n \rightarrow \infty} \frac{\nu_n}{n^{\mathbf{b}}} = 0. \quad (3.10)$$

To choose a weight vector $\lambda \in \Lambda$ in (3.8) we will use the following risk

$$\text{Err}_d(\lambda) = \|\widehat{S}_\lambda - S\|_d^2 = \frac{b-a}{d} \sum_{l=1}^d (\widehat{S}_\lambda(z_l) - S(z_l))^2. \quad (3.11)$$

Using (3.3) and (3.8) it can be represented as

$$\text{Err}_d(\lambda) = \sum_{j=1}^d \lambda^2(j) \widehat{\theta}_{j,d}^2 - 2 \sum_{j=1}^d \lambda(j) \widehat{\theta}_{j,d} \theta_{j,d} + \sum_{j=1}^d \theta_{j,d}^2. \quad (3.12)$$

Since the coefficients $\theta_{j,d}$ are unknown we can't minimize this risk directly to obtain an optimal weight vector. To modify it we set

$$\tilde{\theta}_{j,d} = \widehat{\theta}_{j,d}^2 - \frac{b-a}{d} \mathbf{s}_{j,d} \quad \text{with} \quad \mathbf{s}_{j,d} = \frac{b-a}{d} \sum_{l=1}^d \sigma_l^2 \phi_j^2(z_l). \quad (3.13)$$

Note here that in view of (2.31) - (3.2) the last term can be estimated as

$$\sigma_{0,*} \leq \mathbf{s}_{j,d} \leq \sigma_{1,*}. \quad (3.14)$$

Now, we modify the risk (3.12) as

$$J_d(\lambda) = \sum_{j=1}^d \lambda^2(j) \widehat{\theta}_{j,d}^2 - 2 \sum_{j=1}^d \lambda(j) \tilde{\theta}_{j,d} + \delta P_d(\lambda), \quad (3.15)$$

where the coefficient $0 < \delta < 1$ will be chosen later and the penalty term is

$$P_d(\lambda) = \frac{b-a}{d} \sum_{j=1}^d \lambda^2(j) \mathbf{s}_{j,d} \quad (3.16)$$

Now using (3.15) we define the sequential model selection procedure as

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_d(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}. \quad (3.17)$$

To study the efficiency property we specify the weight coefficients $(\lambda(j))_{1 \leq j \leq n}$ as it is proposed, for example, in [15]. First, for some $0 < \varepsilon < 1$ introduce the two dimensional grid to adapt to the unknown parameters (regularity and size) of the Sobolev ball, i.e. we set

$$\mathcal{A} = \{1, \dots, k^*\} \times \{\varepsilon, \dots, \mathbf{m}\varepsilon\}, \quad (3.18)$$

where $\mathbf{m} = [1/\varepsilon^2]$. 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

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} k^*(n) = +\infty, \quad \lim_{n \rightarrow \infty} \frac{k^*(n)}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^{\mathbf{b}} \varepsilon(n) = +\infty \end{array} \right. \quad (3.19)$$

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

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k_0^* + [\sqrt{\ln n}], \quad (3.20)$$

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

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

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j < j_*\}} + (1 - (j/\omega_\alpha)^\beta) \mathbf{1}_{\{j_* \leq j \leq \omega_\alpha\}}, \quad (3.21)$$

where $j_* = 1 + \lceil \ln n \rceil$ and

$$\omega_\alpha = \left(\frac{(\beta + 1)(2\beta + 1)}{\pi^{2\beta} \beta} \mathbf{t} n \right)^{1/(2\beta+1)}.$$

Now we define the set Λ as

$$\Lambda = \{\lambda_\alpha, \alpha \in \mathcal{A}\}. \quad (3.22)$$

Note that these weight coefficients are used in [23, 24] 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 (3.19) imply the condition (3.10). In [4] we showed the following result.

Theorem 3.1. *Assume that the conditions (2.2) and (3.10) hold. Then for any $n \geq 3$, any $S \in \Theta_{\epsilon, L}$ and any $0 < \delta \leq 1/12$, the procedure (3.17) with the coefficients (3.22) satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{(1 + 4\delta)(1 + \delta)^2}{1 - 6\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{B}_n^*}{\delta n}, \quad (3.23)$$

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

$$\lim_{n \rightarrow \infty} \frac{\mathbf{B}_n^*}{n^{\mathbf{b}}} = 0.$$

Remark 3.1. *In this paper we will use the inequality (3.23) to study efficiency properties for the model selection procedure (3.17) with the weight coefficients (3.22) in adaptive setting, i.e. in the case when the regularity of the function S (1.2) is unknown.*

4 Main results

First, to study the minimax properties for the estimation problem for the model (1.2) we need to introduce some functional class. To this end for any

fixed $\mathbf{r} > 0$ and $\mathbf{k} \geq 2$ we set

$$\mathcal{W}_{\mathbf{k},\mathbf{r}} = \left\{ f \in \Theta_{\epsilon,L} : \sum_{j=1}^{+\infty} \mathbf{a}_j \theta_j^2 \leq \mathbf{r} \right\}, \quad (4.1)$$

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

$$\theta_j = \int_a^b f(x) \phi_j(x) dx$$

and $(\phi_j)_{j \geq 1}$ is the trigonometric basis defined in (3.1). It is clear that we can represent this functional class as the Sobolev ball in $\Theta_{\epsilon,L}$, i.e.

$$\mathcal{W}_{\mathbf{k},\mathbf{r}} = \left\{ f \in \Theta_{\epsilon,L} : \sum_{j=0}^{\mathbf{k}} \|f^{(j)}\|^2 \leq \mathbf{r} \right\}. \quad (4.2)$$

Now, for this set we define the normalizing coefficients

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

and

$$\varsigma_* = \varsigma_*(S) = \int_a^b (1 - S^2(u)) du. \quad (4.4)$$

It is well known that for nonparametric adaptive estimation problems in regression models with the functions $S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}$ the minimax convergence rate is $n^{-2\mathbf{k}/(2\mathbf{k}+1)}$ (see, for example, [15, 22] and the references therein). Our goal in this paper is to show the same property for the non parametric autoregressive models (1.2). To this end, first we obtain the lower bound for the normalized risk (1.4) in the class of all estimators Ξ_n , i.e. any measurable function with respect to the observations $\sigma\{y_1, \dots, y_n\}$.

Theorem 4.1. *For the model (1.2) with the noise distribution from the class \mathcal{P} defined in (2.1)*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n \in \Xi_n} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathcal{R}^*(\hat{S}_n, S) \geq \mathbf{l}_*(\mathbf{r}), \quad (4.5)$$

where $v(S) = \varsigma_*^{-2\mathbf{k}/(2\mathbf{k}+1)}$.

Now to study the procedure (3.17) we have to add some condition on the penalty coefficient δ which provides sufficiently small penalty term in (3.15). **A₂)** Assume that the parameter δ is a function of n , i.e. $\delta = \delta_n$ such that

$$\lim_{n \rightarrow \infty} \delta_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\delta_n}{n^{\mathbf{b}}} = 0 \quad (4.6)$$

for any $\mathbf{b} > 0$.

Theorem 4.2. Assume that the conditions **A₁)** – **A₂)** hold. The model selection procedure \widehat{S}_* defined in (3.17) with the penalty coefficient satisfying the hypothesis **A₂)** admits the following asymptotic upper bound

$$\limsup_{n \rightarrow \infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathcal{R}(\widehat{S}_*, S) \leq \mathbf{l}_*(\mathbf{r}). \quad (4.7)$$

Corollary 4.3. Assume that the conditions **A₁)** – **A₂)** hold. The model selection procedure \widehat{S}_* defined in (3.17) with the penalty coefficient satisfying the condition **A₂)** is efficient, i.e.

$$\lim_{n \rightarrow \infty} \frac{\inf_{\widehat{S}_n \in \Xi_n} \sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathcal{R}^*(\widehat{S}_n, S)}{\sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathcal{R}(\widehat{S}_*, S)} = 1. \quad (4.8)$$

Moreover,

$$\lim_{n \rightarrow \infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathcal{R}(\widehat{S}_*, S) = \mathbf{l}_*(\mathbf{r}). \quad (4.9)$$

Remark 4.1. Note that the limit equalities (4.8) and (4.9) imply that the function $\mathbf{l}_*(\mathbf{r})/v(S)$ is the minimal value of the normalised asymptotic quadratic robust risk, i.e. the Pinsker constant in this case. We remind that the coefficient $\mathbf{l}_*(\mathbf{r})$ is the well known Pinsker constant for the "signal+standard white noise" model obtained in [28]. Therefore, the Pinsker constant for the model (1.2) 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 nonparametric autoregressive model (1.2) with the $(0, 1)$ gaussian i.i.d. random variable $(\xi_j)_{1 \leq j \leq n}$ and the parametric linear function S , i.e.

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

We assume that the functions $(\psi_i)_{1 \leq i \leq d}$ are orthogonal with respect to the scalar product (3.2).

Let now \mathbf{P}_θ^n be the distribution in \mathbb{R}^n of the observations $y = (y_1, \dots, y_n)$ in the model (1.2) with the function (5.1) and ν_ξ^n be the distribution in \mathbb{R}^n of the gaussian vector (ξ_1, \dots, ξ_n) . In this case the Radon - Nykodim density is given as

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

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 $] -\mathbf{N}_j, \mathbf{N}_j[$, i.e. $\mathbf{u}_j(z) > 0$ for any $-\mathbf{N}_j < z < \mathbf{N}_j$ and $\mathbf{u}_j(z) = 0$ for all $|z| \geq \mathbf{N}_j$, such that the Fisher information is finite, i.e.

$$I_j = \int_{-\mathbf{N}_j}^{\mathbf{N}_j} \frac{\dot{\mathbf{u}}_j^2(z)}{\mathbf{u}_j(z)} dz < \infty. \quad (5.3)$$

Now, we set

$$\mathcal{Y} =] -\mathbf{N}_1, \mathbf{N}_1[\times \dots \times] -\mathbf{N}_d, \mathbf{N}_d[\subseteq \mathbb{R}^d. \quad (5.4)$$

Let $g(\theta)$ be a continuously differentiable $\mathcal{Y} \rightarrow \mathbb{R}$ function such that, for each $1 \leq j \leq d$,

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

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

$$\begin{aligned} \tilde{\mathbf{E}} H &= \int_{\mathcal{Y}} \left(\int_{\mathbb{R}^n} H(y, \theta) d\mathbf{P}_\theta^n \right) \mathbf{u}(\theta) d\theta \\ &= \int_{\mathcal{Y}} \left(\int_{\mathbb{R}^n} H(y, \theta) f_n(y, \theta) \mathbf{u}(\theta) d\nu_\xi^n \right) d\theta. \end{aligned} \quad (5.6)$$

Let \mathcal{F}_n^y be the field generated by the observations (1.2), i.e. $\mathcal{F}_n^y = \sigma\{y_1, \dots, y_n\}$. Now we study the Gaussian the model (1.2) with the function (5.1).

Lemma 5.1. For any \mathcal{F}_n^y -measurable square integrable $\mathbb{R}^n \rightarrow \mathbb{R}$ function \widehat{g}_n and for any $1 \leq j \leq d$, the mean square accuracy of the function $g(\cdot)$ with respect to the distribution (5.6) can be estimated from below as

$$\widetilde{\mathbf{E}}(\widehat{g}_n - g(\theta))^2 \geq \frac{\bar{g}_j^2}{\widetilde{\mathbf{E}}\Psi_{n,j} + I_j}, \quad (5.7)$$

where $\Psi_{n,j} = \sum_{l=1}^n \psi_j^2(x_l) y_{l-1}^2$ and $\bar{g}_j = \int_{\mathcal{Y}} g'_j(\theta) \mathbf{u}(\theta) d\theta$.

Proof. First, for any $\theta \in \mathcal{Y}$ we set

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

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

$$\begin{aligned} \widetilde{\mathbf{E}} \left((\widehat{g}_n - g(\theta)) \widetilde{U}_j \right) &= \int_{\mathbb{R}^n \times \mathcal{Y}} (\widehat{g}_n(y) - g(\theta)) \frac{\partial}{\partial \theta_j} (f_n(y, \theta) \mathbf{u}(\theta)) d\theta d\nu_{\xi}^{(n)} \\ &= \int_{\mathbb{R}^n \times \mathcal{Y}_j} \left(\int_{-\mathbf{N}_j}^{\mathbf{N}_j} g'_j(\theta) f_n(y, \theta) \mathbf{u}(\theta) d\theta_j \right) \left(\prod_{i \neq j} d\theta_i \right) d\nu_{\xi}^{(n)} = \bar{g}_j, \end{aligned}$$

where

$$\mathcal{Y}_j = \prod_{i \neq j}] - \mathbf{N}_i, \mathbf{N}_i[.$$

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

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

To study the denominator on the right 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 $\theta \in \mathbb{R}^d$,

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

and

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

Using the equality

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

we get

$$\tilde{\mathbf{E}} \tilde{U}_j^2 = \tilde{\mathbf{E}} \Psi_{n,j} + I_j,$$

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

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

6 Lower bound

First, taking into account that the $(0, 1)$ gaussian density \mathbf{p}_0 belongs to the class (2.1), we obtain that

$$\mathcal{R}^*(\hat{S}_n, S) \geq \mathcal{R}_{\mathbf{p}_0}(\hat{S}_n, S). \quad (6.1)$$

Now, using the definition (4.3) we set

$$d = d_n = \left\lceil \frac{\mathbf{k} + 1}{\mathbf{k}} \left(\frac{n}{b - a} \right)^{1/(2\mathbf{k}+1)} \mathbf{l}_*(\mathbf{r}_\rho) \right\rceil \quad (6.2)$$

where $\mathbf{r}_\rho = \rho \mathbf{r}$, $0 < \rho < 1$ and

$$\mathbf{l}_*(\mathbf{r}_\rho) = ((1 + 2\mathbf{k})\mathbf{r}_\rho)^{1/(2\mathbf{k}+1)} \left(\frac{\mathbf{k}}{\pi(\mathbf{k} + 1)} \right)^{2\mathbf{k}/(2\mathbf{k}+1)} = \rho^{1/(2\mathbf{k}+1)} \mathbf{l}_*(\mathbf{r}).$$

For any vector $\kappa = (\kappa_j)_{1 \leq j \leq d} \in \mathbb{R}^d$, we set

$$S_\kappa(x) = \sum_{j=1}^{d_n} \kappa_j \phi_j(x), \quad (6.3)$$

where $(\phi_j)_{1 \leq j \leq d_n}$ is the trigonometric basis defined in (3.1). To define the bayesian risk we choose a prior distribution on \mathbb{R}^d as

$$\kappa = (\kappa_j)_{1 \leq j \leq d_n} \quad \text{and} \quad \kappa_j = s_j \eta_j^*, \quad (6.4)$$

where η_j^* are i.i.d. random variables with the continuously differentiable density $\rho_n(\cdot)$ defined in Lemma A.1 with $\mathbf{N} = \ln n$ and $n > e^2$,

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

In the sequel we will denote by μ_κ the distribution of the vector (6.4) in \mathbb{R}^d . It is clear that almost sure the function (6.3) can be bounded as

$$\max_{a \leq x \leq b} \left(|S_\kappa(x)| + |\dot{S}_\kappa(x)| \right) \leq \mathbf{c}_* \frac{\ln n}{\sqrt{n}} \sum_{j=1}^{d_n} j \left(\frac{d_n}{j}\right)^{\mathbf{k}/2} := \delta_n^*, \quad (6.5)$$

where

$$\mathbf{c}_* = \sqrt{\frac{2}{b-a} \frac{b-a+\pi}{b-a}}.$$

One can, check directly, that $\delta_n^* \rightarrow 0$ as $n \rightarrow \infty$ for $\mathbf{k} \geq 2$. Therefore, for sufficiently large n the function (6.3) belongs to the class (2.4).

Now, for any function f , we denote by $\mathbf{h}(f)$ its projection in $\mathbf{L}_2[a, b]$ onto the ball $\mathcal{W}_{\mathbf{r}}^* = \{f \in \mathbf{L}_2[a, b] : \|f\| \leq \mathbf{r}\}$, i.e.

$$\mathbf{h}(f) = \frac{\mathbf{r}}{\max(\mathbf{r}, \|f\|)} f.$$

Since $\mathcal{W}_{\mathbf{k}, \mathbf{r}} \subset \mathcal{W}_{\mathbf{r}}^*$, we obtain that for any function $S \in \mathcal{W}_{\mathbf{k}, \mathbf{r}}$

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

Therefore, for sufficiently large n

$$\begin{aligned} \sup_{S \in \mathcal{W}_{\mathbf{k}, \mathbf{r}}} v(S) \mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S) &\geq \int_{\mathbf{D}_n} v(S_z) \mathbf{E}_{\mathbf{p}_0, S_z} \|\widehat{\mathbf{h}}_n - S_z\|^2 \mu_\kappa(dz) \\ &\geq v_* \int_{\mathbf{D}_n} \mathbf{E}_{\mathbf{p}_0, S_z} \|\widehat{\mathbf{h}}_n - S_z\|^2 \mu_\kappa(dz), \end{aligned} \quad (6.6)$$

where

$$\mathbf{D}_n = \left\{ z = (z_1, \dots, z_d)' \in \mathbb{R}^d : \sum_{j=1}^d \mathbf{a}_j z_j^2 \leq \mathbf{r} \right\}$$

and

$$v_* = \inf_{|S|_* \leq \delta_n^*} v(S) \rightarrow \frac{1}{(b-a)^{2k/(2k+1)}} \quad \text{as } n \rightarrow \infty.$$

Using the distribution μ_κ we introduce the following Bayes risk

$$\tilde{\mathcal{R}}_0(\widehat{S}_n) = \tilde{\mathbf{E}}_0 \|\widehat{S}_n - S_z\|^2,$$

where for any random variable measurable ξ with respect to $\sigma\{y_1, \dots, y_n\}$

$$\tilde{\mathbf{E}}_0 \xi = \int_{\mathbb{R}^d} \mathbf{E}_{\mathbf{p}_0, S_z} \xi \mu_\kappa(dz). \quad (6.7)$$

Now taking into account that $\|\widehat{\mathbf{h}}_n\|^2 \leq r$ we get

$$\sup_{S \in \mathcal{W}_{\mathbf{k}, r}} v(S) \mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S) \geq v_* \tilde{\mathcal{R}}_0(\widehat{\mathbf{h}}_n) - 2v_* \mathbf{R}_{0,n} \quad (6.8)$$

with

$$\mathbf{R}_{0,n} = \int_{\mathbf{D}_n^c} (r + \|S_z\|^2) \mu_\kappa(dz) = \int_{\mathbf{D}_n^c} \left(r + \sum_{j=1}^d z_j^2 \right) \mu_\kappa(dz).$$

In Lemma A.3 we study the last term in this inequality. Now it is easy to see that for any $z = (z_1, \dots, z_d)' \in \mathbb{R}^d$

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

where $\widehat{z}_j = \int_a^b \widehat{\mathbf{h}}_n(t) \phi_j(t) dt$.

Now, we will use Lemma 5.1 with $\mathbf{u}_j(z) = \rho_n(z/s_j)/s_j$ and $g(\theta) = \theta_j$ for $1 \leq l \leq d$. It is clear that this is probability density with the support $[-s_j \ln n, s_j \ln n]$ and the Fisher information

$$I_j = s_j^{-2} J_n \quad \text{and} \quad J_n = \int_{-\ln n}^{\ln n} \frac{(\dot{\rho}_n(t))^2}{\rho_n(t)} dt.$$

Therefore, Lemma 5.1 implies that for any $1 \leq j \leq d$ and any \mathcal{F}_n^y measurable random variable $\widehat{\kappa}_j$

$$\tilde{\mathbf{E}}_0 (\widehat{\kappa}_j - \kappa_j)^2 \geq \frac{1}{\tilde{\mathbf{E}}_0 \Psi_{n,j} + s_j^{-2} J_n} \quad \text{and} \quad \Psi_{n,j} = \sum_{l=1}^n \phi_j^2(x_l) y_{l-1}^2.$$

Therefore, the Bayes risk can be estimated from below as

$$\tilde{\mathcal{R}}_0(\hat{\mathbf{h}}_n) \geq \sum_{j=1}^{d_n} \frac{1}{\tilde{\mathbf{E}}_0 \Psi_{n,j} + s_j^{-2} J_n} = \frac{b-a}{n} \sum_{j=1}^{d_n} \frac{1}{\bar{\Psi}_{n,j} + (s_j^*)^{-1} J_n},$$

where

$$\bar{\Psi}_{n,j} = \frac{(b-a)}{n} \tilde{\mathbf{E}}_0 \Psi_{n,j} = \frac{(b-a)}{n} \sum_{l=1}^n \phi_j^2(x_l) \tilde{\mathbf{E}}_0 y_{l-1}^2. \quad (6.9)$$

Note here that in view of Lemma 5.1 and Lemma A.2 the term $J_n \rightarrow 1$ and uniformly over $1 \leq j \leq d$ the term $\bar{\Psi}_{n,j} \rightarrow 1$ as $n \rightarrow \infty$. Therefore, for any $0 < \rho_1 < 1$ and for sufficiently large $n \geq 1$ we obtain that

$$\max_{1 \leq j \leq d} \bar{\Psi}_{n,j} \leq 1 + \rho_1 \quad \text{and} \quad J_n \leq 1 + \rho_1,$$

i.e.

$$\tilde{\mathcal{R}}_0(\hat{\mathbf{h}}) \geq \frac{b-a}{(1+\rho_1)n} \sum_{j=1}^{d_n} \frac{s_j^*}{1+s_j^*} = \frac{b-a}{(1+\rho_1)n} \sum_{j=1}^{d_n} \left(1 - \frac{j^{\mathbf{k}}}{d_n^{\mathbf{k}}}\right).$$

Note here, that

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d \left(1 - \frac{j^{\mathbf{k}}}{d^{\mathbf{k}}}\right) = \int_0^1 (1-t^{\mathbf{k}}) dt = \frac{\mathbf{k}}{\mathbf{k}+1}.$$

Therefore, for sufficiently large n

$$\tilde{\mathcal{R}}_0(\hat{\mathbf{h}}) \geq \frac{(b-a)(1-\rho_1)d_n}{(1+\rho_1)n} \frac{\mathbf{k}}{\mathbf{k}+1}.$$

Therefore, using this bound in (6.8) and through the definition (6.2) and Lemma A.3 and the inequality (6.1) we obtain that for any $0 < \rho, \rho_1 < 1$

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n \in \Xi_n} n^{\frac{2k}{2k+1}} \sup_{S \in \mathcal{W}_{\mathbf{k}, \mathbf{r}}} v(S) \mathcal{R}^*(\hat{S}_n, S) \geq \frac{\rho^{\frac{1}{2k+1}}(1-\rho_1)}{1+\rho_1} \mathbf{l}_*(\mathbf{r}).$$

Taking here limit as $\rho \rightarrow 1$ and $\rho_1 \rightarrow 0$ 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 $\mathcal{W}_{\mathbf{k},\mathbf{r}}$ with known parameters k , r and ς_* defined in (4.4). In this case we use the estimator from family (3.22)

$$\tilde{S} = \widehat{S}_{\tilde{\alpha}} \quad \text{and} \quad \tilde{\alpha} = (\mathbf{k}, \tilde{\mathbf{t}}_n), \quad (7.1)$$

where $\tilde{\mathbf{t}}_n = [\bar{r}(S)/\varepsilon] \varepsilon$, $\bar{r}(S) = \mathbf{r}/\varsigma_*$ and $\varepsilon = 1/\ln n$. Note that for sufficiently large n , the parameter $\tilde{\alpha}$ belongs to the set (3.18). In this section we obtain the upper bound for the empiric risk (3.11).

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

$$\limsup_{n \rightarrow \infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}} v(S) \mathbf{E}_{\mathbf{p},S} \|\tilde{S} - S\|_d^2 \mathbf{1}_\Gamma \leq \mathbf{1}_*(\mathbf{r}). \quad (7.2)$$

Proof. We denote $\tilde{\lambda} = \lambda_{\tilde{\alpha}}$ and $\tilde{\omega} = \omega_{\tilde{\alpha}}$. 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

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

where

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

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

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

Taking into account here the definition (3.6), we can rewrite this inequality as

$$2 \left| \sum_{j=1}^d (1 - \tilde{\lambda}(j)) \tilde{\lambda}(j) \theta_{j,d} \varpi_{j,d} \right| \leq \tilde{\varepsilon} \sum_{j=1}^d (1 - \tilde{\lambda}(j))^2 \theta_{j,d}^2 + \frac{\varpi_n^*}{\tilde{\varepsilon} n}.$$

Therefore,

$$\|\tilde{S} - S\|_d^2 \leq (1 + \tilde{\varepsilon}) \sum_{j=1}^d (1 - \tilde{\lambda}(j))^2 \theta_{j,d}^2 - 2M_n + \frac{\varpi_n^*}{\tilde{\varepsilon} n} + \sum_{j=1}^d \tilde{\lambda}^2(j) \zeta_{j,d}^2.$$

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

$$\sum_{j=1}^d \tilde{\lambda}^2(j) \zeta_{j,d}^2 \leq \frac{(1 + \tilde{\varepsilon})(b - a)}{d} \sum_{j=1}^d \tilde{\lambda}^2(j) \eta_{j,d}^2 + (1 + \tilde{\varepsilon}^{-1}) \frac{\varpi_n^*}{n}.$$

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

$$\|\tilde{S}_n - S\|_d^2 \leq (1 + \tilde{\varepsilon}) \Upsilon_n(S) - 2M_n + (1 + \tilde{\varepsilon}) \mathbf{U}_n + \frac{3\varpi_n^*}{\tilde{\varepsilon} n}, \quad (7.3)$$

where

$$\Upsilon_n(S) = \sum_{j=1}^d (1 - \tilde{\lambda}(j))^2 \theta_{j,d}^2 + \frac{\varsigma_*}{d^2} \sum_{j=1}^d \tilde{\lambda}^2(j) \quad (7.4)$$

and

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

We recall that the variance ς_* is defined in (4.4). In view of Lemma A.8

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

where $\sigma_{1,*}$ is given in (2.31). Moreover, using that $\mathbf{E}_{\mathbf{p},S} M_n = 0$, we get

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

Therefore, Proposition 2.2 yields

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

Now, the property (3.7) Lemma A.7 and Lemma A.9 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. \square

Acknowledgements. This paper is supported by the Normandy RIN projects FuMa. The last author was partially supported by the Normandy RIN project MaSyComB and by the Russian Federal Professor program (project no. 1.472.2016/1.4, the Ministry of Education and Science of the Russian Federation)

A Appendix

A.1 Properties of the prior distribution (6.4)

Lemma A.1. *For any $\mathbf{N} > 2$ there exists a continuously differentiable probability density $\rho_{\mathbf{N}}(\cdot)$ on \mathbb{R} with the support on the interval $[-\mathbf{N}, \mathbf{N}]$, i.e. $\rho_{\mathbf{N}}(z) > 0$ for $-\mathbf{N} < z < \mathbf{N}$ and $\rho_{\mathbf{N}}(z) = 0$ for $|z| \geq \mathbf{N}$, such that for any $\mathbf{N} > 2$ the integral $\int_{-\mathbf{N}}^{\mathbf{N}} z \rho_{\mathbf{N}}(z) dz = 0$ and, moreover,*

$$\int_{-\mathbf{N}}^{\mathbf{N}} z^2 \rho_{\mathbf{N}}(z) dz \rightarrow 1 \quad \text{and} \quad J_{\mathbf{N}} = \int_{-\mathbf{N}}^{\mathbf{N}} \frac{(\dot{\rho}_{\mathbf{N}}(z))^2}{\rho_{\mathbf{N}}(z)} dz \rightarrow 1$$

as $\mathbf{N} \rightarrow \infty$.

Proof. First we set

$$V(z) = \frac{1}{v_*} e^{-\frac{1}{1-z^2}} \mathbf{1}_{\{|z| \leq 1\}} \quad \text{and} \quad v_* = \int_{-1}^1 e^{-\frac{1}{1-t^2}} dt. \quad (\text{A.1})$$

It is clear that this function is infinitely times continuously differentiable, such that $V(z) > 0$ for $|z| < 1$, $V(z) = 0$ for $|z| \geq 1$ and $\int_{-1}^1 V(z) dz = 1$. Using this function, for any $\mathbf{N} \geq 2$ we define

$$\chi_{\mathbf{N}}(z) = \int_{-1}^1 \mathbf{1}_{\{|z+u| \leq \mathbf{N}-1\}} V(u) du = \int_{\mathbb{R}} \mathbf{1}_{\{|t| \leq \mathbf{N}-1\}} V(t-z) dt.$$

Using here the properties of the function (A.1) we can obtain directly that $\chi_{\mathbf{N}}(z) = \chi_{\mathbf{N}}(-z)$ for $z \in \mathbb{R}$, $\chi_{\mathbf{N}}(z) = 1$ for $|z| \leq \mathbf{N} - 2$, $\chi_{\mathbf{N}}(z) > 0$ for $\mathbf{N} - 2 < |z| < \mathbf{N}$ and $\chi_{\mathbf{N}}(z) = 0$ for $|z| \geq \mathbf{N}$. Moreover, it is clear that the derivative

$$\dot{\chi}_{\mathbf{N}}(z) = - \int_{\mathbb{R}} \mathbf{1}_{\{|t| \leq \mathbf{N}-1\}} \dot{V}(t-z) dt = - \int_{-1}^1 \mathbf{1}_{\{|u+z| \leq \mathbf{N}-1\}} \dot{V}(u) du.$$

Note here that

$$|\dot{V}(z)| = \frac{2|z|}{(1-z^2)^2} V(z) \leq \mathbf{c}_* \sqrt{V(z)} \quad \text{and} \quad \mathbf{c}_* = \sup_{|z| \leq 1} \frac{2\sqrt{V(z)}}{(1-z^2)^2}.$$

Now through the Bunyakovsky - Cauchy - Schwartz inequality we get that

$$\dot{\chi}_{\mathbf{N}}^2(z) \leq 2\mathbf{c}_* \chi_{\mathbf{N}}(z) \quad \text{for} \quad |z| < \mathbf{N}. \quad (\text{A.2})$$

Now, denoting by $\varphi(z)$ the $(0, 1)$ Gaussian density. we set

$$\rho_{\mathbf{N}}(z) = \frac{\varphi(z)\chi_{\mathbf{N}}(z)}{\rho_{\mathbf{N}}^*} \quad \text{and} \quad \rho_{\mathbf{N}}^* = \int_{-\mathbf{N}}^{\mathbf{N}} \varphi(t)\chi_{\mathbf{N}}(t)dt.$$

It is clear that $\rho_{\mathbf{N}}(z)$ is the continuously differentiable probability density with the support $[-\mathbf{N}, \mathbf{N}]$ such that for any \mathbf{N} the integral $\int_{-\mathbf{N}}^{\mathbf{N}} z\rho_{\mathbf{N}}(z)dz = 0$ and for $\mathbf{N} \rightarrow \infty$

$$\rho_{\mathbf{N}}^* \rightarrow 1 \quad \text{and} \quad \int_{-\mathbf{N}}^{\mathbf{N}} z^2 \rho_{\mathbf{N}}(z)dz \rightarrow 1.$$

Moreover, note that

$$J_{\mathbf{N}} = \frac{1}{\rho_{\mathbf{N}}^*} \int_{-\mathbf{N}}^{\mathbf{N}} \frac{\dot{\varphi}^2(t)}{\varphi(t)} dt + \frac{1}{\rho_{\mathbf{N}}^*} \Delta_{\mathbf{N}},$$

where taking into account that the derivative $\dot{\chi}_{\mathbf{N}}(z) = 0$ for $|z| \leq \mathbf{N} - 2$, the last term can be represented as

$$\Delta_{\mathbf{N}} = 2 \int_{\mathbf{N}-2 \leq |z| \leq \mathbf{N}} \dot{\varphi}(z)\dot{\chi}_{\mathbf{N}}(z)dz + \int_{\mathbf{N}-2 \leq |z| \leq \mathbf{N}} \varphi(z) \frac{\dot{\chi}_{\mathbf{N}}^2(z)}{\chi_{\mathbf{N}}(z)} dz.$$

Using here the bound (A.2), it is easy to conclude that $\Delta_{\mathbf{N}} \rightarrow 0$ as $\mathbf{N} \rightarrow \infty$. Hence Lemma A.1. \square

Lemma A.2. For any $1 \leq j \leq d$

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n-1} |\bar{\Psi}_{n,j} - 1| = 0, \quad (\text{A.3})$$

where the term $\bar{\Psi}_{n,j}$ is given in (6.9).

Proof. First, note that for any $l \geq 1$

$$y_l = y_0 \prod_{i=1}^l S_\kappa(x_i) + \sum_{\iota=1}^l \prod_{i=\iota+1}^l S_\kappa(x_i) \xi_\iota.$$

Therefore,

$$\tilde{\mathbf{E}}_0 y_l^2 = y_0^2 \tilde{\mathbf{E}}_0 \prod_{i=1}^l S_\kappa^2(x_i) + \sum_{\iota=1}^{l-1} \tilde{\mathbf{E}}_0 \prod_{i=\iota+1}^l S_\kappa^2(x_i) + 1$$

and due to (6.5) we obtain that for any $n \geq 1$ for which $\delta_n^* < 1$

$$\sup_{l \geq 1} |\tilde{\mathbf{E}}_0 y_l^2 - 1| \leq (\delta_n^*)^2 \frac{(y_0^2 + 1)}{1 - (\delta_n^*)^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Taking into account that for any $n \geq 3$ and $1 \leq j \leq n-1$

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

we get (A.3). Hence Lemma A.2. \square

Lemma A.3. For any $\mathbf{b} > 0$ and $0 < \rho < 1$

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \mathbf{R}_{0,n} = 0. \tag{A.4}$$

where the term $\mathbf{R}_{0,n}$ is introduced in (6.8).

Proof. First note, that from the definition of term $\mathbf{R}_{0,n}$ in (6.8) through the Bunyakovsky - Cauchy - Schwartz inequality we get that

$$\mathbf{R}_{0,n} \leq \left(\mathbf{r} + \sqrt{3} \sum_{j=1}^d s_j^2 \right) \mu_\kappa(\mathbf{D}_n^c).$$

Therefore, to show (A.4) it suffices to check that for any $\mathbf{b} > 0$

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \mu_\kappa(\mathbf{D}_n^c) = 0. \tag{A.5}$$

To this end note, that from the definition of \mathbf{D}_n in (6.6) we obtain that

$$\mu_\kappa(\mathbf{D}_n^c) \leq \mathbf{P}(\zeta_n > \mathbf{r}),$$

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

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \mathbf{P}(\zeta_n > \mathbf{r}) = 0. \quad (\text{A.6})$$

Indeed, first note, that the definitions (6.2) and (6.4) imply directly

$$\lim_{n \rightarrow \infty} \mathbf{E} \zeta_n = \lim_{n \rightarrow \infty} \sum_{j=1}^d \mathbf{a}_j s_j^2 \mathbf{E} (\eta_1^*)^2 = \lim_{n \rightarrow \infty} \sum_{j=1}^d \mathbf{a}_j s_j^2 = \rho \mathbf{r}.$$

We set now

$$\tilde{\zeta}_n = \zeta_n - \mathbf{E} \zeta_n = \frac{b-a}{n} \sum_{j=1}^{d_n} s_j^* \mathbf{a}_j \tilde{\eta}_j \quad \text{and} \quad \tilde{\eta}_j = (\eta_j^*)^2 - \mathbf{E}(\eta_j^*)^2.$$

So, for sufficiently large n we obtain that for $\mathbf{r}_1 = \mathbf{r}(1 - \rho)/2$

$$\{\zeta_n > \mathbf{r}\} \subset \{\tilde{\zeta}_n > \mathbf{r}_1\}.$$

Through the correlation inequality from [17] and taking into account that $|\tilde{\eta}_j| \leq 2 \ln^2 n$, we obtain that for any $p \geq 2$ there exists some constant $C_p > 0$ for which

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

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

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

This implies (A.6), and therefore, Lemma A.3. \square

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

Lemma A.4. *Let f be an absolutely continuous $[0, 1] \rightarrow \mathbb{R}$ function with $\|f\| < \infty$ and g be a simple $[0, 1] \rightarrow \mathbb{R}$ 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 $\tilde{\varepsilon} > 0$

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

and

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

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

$$\begin{aligned} \|\Delta\|^2 &= \|\Delta\|_d^2 + \sum_{l=1}^d \int_{t_{l-1}}^{t_l} (2\Delta(t_l) (\Delta(t) - \Delta(t_l)) + (\Delta(t) - \Delta(t_l))^2) dt \\ &\leq (1 + \tilde{\varepsilon})\|\Delta\|_d^2 + (1 + \tilde{\varepsilon}^{-1}) \sum_{l=1}^d \int_{t_{l-1}}^{t_l} [\Delta(t_l) - (\Delta(t))]^2 dt \\ &= (1 + \tilde{\varepsilon})\|\Delta\|_d^2 + (1 + \tilde{\varepsilon}^{-1}) \sum_{l=1}^d \int_{t_{l-1}}^{t_l} |f(t_l) - f(t)|^2 dt. \end{aligned}$$

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

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

one comes to the first inequality. Similarly, one can verify the second inequality. Hence Lemma A.4. \square

A.3 Properties of the trigonometric basis.

Lemma A.5. For any $1 \leq j \leq d$ the trigonometric Fourier coefficients $(\theta_{j,d})_{1 \leq j \leq p}$ for the functions S from the class $\mathcal{W}_{\mathbf{k},\mathbf{r}}$ with $\mathbf{k} \geq 1$ satisfy, for any $\tilde{\varepsilon} > 0$, the following inequality

$$\theta_{j,d}^2 \leq (1 + \tilde{\varepsilon})\theta_j^2 + (1 + \tilde{\varepsilon}^{-1})\frac{2\mathbf{r}}{d^{2k}}. \quad (\text{A.7})$$

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 $0 < \tilde{\varepsilon} < 1$

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

By applying Lemma A.4 with $g = 0$, we obtain that

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

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

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

Taking into account here that

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

we obtain that

$$\|\dot{\Delta}_N\|^2 \leq \sum_{l>N} \frac{\mathbf{a}_l}{l^{2(k-1)}} \theta_l^2 \leq \frac{\mathbf{r}}{N^{2(k-1)}}. \quad (\text{A.8})$$

Hence Lemma A.5 \square

Lemma A.6. *For any $d \geq 2$ and $1 \leq N \leq d$, the coefficients $(\theta_{j,d})_{1 \leq j \leq d}$ of functions S from $W_{\mathbf{k},r}$ with $\mathbf{k} \geq 1$ satisfy, for any $\tilde{\varepsilon} > 0$, the following inequality*

$$\sum_{j=N}^d \theta_{j,d}^2 \leq (1 + \tilde{\varepsilon}) \sum_{j \geq N} \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) \frac{r}{d^2 N^{2(k-1)}}. \quad (\text{A.9})$$

Proof. First we note that

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

where $\Delta_N(t) = \sum_{j \geq N} \theta_j \phi_j(t)$. By applying Lemma A.4 and the inequality (A.8), we obtain the bound (A.9). Hence Lemma A.6 \square

A.4 Technical lemmas

Lemma A.7. *The sequence $\Upsilon_n(S)$ satisfies the following upper bound*

$$\limsup_{n \rightarrow \infty} \sup_{S \in \mathcal{W}_{\mathbf{k}, \mathbf{r}}} n^{2\mathbf{k}/(2\mathbf{k}+1)} v(S) \Upsilon_n(S) \leq \mathbf{1}_*(\mathbf{r}). \quad (\text{A.10})$$

Proof. First of all, note that

$$0 < \epsilon^2(b-a) \leq \inf_{S \in \Theta_{\epsilon, L}} \varsigma_* \leq \sup_{S \in \Theta_{\epsilon, L}} \varsigma_* \leq b-a. \quad (\text{A.11})$$

This implies directly that

$$\lim_{n \rightarrow \infty} \sup_{S \in \Theta_{\epsilon, L}} \left| \frac{\tilde{\mathbf{t}}_n}{\bar{r}(S)} - 1 \right| = 0, \quad (\text{A.12})$$

where $\tilde{\mathbf{t}}_n = [\bar{r}(S)/\epsilon] \epsilon$ and $\bar{r}(S) = \mathbf{r}/\varsigma_*$. Moreover, note that

$$n^{2\mathbf{k}/(2\mathbf{k}+1)} v(S) \Upsilon_n(S) \leq n^{2\mathbf{k}/(2\mathbf{k}+1)} v(S) \mathbf{G}_n + \frac{(\varsigma_*)^{1/(2\mathbf{k}+1)}}{n^{1/(2\mathbf{k}+1)}} \sum_{j=1}^d \tilde{\lambda}^2(j)$$

and

$$\mathbf{G}_n = \sum_{j=1}^d (1 - \tilde{\lambda}(j))^2 \theta_{j,d}^2 = \mathbf{G}_{1,n} + \mathbf{G}_{2,n},$$

where

$$\mathbf{G}_{1,n} = \sum_{j=j_*}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_{j,d}^2 \quad \text{and} \quad \mathbf{G}_{2,n} = \sum_{j=[\tilde{\omega}]+1}^d \theta_{j,d}^2.$$

We recall that

$$\tilde{\omega} = \omega_{\tilde{\alpha}} = \left(\varrho_{\mathbf{k}} \tilde{\mathbf{t}}_n n \right)^{1/(2\mathbf{k}+1)} \quad \text{and} \quad \varrho_{\mathbf{k}} = \frac{(\mathbf{k}+1)(2\mathbf{k}+1)}{\pi^{2\mathbf{k}} \mathbf{k}}.$$

Note now, that Lemma A.5 and Lemma A.6 yield

$$\mathbf{G}_{1,n} \leq (1 + \tilde{\epsilon}) \sum_{j=j_*}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_j^2 + 2\mathbf{r}(1 + \tilde{\epsilon}^{-1}) \frac{\tilde{\omega}}{d^{2\mathbf{k}}}.$$

and

$$\mathbf{G}_{2,n} \leq (1 + \tilde{\epsilon}) \sum_{j \geq [\tilde{\omega}]+1} \theta_j^2 + (1 + \tilde{\epsilon}^{-1}) \frac{\mathbf{r}}{d^2 \tilde{\omega}^{2(k-1)}},$$

i.e.

$$\mathbf{G}_n \leq (1 + \tilde{\varepsilon})\mathbf{G}_n^* + 2\mathbf{r}(1 + \tilde{\varepsilon}^{-1})\tilde{\mathbf{G}}_n, \quad (\text{A.13})$$

where

$$\mathbf{G}_n^* = \sum_{j \geq 1} (1 - \tilde{\lambda}(j))^2 \theta_j^2 = \sum_{j \leq \tilde{\omega}} (1 - \tilde{\lambda}(j))^2 \theta_j^2 + \sum_{j > \tilde{\omega}} \theta_j^2 := \mathbf{G}_{1,n}^* + \mathbf{G}_{2,n}^*$$

and

$$\tilde{\mathbf{G}}_n = \frac{\tilde{\omega}}{d^{2k}} + \frac{1}{d^{2k} \tilde{\omega}^{2(k-1)}}.$$

Note, that

$$\begin{aligned} n^{2\mathbf{k}/(2\mathbf{k}+1)} v(S) \mathbf{G}_{1,n}^* &= \sum_{j=j_*}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_j^2 = \frac{v(S)}{(\varrho_{\mathbf{k}} \tilde{\mathbf{I}}_n)^{2\mathbf{k}/(2\mathbf{k}+1)}} \sum_{j=j_*}^{[\tilde{\omega}]} j^2 \theta_j^2 \\ &\leq \frac{v(S)}{(\varrho_{\mathbf{k}} \tilde{\mathbf{I}}_n)^{2\mathbf{k}/(2\mathbf{k}+1)}} \mathbf{a}_{j_*}^* \sum_{j=j_*}^{[\tilde{\omega}]} \mathbf{a}_j \theta_j^2, \end{aligned}$$

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

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

Therefore, from (A.12) we obtain that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{S \in \Theta_{\varepsilon, L}} \frac{n^{2\mathbf{k}/(2\mathbf{k}+1)} v(S) \mathbf{G}_{1,n}^*}{\sum_{j=j_*}^{[\tilde{\omega}]} \mathbf{a}_j \theta_j^2} \\ \leq \frac{1}{\pi^{2k} (\varrho_{\mathbf{k}} r)^{2\mathbf{k}/(2\mathbf{k}+1)}} = \frac{(\varrho_{\mathbf{k}}^*)^{2\mathbf{k}/(2\mathbf{k}+1)}}{(2k\pi r)^{2\mathbf{k}/(2\mathbf{k}+1)}}, \end{aligned}$$

where

$$\varrho_{\mathbf{k}}^* = \frac{2\mathbf{k}^2}{(\mathbf{k}+1)(2\mathbf{k}+1)} = \int_0^1 (1-t^{\mathbf{k}})^2 dt.$$

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

$$\begin{aligned} \mathbf{G}_{2,n}^* &\leq \frac{1}{\pi^{2k} \tilde{\omega}^{2k}} \sum_{j \geq [\tilde{\omega}] + 1} \mathbf{a}_j \theta_j^2 \\ &\leq (1 + \tilde{\varepsilon}) \frac{(\varsigma_* \varrho_{\mathbf{k}}^*)^{2\mathbf{k}/(2\mathbf{k}+1)}}{\pi^{2\mathbf{k}/(2\mathbf{k}+1)} (2krn)^{2\mathbf{k}/(2\mathbf{k}+1)}} \sum_{j \geq [\tilde{\omega}] + 1} \mathbf{a}_j \theta_j^2, \end{aligned}$$

i.e.

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

Moreover, we get directly that

$$\lim_{T \rightarrow \infty} \sup_{S \in \Theta_{\varepsilon, L}} \left| \frac{\sum_{j=1}^d \tilde{\lambda}^2(j)}{n^{1/(2\mathbf{k}+1)} (\bar{r}(S))^{1/(2\mathbf{k}+1)} \mathbf{g}_k} - 1 \right| = 0, \quad (\text{A.14})$$

where

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

So, taking into account that in (A.13)

$$\lim_{n \rightarrow \infty} \sup_{S \in \mathcal{W}_{\mathbf{k}, r}} n^{2\mathbf{k}/(2\mathbf{k}+1)} \tilde{\mathbf{G}}_n = 0,$$

we obtain (A.10) and, hence Lemma A.7. \square

Lemma A.8. *For any non random coefficients $(\mathbf{u}_j)_{1 \leq j \leq d}$*

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

Proof. Using the definition of $\eta_{j,d}$ in (3.5), we obtain that

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

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

Lemma A.9. *Now we show that*

$$\lim_{n \rightarrow \infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \mathcal{W}_{\mathbf{k}, r}} |\mathbf{E}_{\mathbf{p}, S} \mathbf{U}_n \mathbf{1}_\Gamma| = 0. \quad (\text{A.15})$$

Proof. First of all, note that, using the definition of $\mathbf{s}_{j,d}$ in (3.13), we obtain

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

where

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

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

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

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

$$\mathbf{U}_{1,n} = \frac{b-a}{d} \sum_{l=1}^d \frac{d}{H_l} - \varsigma_* \quad \text{and} \quad \mathbf{U}_{2,n} = \sum_{j=1}^d \tilde{\lambda}^2(j) \bar{\mathbf{s}}_{j,d}.$$

Note now, that using Proposition 2.20 and the dominated convergence theorem in the definition (2.17) we obtain that

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

Taking into account that for the functions from the class (2.4) their derivatives are uniformly bounded, we can deduce that

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

i.e.

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

Therefore, taking into account that

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

we obtain that

$$\overline{\lim}_{n \rightarrow \infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S \in \Theta_{\epsilon,L}} \sup_{\mathbf{p} \in \mathcal{P}} |\mathbf{E}_{\mathbf{p},S} \mathbf{U}_n| \leq (b-a) \overline{\lim}_{n \rightarrow \infty} \mathbf{U}_{2,n}^*, \quad (\text{A.16})$$

where

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

Now, using Lemma A.2 from [14] we obtain that

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

Note that from the definition of $\sigma_{1,*}$ in (2.31) it follows that

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

i.e.

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

Therefore, the using this bound in (A.16) implies

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

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

$$\mathbf{E}_{\mathbf{p},S} \eta_{j,d}^4 \leq 64 \check{\mathbf{m}} \sigma_{1,*}^2,$$

where the coefficient $\check{\mathbf{m}}$ is given in (2.30). From this we obtain, that

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

So, Proposition 2.2 implies

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

This implies the quality (A.15), and we obtain Lemma A.9. \square

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*, **38** (4), 437 - 460.
- [5] Bayisa, F.L., Zhou, Z., Cronie, O. and Yu, J. (2019) Adaptive algorithm for sparse signal recovery. *Digital Signal Processing*, **87**, 10 – 18.
- [6] Belitser, E. (2000) : Recursive Estimation of a Drifted Autoregressive Parameter, *The Annals of Statistics* 26: 860-870.
- [7] Beltaief, S., Chernoyarov, O. and Pergamenchtchikov, S. (2019) Model selection for the robust efficient signal processing observed with small Levy noise. - *Ann Inst Stat Math.*, <https://doi.org/10.1007/s10463-019-00726-2>
- [8] Candés E, Tao, T (2007) The Dantzig selector: statistical estimation when p is much larger than n . - *Ann. Statist*, **35** (6), p. 2313 – 2351.
- [9] Dahlhaus, R. (1996) : Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes, *Journal of Nonparametric Statistics* 6: 171-191.
- [10] Dahlhaus, R. (1996) : On the Kullback-Leibler Information Divergence of Locally Stationary Processes, *Stochastic Processes and their Applications* 62: 139-168.
- [11] De Gregorio, A., Iacus, SM (2012) Adaptive LASSO-type estimation for multivariate diffusion processes. *Econom. Theory*, **28** (4), p. 838 – 860.

- [12] Fujimori, K. (2019) The Danzing selector for a linear model of diffusion processes. - *Statistical Inference for Stochastic Processes*, **22**, p. 475 – 498.
- [13] Galtchouk, L. and Pergamenshchikov, S. (2006) : Asymptotically Efficient Estimates for Nonparametric Regression Models, *Statistics and Probability Letters* 76 : 852-860.
- [14] Galtchouk, L.I. and Pergamenshchikov, S.M. (2009) Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21**, 1-16.
- [15] 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.
- [16] Galtchouk, L. and Pergamenshchikov, S. (2011) : Adaptive Sequential Estimation for Ergodic Diffusion Processes in Quadratic Metric, *Journal of Nonparametric Statistics* 23: 255-285.
- [17] 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.
- [18] Galtchouk, L. and Pergamenshchikov, S. (2019) Non asymptotic sharp oracle inequalities for high dimensional ergodic diffusion models. - Preprint, 2019, <https://hal.archives-ouvertes.fr/hal-02387034>
- [19] Galtchouk, L. and Pergamenshchikov, S. (2020) Efficient big data analysis for ergodic diffusion models on the basis of discrete data. - Preprint, 2020, <https://hal.archives-ouvertes.fr/hal-02474675>
- [20] 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.
- [21] 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.
- [22] Konev, V.V. and Pergamenshchikov, S. M. Non-parametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Tomsk State University Journal of Mathematics and Mechanics* , 2009, 4(8), 31-45.

- [23] 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.
- [24] 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.
- [25] Moulines, E., Priouret, P. and Roueff, F. (2005) On Recursive Estimation for Time Varying Autoregressive Processes, *The Annals of Statistics* **33**, 2610 – 2654.
- [26] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L2. - *Annals of Statistics*, **13**, 984 – 997.
- [27] Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. - *J. R. Statist. Soc.*, **Ser B 58** (1), p. 267 – 288.
- [28] Pinsker, M.S. (1981) Optimal filtration of square integrable signals in gaussian noise. - *Problems Inform. Transmission*, **17**, 120-133.