Adaptive 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. Adaptive efficient robust sequential analysis for autoregressive big data models. 2022. hal-03890891

HAL Id: hal-03890891
https://hal.science/hal-03890891v1

Preprint submitted on 8 Dec 2022

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.

Adaptive efficient robust sequential analysis for autoregressive big data models *

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

December 8, 2022

Abstract

In this paper we consider high dimension models based on dependent observations defined through autoregressive processes. For such models we develop an adaptive efficient estimation method via the robust sequential model selection procedures. To this end, we first obtain a van Trees inequality for such models, and then, using this inequality, we probably for the first time obtain a sharp lower bound for the weighted robust risk in an explicit form given by the famous Pinsker constant. Moreover, in getting this lower bound we have found the nonparametric version of the Fisher information for this model. Then, using the weighted least square method and sharp non asymptotic oracle inequalities from (Arkoun O., Brua J.-Y., and Pergamenchtchikov S. 2019. Sequential Analysis 38(4): 437-460), we develop analytic tools to provide the efficiency property in the minimax sense for the proposed estimation procedure, i.e. we show that the upper bound for its risk coincides with the obtained lower bound. It should be emphasized that this property is obtained without using sparse conditions and in the adaptive setting when the parameter dimension and model regularity are unknown. We then study the constructed procedures numerically using Monte Carlo simulations.

MSC: primary 62G08, secondary 62G05
Keywords: Autoregressive big data models; Robust estimation; Sequential estimation; Model selection; Sharp oracle inequality; Pinsker constant; Asymptotic efficiency.

*This research was supported by RSF, project no 22-21-00302 (National Research Tomsk State University, Russia).

†Sup'Biotech, Laboratoire BIRL, 66 Rue Guy Moquet, 94800 Villejuif, France and Univ Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, France, e-mail: ouerdia.arkoun@supbiotech.fr

‡Univ Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, France, e-mail : jean-yves.brua@univ-rouen.fr

§Univ Rouen Normandie, CNRS, LMRS UMR 6085, F-76000 Rouen, 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

We study the observations model defined for $1 \leq j \leq n$ through the following difference equation

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

where the initial value y_0 is a non random known constant, $(\psi_i)_{i\geq 1}$ are known linearly independent functions, $a < b$ are fixed known constants and $(\xi_j)_{j\geq 1}$ are i.i.d. unobservable random variables with an unknown density distribution $\mathbf{p}$ from some functional class which will be specified later.

The problem is to estimate the unknown parameters $(\beta_i)_{1\leq i\leq q}$ in the high dimension setting, i.e. when the number of parameters is greater than the number of observations, i.e. $q > n$. It should be noted that, usually, big data models in discrete time are considered only for the i.i.d. observations (see, for example, [Hastie et al.(2008)] and the references therein). In this paper, we study such model in the dependent observations framework defined by the autoregressive model (1.1). It should also be emphasized, that in all the papers devoted to big data models in discrete time, it is assumed that the number of parameters q is known and, moreover, to provide optimality properties some sparse conditions are assumed. This means that the usual methods (Lasso algorithms or Dantzig selector) cannot be used to estimate, for example, the number of parameters q. To overcome these limitations, in this paper, similarly to the approach proposed in [Galtchouk and Pergamenshchikov (2022)], we study this problem in a nonparametric setting, i.e., we include observations (1.1) in the general model defined by

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

where $S(\cdot) \in \mathcal{L}_2[a,b]$ is unknown function. The nonparametric setting allows to consider the models (1.1) with unknown q or even 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, [Beltaief et al. (2020), Bayisa et al. (2019)]). Thus the problem is now to estimate the function $S(\cdot)$ on the basis of the observations (1.2) under the condition that the noise distribution $\mathbf{p}$ is unknown and belongs to some functional class $\mathcal{P}$. Note that this is well known as time varying autoregressive model (TVAR) (see, for example, [Niedzwiecki (2000)] and the references therein). Such models are very popular in various important fields such as signal and image processing [De Souza and al. (2019)], econometrics [Orbe and al. (2006)], farming [Li and al. (2022)], etc. There is a number of papers that consider these models (see, for example, [Belitser (2000), Fan and Zhang (2008), Luo et al. (2009)] and the references therein). Firstly, minimax estimation problems for the model (1.2) has been treated in [Arkoun and Pergamenchtchikov (2008), Moulines et al. (2005)] in the nonadaptive case, i.e. for the known regularity of the function S. It should be noted that in [Moulines et al. (2005)], the authors found the optimal convergence rate for the Hölder multidimensional functional coefficients, while in [Arkoun and Pergamenchtchikov (2008)], along with the optimal rate, the efficiency of a kernel estimator was also established, that is, a constant determining the sharp lower bound for the pointwise risk and a kernel estimator whose risk coincides with this constant were found. Remark also that, in the parametric case, this constant is the Fisher

information. Then, in [Arkoun (2011), Arkoun and Pergamenchtchikov (2016)], the authors proposed to use the sequential analysis for the adaptive pointwise estimation problem, i.e. in the case when the Hölder regularity is unknown. Moreover, it turned out that only the sequential methods can provide the adaptive estimation for autoregressive models. For these reasons we use in this paper the sequential analysis approach. Moreover, it should be noted that in all these papers the estimation problems are considered only in the pointwise setting when one estimates the functional parameters in fixed points. Unfortunately, these methods do not provide an effective estimation of functions with respect to the integral risk. To develop such methods for the model (1.2), we use in this paper the adaptive sequential model selection procedures from [Arkoun et al. (2019)], which are optimal in the sense of sharp oracle inequalties for the quadratic risk defined by

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

where $\widehat{S}_n$ is an estimator of S based on the observations $(y_j)_{1\le j\le n}$ and $\mathbf{E}_{\mathbf{p},S}$ is the expectation with respect to the distribution law $\mathbf{P}_{\mathbf{p},S}$ of the process 1.2 given the distribution density $\mathbf{p}$ and the function S. Moreover, taking into account that $\mathbf{p}$ is unknown, we use the robust risk defined by

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

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

1.2 Main tools

In this paper, in order to provide an efficient estimation, we develop special efficient estimation methods for the model (1.2), as it is done in [Galtchouk and Pergamenshchikov (2022)] for the diffusion processes. It should be emphasized that these methods cannot be used here, since they are completely based on the uniform geometric ergodicity property. Note that the process (1.2) is not ergodic even in the asymptotic setting. Therefore, to provide an efficient estimation in this case, one needs to develop a new asymptotic estimation method for the model (1.2). To estimate the function S in this model we make use of the model selection procedures proposed in [Arkoun et al. (2019)]. These procedures are based on the optimal pointwise truncated sequential estimators from [Arkoun and Pergamenchtchikov (2016)] for which through the method developed in [Galtchouk and Pergamenshchikov (2009a)] sharp oracle inequalities are obtained. Based on these inequalities, our goal is to show that the estimation procedures constructed in this paper are efficient in an adaptive formulation for the robust risk (1.3). To do this, first of all, it is necessary to establish a sharp lower bound for this risk, i.e. to provide the best possible estimation accuracy for the model (1.2) which is usually given by the Pinsker constant for quadratic risks (see, for example, in [Pinsker (1981)], [Galtchouk and Pergamenshchikov (2011)], [Pchelintsev et al. (2021)]). The first step consists in obtaining the van Trees inequality. Secondly, to apply this inequality to autoregressive models one needs to choose prior distributions. It should be emphasized here that we cannot apply the usual analytic methods based on the Gaussian prior distributions, since in this case, the observation process is not homogenous, and there is no stationarity as well. Here one needs to choose prior functional distributions of special form with bounded supports.

Then, through the van Trees inequality, we obtain a lower bound for the quadratic risk and to get the sharp bound, we maximize it over all the chosen priori distributions. It turns out that for the model (1.2) the sharp lower bound for the weighted quadratic risk is the Pinsker constant obtained in [Pinsker (1981)] for the signal filtration problem. Moreover, it appears that the weight coefficient in the quadratic risk is the nonparametric Fisher information for the model (1.2). This is the new result in the efficient nonparametric estimation for dependent observations. Eventually, using the oracle inequality from [Arkoun et al. (2019)] and the weighted least square estimation method, we show that for our constructed model selection procedure, the upper bound of its risk asymptotically coincides with the obtained lower bound without using the regularity properties of the unknown function. This means that it is efficient in the adaptive setting with respect to the robust risk (1.3).

1.3 Organization of the paper

The paper is organized as follows. In Section 2, we construct the sequential pointwise estimation procedure which allows us to pass from the autoregression model to the corresponding regression model, then in Section 3 we develop the model selection method. We announce the main results in Section 4. In Section 5 we present Monte-Carlo results which numerically illustrate the behavior of the proposed estimation procedure. In Section 6 we show the van Trees inequality for the model (1.2). We get the lower bound for the robust risk in Section 7 and in Section 8 we obtain the upper bound for the corresponding robust risk. In Conclusion, we summarize all the main results of the paper and compare them with the existing ones. We provide all the auxiliary tools in Appendix.

2 Sequential pointwise estimation method

To estimate the function S in the model (1.2) on the interval $[a,b]$, we use the kernel sequential estimator proposed in [Arkoun and Pergamenchtchikov (2016)] at the points $(z_l)_{1\le l\le d}$ defined by

$$z_l = a + \frac{l}{d}(b-a)\,, \tag{2.1}$$

where d is an integer valued function of n, i.e. $d=d_n$, such that $d/\sqrt{n}\to 1$ as $n\to\infty$. Note that in this case the kernel estimator has the following form

$$\check{S}(z_l) = \frac{\sum_{j=1}^{n} Q_{l,j}\, y_{j-1}\, y_j}{\sum_{j=1}^{n} Q_{l,j}\, y_{j-1}^2} \quad \text{and} \quad Q_{l,j} = Q\left(\frac{x_j - z_l}{h}\right)\,,$$

where $Q(\cdot)$ is a kernel function and h is a bandwidth.
As it is shown in [Arkoun and Pergamenchtchikov (2008)] to provide an efficient pointwise estimation, the kernel function must be chosen as the indicator of the interval $]-1;1]$, i.e. $Q(u)=\mathbf{1}_{]-1,1]}(u)$. This means that we can rewrite the estimator (2) as

$$\check{S}(z_l) = \frac{\sum_{j=k_{1,l}}^{k_{2,l}} y_{j-1}\, y_j}{\sum_{j=k_{1,l}}^{k_{2,l}} y_{j-1}^2}\,, \tag{2.2}$$

where $k_{1,l} = [n\widetilde{z}_l - n\widetilde{h}] + 1$ and $k_{2,l} = [n\widetilde{z}_l + n\widetilde{h}] \wedge n$, $[x]$ is the integer part of x, $\widetilde{z}_l = l/d$ and $\widetilde{h} = h/(b-a)$. In order to use the model selection method from [Arkoun et al. (2019)] we need to obtain the uncorrelated stochastic terms in the kernel estimator of the function S at the points (2.1), i.e. one needs to use the disjoint observations sets $(y_j)_{k_{1,l} \le j \le k_{2,l}}$. To this end it suffices to choose h for which for all $2 \le l \le d$ the bounds $k_{2,l-1} < k_{1,l}$, i.e. we set

$$h = \frac{b-a}{2d} \quad \text{and} \quad \widetilde{h} = \frac{1}{2d}\,. \tag{2.3}$$

Note that the main difficulty is that the kernel estimator is a non linear function of the observations due to the random denominator. To control this denominator we need to assume conditions providing the stability properties for the model (1.2). To obtain the stability (uniformly with respect to the function S) of the 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 [Arkoun and Pergamenchtchikov (2016)] as

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

where $\mathbf{C}_1[a,b]$ is the Banach space of continuously differentiable $[a,b] \to \mathbb{R}$ functions and $|S|_* = \sup_{a \le x \le b} |S(x)|$. As it is shown in [Arkoun et al. (2019)] $\forall S \in \Theta_{\varepsilon,L}$

$$\sum_{j=k_{1,l}}^{k_{2,l}} y_{j-1}^2 \approx \frac{k_{2,l}-k_{1,l}}{1-S^2(z_l)} \quad \text{as} \quad k_{2,l}-k_{1,l} \to \infty\,.$$

Therefore, to replace the denominator in (2.2) with its limit we first need a preliminary estimate of the function $S(z_l)$. We estimate it as

$$\widehat{S}_l = \frac{\sum_{j=k_{1,l}}^{\iota_l} y_{j-1}\, y_j}{\sum_{j=k_{1,l}}^{\iota_l} y_{j-1}^2} \quad \text{with} \quad \iota_l = k_{1,l} + \mathbf{q}\,, \tag{2.5}$$

where $\mathbf{q} = \mathbf{q}_n = [(n\widetilde{h})^{\mu_0}]$ for some $0 < \mu_0 < 1$. Indeed, we cannot use this estimator directly to replace the random denominator since in general it can be close to one. For this reason we use its projection onto the interval $]-1+\widetilde{\varepsilon}, 1-\widetilde{\varepsilon}[$, i.e. $\widetilde{S}_l = \min(\max(\widehat{S}_l, -1+\widetilde{\varepsilon}), 1-\widetilde{\varepsilon})$ and $\widetilde{\epsilon} = (2+\ln n)^{-1}$. Finally, omitting some technical details, we will replace the denominator (2.2) with the threshold H_l defined by

$$H_l = \frac{1-\widetilde{\epsilon}}{1-\widetilde{S}_l^2}(k_{2,l} - \iota_l)\,. \tag{2.6}$$

It should be noted that H_l is a function the observations $y_{k_{1,l}}, \ldots, y_{\iota_l}$. To replace the random denominator in (2.2) with H_l, we use the sequential estimation method through the following stopping time

$$\tau_l = \inf\{k > \iota_l \,:\, \sum_{j=\iota_l+1}^{k} \mathbf{u}_{j,l} \ge H_l\}\,, \tag{2.7}$$

where $\mathbf{u}_{j,l} = y_{j-1}^2 \mathbf{1}_{\{\iota_l+1\le j<k_{2,l}\}} + H_l \mathbf{1}_{\{j=k_{2,l}\}}$. It is clear that $\tau_l \le k_{2,l}$ a.s. Now we define the sequential estimator by

$$S_l^* = \frac{1}{H_l}\left(\sum_{j=\iota_l+1}^{\tau_l-1} y_{j-1}\, y_j \;+\; \varkappa_l\, y_{\tau_l-1}\, y_{\tau_l}\right)\mathbf{1}_{\Gamma_l}\,, \tag{2.8}$$

where $\Gamma_l = \{\tau_l < k_{2,l}\}$ and the correcting coefficient $0 < \varkappa_l \le 1$ is given by

$$\sum_{j=\iota_l+1}^{\tau_l-1} \mathbf{u}_{j,l} + \varkappa_l^2 \mathbf{u}_{\tau_l,l} = H_l\,. \tag{2.9}$$

As in [Arkoun and Pergamenchtchikov (2016)], we study robust properties of this sequential procedure by assuming that in the model (1.2) the i.i.d. random variables $(\xi_j)_{j\ge1}$ have a density $\mathbf{p}$ (with respect to the Lebesgue measure) from the functional class $\mathcal{P}$ defined by

$$\mathcal{P} := \left\{ \mathbf{p}\ge 0 \,:\, \int_{-\infty}^{+\infty} \mathbf{p}(x)\,\mathrm{d}x = 1\,,\quad \int_{-\infty}^{+\infty} x\,\mathbf{p}(x)\,\mathrm{d}x = 0\,,\right.$$
$$\left. \int_{-\infty}^{+\infty} x^2\,\mathbf{p}(x)\,\mathrm{d}x = 1 \quad\text{and}\quad \sup_{l\ge1}\frac{\int_{-\infty}^{+\infty}|x|^{2l}\,\mathbf{p}(x)\,\mathrm{d}x}{l!\,\varsigma^l} \le 1 \right\}, \tag{2.10}$$

where $\varsigma \ge 1$ is some fixed parameter which may be a *slowly increasing function* of the number observation n, i.e. for any $\mathbf{b} > 0$

$$\lim_{n\to\infty} n^{-\mathbf{b}}\,\varsigma(n) = 0\,. \tag{2.11}$$

We can take, for example, $\varsigma(n) = \ln^m n$ for some fixed $m \ge 0$. It should be noted, that the $(0,1)$-Gaussian density $\mathbf{p}_0$ belongs to $\mathcal{P}$. Now we can formulate the following proposition from [Arkoun et al. (2019)] (Theorem 3.1).

Proposition 2.1. *For any* $\mathbf{b} > 0$

$$\lim_{n\to\infty} n^{\mathbf{b}} \max_{1\le l\le d} \sup_{S\in\Theta_{\varepsilon,L}} \sup_{\mathbf{p}\in\mathcal{P}} \mathbf{P}_{\mathbf{p},S}\left(|\widetilde{S}_l - S(z_l)| > \epsilon_0\right) = 0\,, \tag{2.12}$$

where $\epsilon_0 = \epsilon_0(n) \to 0$ *as* $n\to\infty$ *such that* ϵ_0^{-1} *is a slowly increasing function.*

Now we set

$$Y_l = S_l^* \mathbf{1}_\Gamma \quad\text{and}\quad \Gamma = \cap_{l=1}^d \Gamma_l\,. \tag{2.13}$$

In Theorem 3.2 from [Arkoun et al. (2019)] it is shown that the probability of Γ goes to zero uniformly faster than any power of the observations number n, which is formulated in the next proposition.

Proposition 2.2. *For any* $\mathbf{b} > 0$

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

In view of this proposition we can neglect the set Γ^c. So, using the estimators (2.13) on the set Γ we obtain the discrete time regression model

$$Y_l = S(z_l) + \zeta_l \quad \text{and} \quad \zeta_l = \eta_l + \varpi_l \,, \tag{2.14}$$

in which

$$\eta_l = \frac{\sum_{j=\iota_l+1}^{\tau_l-1} \mathbf{u}_{j,l}\,\xi_j + \varkappa_l \mathbf{u}_{\tau_l,l}\,\xi_{\tau_l}}{H_l} \quad \text{and} \quad \varpi_l = \varpi_{1,l} + \varpi_{2,l} \,,$$

where

$$\varpi_{1,l} = \frac{\sum_{j=\iota_l+1}^{\tau_l-1} \mathbf{u}_{j,l}\,(S(x_j) - S(z_l)) + \varkappa_l^2 \mathbf{u}_{\tau_l,l}\,(S(x_{\tau_l}) - S(z_{\tau_l}))}{H_l}$$

and $\varpi_{2,l} = (\varkappa_l - \varkappa_l^2)\mathbf{u}_{\tau_l,l}\,S(x_{\tau_l})/H_l$. Note that the random variables $(\eta_j)_{1\le j\le d}$ (see Lemma A.2 in [Arkoun et al. (2019)]), for any $1 \le l \le d$ and $\mathbf{p} \in \mathcal{P}$, are such that

$$\mathbf{E}_{\mathbf{p},S}\,(\eta_l\,|\mathcal{G}_l) = 0 \,, \quad \mathbf{E}_{\mathbf{p},S}\,(\eta_l^2\,|\mathcal{G}_l) = \sigma_l^2 \quad \text{and} \quad \mathbf{E}_{\mathbf{p},S}\,(\eta_l^4\,|\mathcal{G}_l) \le \mathbf{v}^*\sigma_l^4 \,, \tag{2.15}$$

where $\sigma_l = H_l^{-1/2}$, $\mathcal{G}_l = \sigma\{\eta_1, \ldots, \eta_{l-1}, \sigma_l\}$ and $\mathbf{v}^*$ is a fixed constant. Note also that

$$\sigma_{0,*} \le \min_{1\le l\le d} \sigma_l^2 \le \max_{1\le l\le d} \sigma_l^2 \le \sigma_{1,*} \,, \tag{2.16}$$

where

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

Remark 1. *It should be summarized that we construct the sequential pointwise procedure (2.7) – (2.8) in two steps. First, we preliminary estimate the function $S(z_l)$ in (2.5) with the observations $(y_j)_{k_{1,l}\le j\le \iota_l}$ and through this estimator we replace the random denominator in (2.8) with the threshold H_l in the second step when we construct the estimation procedure on the basis of the observations $(y_j)_{\iota_l<j\le k_{2,l}}$. Also remark that in the deviation (2.14) the main term η_l has a martingale form and the second one, as it is shown in [Arkoun et al. (2019)], is asymptotically small. Finally, it should be emphasized that namely these properties allow us to develop effective estimation methods.*

3 Model selection

Now we use the sequential model selection procedure from [Arkoun et al. (2019)] to estimate the function S for the regression (2.14). To this end, first we choose the trigonometric basis $(\phi_j)_{j\ge1}$ in $\mathcal{L}_2[a,b]$, i.e.

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

where the function $\mathrm{Tr}_j(x) = \cos(x)$ for even j and $\mathrm{Tr}_j(x) = \sin(x)$ for odd j, and $\mathbf{l}_0(x) = (x-a)/(b-a)$. Moreover, we choose the odd number d of regression points (2.1), for example, $d = 2[\sqrt{n}/2]+1$. Then the functions $(\phi_j)_{1\le j\le d}$ are orthonormal for the empirical inner product, i.e.

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

It is clear that the function S can be represented as

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

We define the estimators of the coefficients $(\theta_{j,d})_{1\le j\le d}$ by

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

From (2.14) we obtain immediately the following regression on the set Γ

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

where

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

It should be noted here that

$$\sum_{j=1}^{d} \varpi_{j,d}^2 = \|\varpi\|_d^2 = \frac{b-a}{d} \sum_{l=1}^{d} \varpi_l^2 \le (b-a)\frac{\varpi_n^*}{n}\,, \tag{3.6}$$

where $\varpi_n^* = n \max_{1\le l\le d} \varpi_l^2$. In [Arkoun et al. (2019)] (Theorem 3.3) it is shown that

$$\lim_{n\to\infty} \frac{1}{n^{\mathbf{b}}} \sup_{\mathbf{p}\in\mathcal{P}} \sup_{S\in\Theta_{\epsilon,L}} \mathbf{E}_{\mathbf{p},S} \varpi_n^*\, \mathbf{1}_\Gamma \;=\; 0 \quad \text{for any} \quad \mathbf{b} > 0\,. \tag{3.7}$$

For the model selection procedure we use weighted least squares estimators defined by

$$\widehat{S}_\lambda(t) = \sum_{l=1}^{d} \widehat{S}_\lambda(z_l) \mathbf{1}_{]z_{l-1},z_l]}\,, \quad \widehat{S}_\lambda(z_l) \;=\; \sum_{j=1}^{d} \lambda(j)\, \widehat{\theta}_{j,d}\, \phi_j(z_l)\, \mathbf{1}_\Gamma\,, \tag{3.8}$$

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

$\mathbf{H}_1$) : *Assume that the number of the weight vector ν is a slowly increasing function of n.*

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

$$\mathrm{Err}_d(\lambda) = \|\widehat{S}_\lambda - S\|_d^2 = \frac{b-a}{d} \sum_{l=1}^{d} (\widehat{S}_\lambda(z_l) - S(z_l))^2\,. \tag{3.9}$$

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

$$\mathrm{Err}_d(\lambda) = \sum_{j=1}^{d} \lambda^2(j)\widehat{\theta}_{j,d}^2 \,-\, 2\sum_{j=1}^{d} \lambda(j)\widehat{\theta}_{j,d}\,\theta_{j,d} \,+\, \sum_{j=1}^{d} \theta_{j,d}^2\,. \tag{3.10}$$

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

$$\widetilde{\theta}_{j,d} = \widehat{\theta}^2_{j,d} - \frac{b-a}{d}\,\mathbf{s}_{j,d} \quad \text{with} \quad \mathbf{s}_{j,d} = \frac{b-a}{d}\sum_{l=1}^{d}\sigma_l^2\,\phi_j^2(z_l)\,. \tag{3.11}$$

Note here that in view of (2.16) - (3.2), the last term can be estimated as

$$\sigma_{0,*} \le \mathbf{s}_{j,d} \le \sigma_{1,*}\,. \tag{3.12}$$

Now, we modify the risk (3.10) as

$$J_d(\lambda) \;=\; \sum_{j=1}^{d}\lambda^2(j)\widehat{\theta}^2_{j,d} \;-\; 2\sum_{j=1}^{d}\lambda(j)\,\widetilde{\theta}_{j,d} \;+\; \delta P_d(\lambda)\,, \tag{3.13}$$

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

$$P_d(\lambda) = \frac{b-a}{d}\sum_{j=1}^{d}\lambda^2(j)\mathbf{s}_{j,d}\,. \tag{3.14}$$

Now using (3.13) we define the sequential model selection procedure as

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda\in\Lambda} J_d(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,. \tag{3.15}$$

To study the efficiency property we specify the weight coefficients $(\lambda(j))_{1\le j\le n}$ as it is proposed, for example, in [Galtchouk and Pergamenshchikov (2009b)]. First, for some $0 < \varepsilon < 1$, we introduce the two dimensional grid to adapt to the unknown parameters (regularity and size) of the Sobolev ball, i.e. we set

$$\mathcal{A} = \{1,\ldots,k^*\}\times\{\varepsilon,\ldots,\mathbf{m}\varepsilon\}\,, \tag{3.16}$$

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

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty}\dfrac{k^*(n)}{\ln n} = 0\,, \\ \\ \lim_{n\to\infty}\varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\mathbf{b}}\varepsilon(n) \;= +\infty \end{cases} \tag{3.17}$$

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

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k_0^* + [\sqrt{\ln n}]\,, \tag{3.18}$$

where $k_0^* \ge 0$ is some fixed integer number. For each $\alpha = (k,\mathbf{t}) \in \mathcal{A}$, we introduce the weight sequence $\lambda_\alpha = (\lambda_\alpha(j))_{1\le j\le d}$ with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\le j<j_*\}} + \left(1-(j/\omega_\alpha)^k\right)\,\mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}\,, \tag{3.19}$$

where

$$\omega_\alpha = \omega_* + (b-a)^{2k/(2k+1)} \left(\mathrm{d}_k^* \, \mathbf{t} \, n\right)^{1/(2k+1)} \quad \text{and} \quad \mathrm{d}_k^* = (k+1)(2k+1)/(\pi^{2k} k)\,.$$

Here, j_* and ω_* are such that $j_* \to \infty$, $j_* = \mathrm{o}\left((n/\varepsilon)^{1/(2k+1)}\right)$ and $\omega_* = \mathrm{O}(j_*)$ as $n \to \infty$. In this case we define the weight set as $\Lambda = \{\lambda_\alpha\,, \alpha \in \mathcal{A}\}$. Note, that these weight coefficients are used in [Konev and Pergamenshchikov (2012), Konev and Pergamenshchikov (2015)] 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 (3.17) imply the condition $\mathbf{H}_1$). In [Arkoun et al. (2019)] we showed the following result.

Theorem 3.1. *Assume that the conditions (2.11) and $\mathbf{H}_1$) hold. Then for any $n \geq 3$, any $S \in \Theta_{\varepsilon,L}$ and any $0 < \delta \leq 1/12$, the procedure (3.15) with the coefficients (3.19) satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{(1+4\delta)(1+\delta)^2}{1-6\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{B}_n^*}{\delta n}\,, \tag{3.20}$$

where the term $\mathbf{B}_n^*$ is a slowly increasing function.

Remark 2. *In this paper we will use the inequality (3.20) to study efficiency properties for the model selection procedure (3.15) with the weight coefficients (3.19) 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 of 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

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

where $\mathbf{a}_j = 1 + \sum_{l=1}^{\mathbf{k}} (2\pi[j/2]/(b-a))^{2l}$, $(\theta_j)_{j\geq 1}$ are the trigonometric Fourier coefficients in $\mathcal{L}_2[a,b]$, i.e. $\theta_j = (f, \phi_j) = \int_a^b f(x)\phi_j(x)\mathrm{d}x$ and $(\phi_j)_{j\geq 1}$ is the trigonometric basis defined in (3.1). It is clear that we can represent this functional class as the Sobolev ball

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

Now, for this set we define the normalizing coefficients

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

and

$$\varsigma_* = \varsigma_*(S) = \int_a^b (1 - S^2(u))\mathrm{d}u\,. \tag{4.2}$$

It is well known that 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, [Galtchouk and Pergamenshchikov (2009b), Konev and Pergamenshchikov (2009)] and the references therein). Our goal in this paper is to show the same property for the nonparametric autoregressive model (1.2). First we have to obtain a lower bound for the risk (1.3) over all possible estimators Ξ_n, i.e. any measurable function of the observations $(y_1, \ldots, y_n)$.

Theorem 4.1. *The robust risk (1.3) normalized by the coefficient $\upsilon(S) = \varsigma_*^{-2\mathbf{k}/(2\mathbf{k}+1)}$ can be estimated from below as*

$$\liminf_{n\to\infty} \inf_{\widehat{S}_n\in\Xi_n} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\mathcal{R}^*(\widehat{S}_n, S) \geq (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)} \mathbf{l}_{\mathbf{k},\mathbf{r}} \,. \tag{4.3}$$

Now to study the procedure (3.15) we have to add some condition on the penalty coefficient δ which provides sufficiently small penalty term in (3.13).

$\mathbf{H}_2$) : *Assume that the parameter $\delta = \delta_n$ is a function of n which goes to zero as $n \to \infty$ such that δ_n^{-1} is slowly increasing.*

Theorem 4.2. *Assume that the conditions $\mathbf{H}_1$) and $\mathbf{H}_2$) hold. Then the model selection procedure $\widehat{S}_*$ defined in (3.15) with the weight vectors (3.19) admits the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\, \mathcal{R}(\widehat{S}_*, S) \leq (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\, \mathbf{l}_{\mathbf{k},\mathbf{r}} \,.$$

Now, Theorems 4.1 - 4.2 imply the following efficiency property.

Corollary 4.1. *Assume that the conditions $\mathbf{H}_1$) – $\mathbf{H}_2$) hold. The model selection procedure $\widehat{S}_*$ defined in (3.15) and (3.19) is efficient, i.e.*

$$\lim_{n\to\infty} \frac{\inf_{\widehat{S}_n\in\Xi_n} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\, \mathcal{R}^*(\widehat{S}_n, S)}{\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\mathcal{R}(\widehat{S}_*, S)} = 1 \,. \tag{4.4}$$

Moreover,

$$\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \inf_{\widehat{S}_n\in\Xi_n} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\, \mathcal{R}^*(\widehat{S}_n, S) = (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\, \mathbf{l}_{\mathbf{k},\mathbf{r}}$$

and

$$\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\, \mathcal{R}^*(\widehat{S}_*, S) = (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\, \mathbf{l}_{\mathbf{k},\mathbf{r}} \,.$$

Remark 3. *It should be noted that the inequality (4.3) can be represented as*

$$\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \inf_{\widehat{S}_n\in\Xi_n} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} (\mathcal{I}(S))^{2\mathbf{k}/(2\mathbf{k}+1)}\, \overline{\mathcal{R}}(\widehat{S}_n, S) \geq (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\, \mathbf{l}_{\mathbf{k},\bar{\mathbf{r}}} \,, \tag{4.5}$$

where $\bar{\mathbf{r}} = \mathbf{r}/(b-a)$, $\mathcal{I}(S) = (b-a)\left(\int_a^b (1 - S^2(x))\mathrm{d}x\right)^{-1}$ and

$$\overline{\mathcal{R}}(\widehat{S}_n, S) = \sup_{\mathbf{p}\in\mathcal{P}} \mathbf{E}_{\mathbf{p},S} \frac{1}{b-a} \int_a^b \left(\widehat{S}_n(x) - S(x)\right)^2 \mathrm{d}x \,.$$

Note also that the function $\mathcal{I}(S)$ is the nonparametric version of the asymptotic Fisher information. We remind that for the parametric model (1.2), i.e. when $S \equiv \vartheta$ and $|\vartheta| < 1$, the Fisher information is $(1-\vartheta^2)^{-1}$ (see, for example, [Amderson1970]). Moreover, the coefficient $(b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\,\mathbf{l}_{\mathbf{k},\bar{\mathbf{r}}}$ is the well known Pinsker constant for the "signal+white noise" model with the noise intensity $b-a$ and the radius $\bar{\mathbf{r}}$ (see [Pinsker (1981)]). Therefore the lower bound (4.5) is the nonparametric version of the Rao-Cramer inequality for the models (1.2).

Now we assume that in the model (1.1) the functions $(\psi_i)_{i\geq 1}$ are orthonormal in $\mathcal{L}_2[a,b]$, i.e. $(\psi_i,\psi_j) = \mathbf{1}_{\{i=j\}}$. We use the estimators (3.8) to estimate the parameters $\beta = (\beta_i)_{i\geq 1}$ as $\widehat{\beta}_\lambda = (\widehat{\beta}_{\lambda,i})_{i\geq 1}$ and $\widehat{\beta}_{\lambda,i} = (\psi_i, \widehat{S}_\lambda)$. Then, similarly we use the selection model procedure (3.15) as

$$\widehat{\beta}_* = (\widehat{\beta}_{*,i})_{i\geq 1} \quad \text{and} \quad \widehat{\beta}_{*,i} = (\psi_i, \widehat{S}_*)\,. \tag{4.6}$$

It is clear that in this case $|\widehat{\beta}_\lambda - \beta|^2 = \sum_{i=1}^{\infty}(\widehat{\beta}_{\lambda,i}-\beta_i)^2 = \|\widehat{S}_\lambda - S\|^2$ and $|\widehat{\beta}_* - \beta|^2 = \|\widehat{S}_* - S\|^2$.

Note that Theorem 3.1 implies that the estimator (4.6) is optimal in the sharp oracle inequality sense which is established in the following theorem.

Theorem 4.3. *For any $S \in \Theta_{\varepsilon,L}$, $n \geq 3$ and $0 < \delta \leq 1/12$,*

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

where $\mathcal{R}^*(\widehat{\beta},\beta) = \sup_{\mathbf{p}\in\mathcal{P}}\mathbf{E}_{\mathbf{p},S}|\widehat{\beta}-\beta|^2$ and $\mathbf{B}^*_n$ satisfies the limit property mentioned in Theorem 3.1.

Remark now that Theorems 4.1 and 4.2 imply the efficiency property for the estimate (4.6) based on the model selection procedure (3.15) constructed with the penalty threshold δ satisfying the condition $\mathbf{H}_2$).

Theorem 4.4. *Then the estimate (4.6) is asymptotically efficient, i.e.*

$$\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\mathcal{R}^*(\widehat{\beta}_*,\beta) = (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\,\mathbf{l}_{\mathbf{k},\mathbf{r}}$$

and

$$\lim_{n\to\infty}\frac{\inf_{\widehat{\beta}_n\in\Xi_n}\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}}\upsilon(S)\mathcal{R}^*(\widehat{\beta}_n,\beta)}{\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}}\upsilon(S)\mathcal{R}^*(\widehat{\beta}_*,\beta)} = 1\,, \tag{4.7}$$

where Ξ_n is the set estimators for β based on the observations $(y_j)_{1\leq j\leq n}$.

Remark 4. *It should be noted that we obtain the efficiency property (4.7) for the big data autoregressive model (1.2) without using the parameter dimension q or sparse conditions usually used for such models (see, for example, [Hastie et al.(2008)]).*

5 Monte-Carlo simulations

In this section we present the numeric results obtained through the Python software for the model (1.2) in which $(\xi_j)_{1\leq j\leq n}$ are i.i.d. $\mathcal{N}(0,1)$ random variables and $0 \leq x \leq 1$, i.e. $a = 0$ and $b = 1$. In this case we simulate the model selection procedure (3.15) with the weights

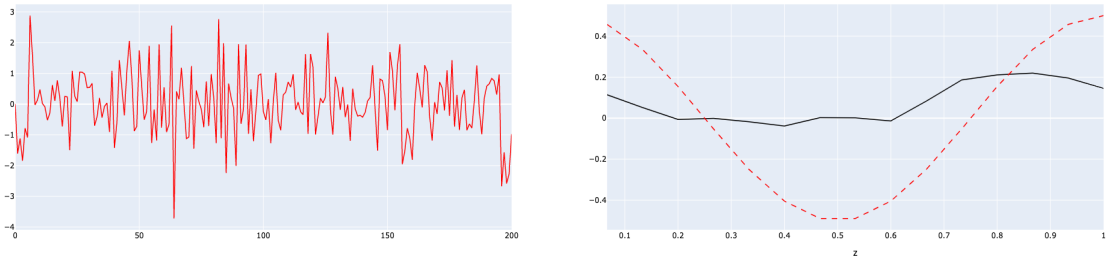
(3.19) in which $k^* = 150 + \sqrt{\ln n}$, $\mathbf{m} = [\ln^2 n]$, $\varepsilon = 1/\ln n$. Moreover, the parameters j_* and ω_* are chosen as

$$j_* = \frac{\underline{\omega}}{200 + \ln \underline{\omega}}\,, \qquad \underline{\omega} = \ln n + \left(\frac{(k+1)(2k+1)}{\pi^{2k} k} \mathbf{t}\, n \right)^{1/(2k+1)}$$

and $\omega_* = j_* + \ln n$. First we study the model (1.2) with $S_1(x) = 0,5\cos(2\pi x)$ and then for the function

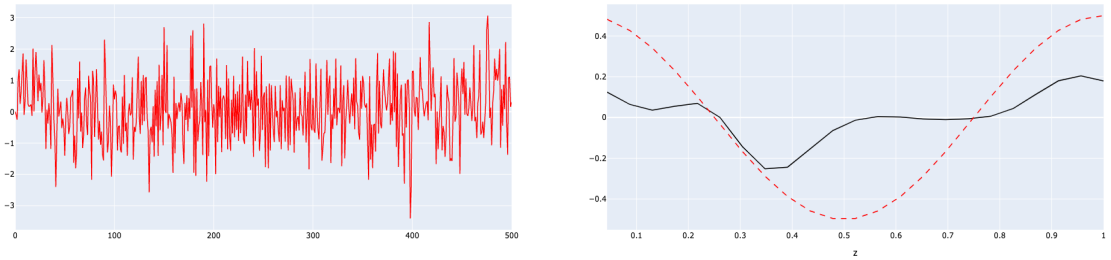
$$S_2(x) = 0,1 + \sum_{j=1}^{q} \frac{\cos(2\pi j x)}{(j+3)^2} \quad \text{and} \quad q = 100000\,.$$

In the model selection procedure we use $d = 2[\sqrt{n}/2] + 1$ points in (2.1).



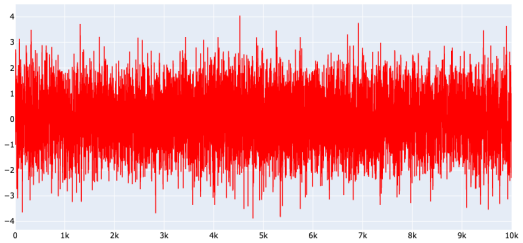
(a) Observations for $n = 200$ (b) Estimator of S_1 for $n = 200$

Figure 1: Model selection for $n = 200$

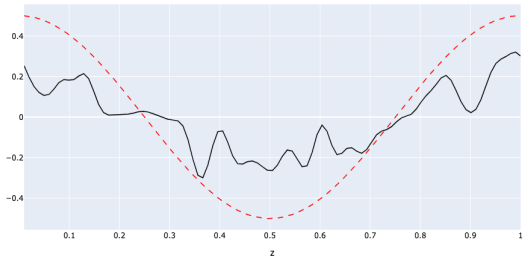


(a) Observations for $n = 500$ (b) Estimator of S_1 for $n = 500$

Figure 2: Model selection for $n = 500$

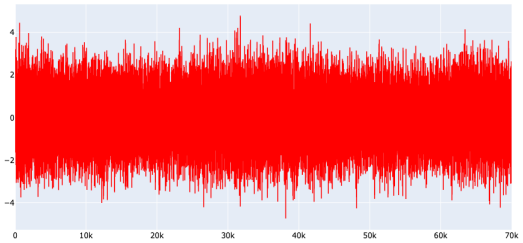


(a) Observations for $n = 10000$

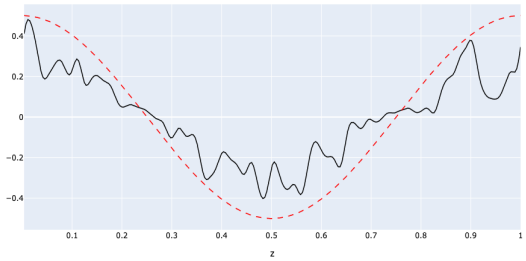


(b) Estimator of S_1 for $n = 10000$

Figure 3: Model selection for $n = 10000$



(a) Observations for $n = 70000$



(b) Estimator of S_1 for $n = 70000$

Figure 4: Model selection for $n = 70000$

Figures 1–4 show the behavior of the function S_1 and its estimators by the model selection procedure (3.15) depending on the observations number n. On the figures (a) the observations are given and on the figures (b) the red dotted is the regression function and the black full line is its estimator at the points (2.1). Then we calculate the empiric risks as

$$\widehat{\mathcal{R}} = \frac{1}{d}\sum_{j=1}^{d} \widehat{\mathbf{E}}\left(\widehat{S}_n(z_j) - S(z_j)\right)^2, \tag{5.1}$$

where the expectation is taken as an average over $M = 50$ replications, i.e.

$$\widehat{\mathbf{E}}\left(\widehat{S}_n(.) - S(.)\right)^2 = \frac{1}{M}\sum_{l=1}^{M}\left(\widehat{S}_n^l(\cdot) - S(\cdot)\right)^2.$$

We use also the relative risk

$$\widehat{\mathcal{R}}_* = \frac{\widehat{\mathcal{R}}}{\|S\|_n^2} \quad \text{and} \quad \|S\|_n^2 = \frac{1}{n}\sum_{j=1}^{n} S^2(x_j)\,. \tag{5.2}$$

The tables below give the values for the sample risks (5.1) and (5.2) for different numbers of observations n.

Table 1: Empirical risks for S_1

n	$\widehat{\mathcal{R}}$	$\widehat{\mathcal{R}}_*$
200	0.135	0.98
500	0.0893	0.624
10000	0.043	0.362
70000	0.03523	0.281

Table 2: Empirical risks for S_2

n	$\widehat{\mathcal{R}}$	$\widehat{\mathcal{R}}_*$
200	0.0821	5.685
500	0.0386	2.623
10000	0.0071	0.516
70000	0.0067	0.419

Remark 5. *From numerical simulations of the procedure (3.15) with various observation durations n and for the different functions S we may conclude that the quality of the estimation improves as the number of observations increases.*

6 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 $(\xi_l)_{1\le l\le n}$ and the parametric linear function S, i.e.

$$S_\theta(x)=\sum_{j=1}^{d}\theta_j\,\psi_j(x)\,,\quad \theta=(\theta_1,\ldots,\theta_d)'\in\mathbb{R}^d\,. \tag{6.1}$$

We assume that the functions $(\psi_j)_{1\le j\le d}$ are orthogonal with respect to the scalar product (3.2). 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 (1.2) with the function (6.1) 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 by

$$f_n(y,\theta)=\frac{\mathrm{d}\mathbf{P}_\theta^{(n)}}{\mathrm{d}\nu_\xi^n}=\exp\left\{\sum_{l=1}^{n}S_\theta(x_l)y_{l-1}y_l-\frac{1}{2}\sum_{j=1}^{n}S_\theta^2(x_l)y_{l-1}^2\right\}. \tag{6.2}$$

Let ϱ be a prior distribution density on $\mathbb{R}^d$ for the parameter θ of the following form $\varrho(\theta)=\prod_{j=1}^{d}\varrho_j(\theta_j)$, where ϱ_j is a some probability density in $\mathbb{R}$ with continuously derivative $\dot\varrho_j$ for which the Fisher information is finite, i.e.

$$I_j=\int_{\mathbb{R}}\frac{\dot\varrho_j^2(z)}{\varrho_j(z)}\,\mathrm{d}z<\infty\,. \tag{6.3}$$

Let $g(\theta)$ be a continuously differentiable $\mathbb{R}^d\to\mathbb{R}$ function such that

$$\lim_{|\theta_j|\to\infty}g(\theta)\,\varrho_j(\theta_j)=0\quad\text{and}\quad\int_{\mathbb{R}^d}|g_j'(\theta)|\,\mathbf{u}(\theta)\,\mathrm{d}\theta<\infty\,, \tag{6.4}$$

where $g_j'(\theta)=\partial g(\theta)/\partial\theta_j$.
For any $Q(\mathbb{R}^{n+d})$-measurable integrable function $H=H(y,\theta)$ we denote

$$\widetilde{\mathbf{E}}\,H=\int_{\mathbb{R}^{n+d}}H(y,\theta)\,f_n(y,\theta)\,\varrho(\theta)\mathrm{d}\nu_\xi^{(n)}\,\mathrm{d}\theta\,. \tag{6.5}$$

Let $\mathcal{F}_n^y$ be the field generated by the observations (1.2), i.e. $\mathcal{F}_n^y=\sigma\{y_1,\ldots,y_n\}$. Now we study the Gaussian the model (1.2) with the function (6.1).

Proposition 6.1. *For any $\mathcal{F}_n^y$-measurable square integrable $\mathbb{R}^n\to\mathbb{R}$ function $\widehat{g}_n$ and for any $1\le j\le d$, the mean square accuracy of the function $g(\cdot)$ with respect to the distribution (6.5) can be estimated from below as*

$$\widetilde{\mathbf{E}}(\widehat{g}_n-g(\theta))^2\ge\frac{\overline{g}_j^2}{\widetilde{\mathbf{E}}\,\Psi_{n,j}+I_j}\,, \tag{6.6}$$

where $\Psi_{n,j}=\sum_{l=1}^{n}\psi_j^2(x_l)\,y_{l-1}^2$ and $\overline{g}_j=\int_{\mathbb{R}^d}g_j'(\theta)\,\varrho(\theta)\,\mathrm{d}\theta$.

Proof. First, for any $\theta \in \mathbb{R}^d$ we set

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

Taking into account the condition (6.4) and integrating by parts we get

$$\widetilde{\mathbf{E}}\left((\widehat{g}_n - g(\theta))\widetilde{U}_j\right) = \int_{\mathbb{R}^{n+d}} (\widehat{g}_n(y) - g(\theta)) \frac{\partial}{\partial \theta_j} \left(f_n(y,\theta)\mathbf{u}(\theta)\right) \mathrm{d}\theta \, \mathrm{d}\nu_\xi^{(n)}$$

$$= \int_{\mathbb{R}^{n+d-1}} \left(\int_{-\infty}^{+\infty} g_j'(\theta) \, f_n(y,\theta)\mathbf{u}(\theta) \mathrm{d}\theta_j \right) \left(\prod_{i \neq j} \mathrm{d}\theta_i \right) \mathrm{d}\nu_\xi^{(n)} = \overline{g}_j \,.$$

Now by the Bouniakovskii-Cauchy-Schwarz inequality we obtain the following lower bound for the quadratic risk $\widetilde{\mathbf{E}}(\widehat{g}_n - g(\theta))^2 \geq \overline{g}_j^2 / \widetilde{\mathbf{E}}\widetilde{U}_j^2$. To study the denominator in this inequality note that in view of the representation (6.2)

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

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

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

we get $\widetilde{\mathbf{E}} \, \widetilde{U}_j^2 = \widetilde{\mathbf{E}} \, \Psi_{n,j} + I_j$. Hence Lemma 6.1. $\square$

7 Lower bound

First, taking into account that the $(0,1)$-Gaussian density $\mathbf{p}_0$ belongs to the class (2.10), we get $\mathcal{R}^*(\widehat{S}_n, S) \geq \mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S)$. Now, according to the general lower bound methods (see, for example, [Galtchouk and Pergamenshchikov (2009b)]) one needs to estimate this risk from below by some bayesian risk and then to apply the van Trees inequality. To define the bayesian risk we need to choose a prior distribution on $\mathcal{W}_{\mathbf{k},\mathbf{r}}$. To this end, first for any vector $\kappa = (\kappa_j)_{1 \leq j \leq d} \in \mathbb{R}^d$, we set

$$S_\kappa(x) = \sum_{j=1}^{d} \kappa_j \, \phi_j(x) \,, \tag{7.1}$$

where $(\phi_j)_{1 \leq j \leq d}$ is the trigonometric basis defined in (3.1). Now, to use the van Trees inequality we introduce the Bayes risk as

$$\widetilde{\mathcal{R}}_0(\widehat{S}_n) = \widetilde{\mathbf{E}}_0 \, \|\widehat{S}_n - S_z\|^2 = \int_{\mathbb{R}^d} \mathbf{E}_{\mathbf{p}_0, S_z} \, \|\widehat{S}_n - S_z\|^2 \, \mu_\kappa(\mathrm{d}z) \,, \tag{7.2}$$

where μ_κ is the distribution in $\mathbb{R}^d$ of the random vector $\kappa = (\kappa_j)_{1\le j\le d}$ defined by its components as

$$\kappa_j = s_j\,\eta_j^*\,, \tag{7.3}$$

where $(s_j)_{1\le j\le d}$ are some positive coefficients which will be specified later, $(\eta_j^*)_{1\le j\le d}$ 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 $n > e^2$. As to the number of the terms d in (7.1) similarly to [Pinsker (1981)] we choose it as a function of n, i.e. $d = d_n$, such that $d_n \to \infty$ as $n \to \infty$ and, moreover, this function must belong to the class $\mathcal{W}_{\mathbf{k},\mathbf{r}}$ with a probability approaching one faster than any power function of n^{-1}. For this, as we will see below in Lemma A.3, it is necessary to provide for a fixed arbitrary $0 < \rho < 1$ the following property

$$\limsup_{n\to\infty} \sum_{j=1}^{d} \mathbf{a}_j\, s_j^2 \le \rho\mathbf{r}\,.$$

Taking into account that $\mathbf{a}_j \approx (\pi j/(b-a))^{2\mathbf{k}}$ for $j \to \infty$, this condition can be rewritten as

$$\sum_{j=1}^{d} j^{2\mathbf{k}}\, s_j^2 \le \rho\mathbf{r}(b-a)^{2\mathbf{k}}/\pi^{2\mathbf{k}}\,. \tag{7.4}$$

It is clear that almost surely the function (7.1) can be bounded as

$$\max_{a\le x\le b}\Big(|S_\kappa(x)| + |\dot{S}_\kappa(x)|\Big) \le \mathbf{c}_*\,\delta_n^*\,, \tag{7.5}$$

where

$$\mathbf{c}_* = \sqrt{2}(b-a+\pi)/(b-a)^{3/2} \quad \text{and} \quad \delta_n^* = \ln n \sum_{j=1}^{d} j\, s_j\,.$$

In the sequel we choose the number of the terms d in (7.1) and the coefficients s_j in (7.3) such that $\delta_n^* \to 0$ as $n \to \infty$. Therefore, under this condition for sufficiently large n the function (7.1) belongs to the class (2.4) almost surely. Now, for all $f \in \mathcal{L}_2[a,b]$, we denote by $\mathbf{h}(f)$ its projection in $\mathcal{L}_2[a,b]$ onto the ball $\mathcal{W}_{\mathbf{r}}^* = \{f \in \mathcal{L}_2[a,b] : \|f\| \le \mathbf{r}\}$, i.e. $\mathbf{h}(f) = (\mathbf{r}/\max(\mathbf{r}, \|f\|))f$. Since $\mathcal{W}_{\mathbf{k},\mathbf{r}} \subset \mathcal{W}_{\mathbf{r}}^*$, we have $\|\widehat{S}_n - S\|^2 \ge \|\widehat{\mathbf{h}}_n - S\|^2$ for $S \in \mathcal{W}_{\mathbf{k},\mathbf{r}}$, where $\widehat{\mathbf{h}}_n = \mathbf{h}(\widehat{S}_n)$. This implies that for sufficiently large n

$$\begin{aligned}\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\,\mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S) &\ge \int_\Gamma \upsilon(S_z)\mathbf{E}_{\mathbf{p}_0,S_z}\|\widehat{\mathbf{h}}_n - S_z\|^2\,\mu_\kappa(\mathrm{d}z)\\ &\ge \upsilon_* \int_\Gamma \mathbf{E}_{\mathbf{p}_0,S_z}\|\widehat{\mathbf{h}}_n - S_z\|^2\,\mu_\kappa(\mathrm{d}z)\,,\end{aligned} \tag{7.6}$$

where $\Gamma = \Big\{z \in \mathbb{R}^d : \sum_{j=1}^d \mathbf{a}_j z_j^2 \le \mathbf{r}\Big\}$ and $\upsilon_* = \inf_{|S|_*\le \mathbf{c}_*\delta_n^*} \upsilon(S)$. Now taking into account that $\|\widehat{\mathbf{h}}_n\|^2 \le r$, we get

$$\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\,\mathcal{R}_{\mathbf{p}_0}(\widehat{S}_n, S) \ge \upsilon_*\,\widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}_n) - 2\,\upsilon_*\,\mathbf{R}_{0,n} \tag{7.7}$$

and $\mathbf{R}_{0,n} = \int_{\Gamma^c} (r + \|S_z\|^2)\, \mu_\kappa(\mathrm{d}z) = \int_{\Gamma^c} \left(\mathbf{r} + |z|^2\right) \mu_\kappa(\mathrm{d}z)$. Remark here that this term is studied in Lemma A.3. Moreover, note also that for any $z \in \mathbb{R}^d$ we get $\|\widehat{\mathbf{h}}_n - S_z\|^2 \geq \sum_{j=1}^d (\widehat{z}_j - z_j)^2$ with $\widehat{z}_j = \left(\widehat{\mathbf{h}}_n\,, \phi_j\right)$. Furthermore, from Lemma 6.1 with $g(\theta) = \theta_j$ it follows that for any $1 \leq j \leq d$ and any $\mathcal{F}_n^y$ measurable random variable $\widehat{\kappa}_j$

$$\widetilde{\mathbf{E}}_0(\widehat{\kappa}_j - \kappa_j)^2 \geq \frac{1}{\widetilde{\mathbf{E}}_0\, \Psi_{n,j} + s_j^{-2} J_n}\,,$$

where $\Psi_{n,j} = \sum_{l=1}^n \phi_j^2(x_l) y_{l-1}^2$ and $J_n = \int_{-\ln n}^{\ln n} (\dot{\rho}_n(t))^2/\rho_n(t)\mathrm{d}t$. Therefore, the Bayes risk can be estimated from below as

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

where

$$\overline{\Psi}_{n,j} = \frac{(b-a)}{n} \sum_{l=1}^n \phi_j^2(x_l) \widetilde{\mathbf{E}}_0\, y_{l-1}^2 \quad \text{and} \quad s_j^* = \frac{n}{b-a} s_j^2\,. \tag{7.8}$$

Note here that in view of Lemmas A.1 and A.2 for any $0 < \rho_1 < 1$ and sufficiently large n we have $J_n \leq 1 + \rho_1$ and $\max_{1\leq j\leq d} \overline{\Psi}_{n,j} \leq 1 + \rho_1$, hence,

$$n^{2\mathbf{k}/(2\mathbf{k}+1)}\, \widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) \geq \frac{(b-a)}{(1+\rho_1)\, n^{1/(2\mathbf{k}+1)}} \sum_{j=1}^d \frac{s_j^*}{1+s_j^*}\,. \tag{7.9}$$

Now we choose $(s_j^*)_{1\leq j\leq d}$ to maximise this lower bound under the condition (7.4), i.e.

$$\max_{v_1\geq 0,\ldots,v_d\geq 0} \sum_{j=1}^d \frac{v_j}{1+v_j} \quad \text{subjected to} \quad \sum_{j=1}^d j^{2\mathbf{k}}\, v_j \leq \mathbf{B}\, n\,, \quad \mathbf{B} = \frac{\rho \mathbf{r} (b-a)^{2\mathbf{k}-1}}{\pi^{2\mathbf{k}}}\,. \tag{7.10}$$

Using the Lagrange multipliers method, the solution of this problem can be given by

$$s_j^* = \mathbf{u}_d \left(\frac{d}{j}\right)^{\mathbf{k}} - 1 \quad \text{and} \quad \mathbf{u}_d = \frac{\mathbf{B}\, n + \sum_{l=1}^d l^{2\mathbf{k}}}{d^{\mathbf{k}} \sum_{l=1}^d l^{\mathbf{k}}}\,. \tag{7.11}$$

Therefore, in (7.3) we choose $s_j = \sqrt{(b-a)s_j^*/n}$ and d such, that

$$\mathbf{u}_d \geq 1\,. \tag{7.12}$$

Using this in (7.9), we deduce that

$$n^{2\mathbf{k}/(2\mathbf{k}+1)}\, \widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) \geq \frac{(b-a)d}{(1+\rho_1)\, n^{1/(2\mathbf{k}+1)}} \left(1 - \frac{1}{\mathbf{u}_d} \sum_{j=1}^d \left(\frac{j}{d}\right)^{\mathbf{k}} \frac{1}{d}\right)\,.$$

To obtain here a positive lower bound, we set $d = d_n = [Dn^{1/(2k+1)}]$ for some $D > 0$. Then

$$\lim_{n\to\infty} \mathbf{u}_d = \frac{\mathbf{B}(\mathbf{k}+1)}{D^{2\mathbf{k}+1}} + \frac{\mathbf{k}+1}{2\mathbf{k}+1} := \mathbf{u}_*(D)$$

and, using the definitions (7.11) in (7.5) we obtain that for any $\mathbf{k} \geq 2$

$$\delta_n^* \leq \sqrt{(b-a)\mathbf{u}_d}\,\frac{\ln n}{\sqrt{n}}\,d^{\mathbf{k}/2}\sum_{j=1}^{d} j^{1-\mathbf{k}/2} \to 0 \quad \text{as} \quad n \to \infty\,.$$

Moreover, in this case

$$\liminf_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)}\,\widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) \geq \frac{b-a}{(1+\rho_1)}\,L_*(D) \quad \text{and} \quad L_*(D) = D - \frac{D}{\mathbf{u}_*(D)(k+1)}\,.$$

Note here that the derivative $L_*'(D) = \big(1-\mathbf{u}_*^{-1}(D)\big)^2 \geq 0$. In view of the condition (7.12) we obtain that $\mathbf{u}_*(D) \geq 1$, i.e.

$$0 < D \leq \left(\frac{(2\mathbf{k}+1)(\mathbf{k}+1)}{\mathbf{k}}\mathbf{B}\right)^{1/(2k+1)} := D_*\,.$$

Therefore,

$$\max_{0<D\leq D_*} L_*(D) = L_*(D_*) = D_*\frac{k}{k+1} = \rho^{1/(2\mathbf{k}+1)}\,(b-a)^{(2\mathbf{k}-1)/(2\mathbf{k}+1)}\mathbf{l}_{\mathbf{k},\mathbf{r}}$$

and we get that for any $0 < \rho < 1$

$$\liminf_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)}\,\widetilde{\mathcal{R}}_0(\widehat{\mathbf{h}}) \geq \frac{\rho^{1/(2\mathbf{k}+1)}\,(b-a)^{4\mathbf{k}/(2\mathbf{k}+1)}}{(1+\rho_1)}\mathbf{l}_{\mathbf{k},\mathbf{r}}\,.$$

Taking into account that $\upsilon_* \to (b-a)^{-2\mathbf{k}/(2\mathbf{k}+1)}$ as $n \to \infty$ in (7.7), we conclude through Lemma A.3, that for any $0 < \rho < 1$ and $\rho_1 > 0$

$$\liminf_{n\to\infty} \inf_{\widehat{S}_n\in\Xi_n} n^{\frac{2k}{2k+1}} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \upsilon(S)\,\mathcal{R}^*(\widehat{S}_n, S) \geq \frac{\rho^{\frac{1}{2k+1}}}{1+\rho_1}\,(b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\,\mathbf{l}_{\mathbf{k},\mathbf{r}}\,.$$

Taking here the limits as $\rho \to 1$ and $\rho_1 \to 0$ we come to the Theorem 4.1. □

8 Upper bound

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 (4.2). To do this, we denote

$$\widetilde{\omega} = \omega_{\widetilde{\alpha}} \quad \text{and} \quad \widetilde{\alpha} = (\mathbf{k}, \widetilde{\mathbf{t}}_n)\,, \tag{8.1}$$

where $\widetilde{\mathbf{t}}_n = [\widetilde{r}(S)/\varepsilon]\,\varepsilon$, $\widetilde{r}(S) = \mathbf{r}/\varsigma_*$ and $\varepsilon = 1/\ln n$. In this case we use the estimator (3.8) with the corresponding weight vector defined in (3.19), i.e.

$$\widetilde{S} = \widehat{S}_{\widetilde{\lambda}} \quad \text{and} \quad \widetilde{\lambda} = \lambda_{\widetilde{\alpha}}\,. \tag{8.2}$$

Note, that for sufficiently large n, for which $k^* \geq \mathbf{k}$ and $\mathbf{m} \geq [\widetilde{r}(S)/\varepsilon]$ the parameter $\widetilde{\alpha}$ belongs to the set (3.16). In this section we study the risk of the estimator (8.2). To this end we need firstly to analyse the asymptotic behavior of the sequance

$$\Upsilon_n(S) = \sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 + \frac{\varsigma_*}{n}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,. \tag{8.3}$$

Proposition 8.1. *The sequence $\Upsilon_n(S)$ is bounded from above*

$$\limsup_{n\to\infty} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} n^{2\mathbf{k}/(2\mathbf{k}+1)}\, \upsilon(S)\, \Upsilon_n(S) \,\leq\, (b-a)^{2\mathbf{k}/(2\mathbf{k}+1)} \mathbf{l}_{\mathbf{k},\mathbf{r}}\,. \tag{8.4}$$

Proof. First, note that $0 < \epsilon^2(b-a) \leq \inf_{S\in\Theta_{\varepsilon,L}} \varsigma_* \leq \sup_{S\in\Theta_{\varepsilon,L}} \varsigma_* \leq b-a$. This implies directly that

$$\lim_{n\to\infty} \sup_{S\in\Theta_{\varepsilon,L}} \left|\widetilde{\mathbf{t}}_n/\widetilde{r}(S) - 1\right| = 0\,, \tag{8.5}$$

where $\widetilde{\mathbf{t}}_n = [\widetilde{r}(S)/\varepsilon]\,\varepsilon$ and $\widetilde{r}(S) = \mathbf{r}/\varsigma_*$. Note now that

$$\Upsilon_n(S) = \mathbf{G}_n + \frac{\varsigma_*}{n}\sum_{j=1}^{d} \widetilde{\lambda}^2(j)\,,$$

where $\mathbf{G}_n = \sum_{j=j_*}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\,\theta_{j,d}^2 + \sum_{j=[\widetilde{\omega}]+1}^{d} \theta_{j,d}^2 := \mathbf{G}_{1,n} + \mathbf{G}_{2,n}$. Moreover Lemma A.5 and Lemma A.6 yield

$$\mathbf{G}_{1,n} \leq (1+\widetilde{\varepsilon}) \sum_{j=j_*}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\,\theta_j^2 + 4\mathbf{r}(1+\widetilde{\varepsilon}^{-1})\,\frac{(b-a)^{2\mathbf{k}}\widetilde{\omega}}{d^{2k}}$$

and

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

i.e. $\mathbf{G}_n \leq (1+\widetilde{\varepsilon})\mathbf{G}_n^* + 4\mathbf{r}(b-a)^{2\mathbf{k}}(1+\widetilde{\varepsilon}^{-1})\,\widetilde{\mathbf{G}}_n$, where

$$\mathbf{G}_n^* = \sum_{j\geq1}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 = \sum_{j\leq\widetilde{\omega}}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 + \sum_{j>\widetilde{\omega}} \theta_j^2 := \mathbf{G}_{1,n}^* + \mathbf{G}_{2,n}^*$$

and $\widetilde{\mathbf{G}}_n = \widetilde{\omega}\,d^{-2\mathbf{k}} + d^{-2}\widetilde{\omega}^{-2(\mathbf{k}-1)}$. One also has

$$\mathbf{G}_{1,n}^* = \frac{1}{\widetilde{\omega}^{2\mathbf{k}}}\sum_{j=j_*}^{[\widetilde{\omega}]} j^{2\mathbf{k}}\,\theta_j^2 \leq \frac{\mathbf{u}_n^*}{\widetilde{\omega}^{2\mathbf{k}}}\sum_{j=j_*}^{[\widetilde{\omega}]} \mathbf{a}_j\,\theta_j^2\,,$$

where $\mathbf{u}_n^* = \sup_{j\geq j_*} j^{2k}/\mathbf{a}_j$. It is clear that $\lim_{n\to\infty} \mathbf{u}_n^* = (b-a)^{2\mathbf{k}}\pi^{-2k}$. Therefore, for any $0<\widetilde{\varepsilon}<1$ and for sufficiently large n, uniformly in $S\in\Theta_{\varepsilon,L}$

$$\mathbf{G}_{1,n}^* \leq (1+\widetilde{\varepsilon})\frac{(b-a)^{2\mathbf{k}}}{\pi^{2\mathbf{k}}\widetilde{\omega}^{2\mathbf{k}}}\sum_{j=j_*}^{[\widetilde{\omega}]} \mathbf{a}_j\,\theta_j^2\,.$$

Similarly, for sufficiently large n, the term $\mathbf{G}_{2,n}^*$ can be estimated as

$$\mathbf{G}_{2,n}^* \leq \frac{1}{\mathbf{a}_{[\widetilde{\omega}]+1}}\sum_{j\geq[\widetilde{\omega}]+1} \mathbf{a}_j\,\theta_j^2 \leq (1+\widetilde{\varepsilon})\frac{(b-a)^{2\mathbf{k}}}{\pi^{2\mathbf{k}}\widetilde{\omega}^{2\mathbf{k}}}\sum_{j\geq[\widetilde{\omega}]+1} \mathbf{a}_j\,\theta_j^2\,,$$

i.e.

$$\mathbf{G}_n^* \le (1+\widetilde{\varepsilon})\frac{(b-a)^{2\mathbf{k}}}{\pi^{2\mathbf{k}}\widetilde{\omega}^{2\mathbf{k}}}\sum_{j\ge 1}\mathbf{a}_j\,\theta_j^2 \le (1+\widetilde{\varepsilon})\frac{(b-a)^{2\mathbf{k}}}{\pi^{2\mathbf{k}}\widetilde{\omega}^{2\mathbf{k}}}\,\mathbf{r}\,.$$

Furthermore, we have that uniformly in $S\in\Theta_{\varepsilon,L}$

$$\lim_{n\to\infty}\frac{1}{\widetilde{\omega}}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)=\lim_{n\to\infty}\frac{1}{\widetilde{\omega}}\sum_{j=j_*}^{\widetilde{\omega}}\left(1-(j/\widetilde{\omega})^{\mathbf{k}}\right)^2=\int_0^1(1-u^{\mathbf{k}})^2\mathrm{d}u=\frac{2\mathbf{k}^2}{(\mathbf{k}+1)(2\mathbf{k}+1)}:=\iota_{\mathbf{k}}\,.$$

Therefore, taking into account that $\lim_{n\to\infty}\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}}n^{2\mathbf{k}/(2\mathbf{k}+1)}\widetilde{\mathbf{G}}_n=0$ for $\mathbf{k}\ge 2$, we obtain, that for sufficiently large n and uniformly in $S\in\Theta_{\varepsilon,L}$

$$\Upsilon_n(S)\le(1+\widetilde{\varepsilon})\Psi(\widetilde{\omega})\quad\text{and}\quad\Psi(v)=\frac{(b-a)^{2\mathbf{k}}}{\pi^{2\mathbf{k}}v^{2\mathbf{k}}}\,\mathbf{r}+\frac{\varsigma_*\iota_{\mathbf{k}}}{n}\,v\,.$$

Note that

$$\inf_{v\in\mathbb{R}}\Psi(v)=\Psi(v_*)=(\varsigma_*(b-a))^{2\mathbf{k}/(2\mathbf{k}+1)}\,\mathbf{l}_{\mathbf{k},\mathbf{r}}$$

and

$$v_*=(b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\left(d_{\mathbf{k}}^*\widetilde{r}(S)\,n\right)^{1/(2\mathbf{k}+1)}\,.$$

Then remind that $\widetilde{\omega}=\widetilde{\omega}_*+(b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\left(d_{\mathbf{k}}^*\widetilde{\mathbf{t}}_n\,n\right)^{1/(2\mathbf{k}+1)}$. Therefore, the limit property (8.5) implies the upper bound (8.4). □

Theorem 8.1. *The estimator $\widetilde{S}$ constructed on the trigonometric basis satisfies the following asymptotic upper bound*

$$\limsup_{n\to\infty}n^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}}v(S)\,\mathbf{E}_{\mathbf{p},S}\|\widetilde{S}-S\|_d^2\,\mathbf{1}_\Gamma\le(b-a)^{2\mathbf{k}/(2\mathbf{k}+1)}\mathbf{l}_{\mathbf{k},\mathbf{r}}\,.\tag{8.6}$$

Proof. Now we recall that the Fourier coefficients on the set Γ

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

Hence, 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 $M_n=\sqrt{b-a}\sum_{j=1}^{d}(1\;-\;\widetilde{\lambda}(j))\,\widetilde{\lambda}(j)\theta_{j,d}\,\eta_{j,d}/\sqrt{d}$. Now, for any $0<\varepsilon_1<1$,

$$\begin{aligned}2\left|\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))\widetilde{\lambda}(j)\theta_{j,d}\,\varpi_{j,d}\right|&\le\varepsilon_1\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2+\varepsilon_1^{-1}\sum_{j=1}^{d}\varpi_{j,d}^2\\&\le\varepsilon_1\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2+\frac{(b-a)\varpi_n^*}{\varepsilon_1 n}\,,\end{aligned}$$

where ϖ_n^* is defined in (3.6). Therefore,

$$\|\widetilde{S} - S\|_d^2 \le (1+\varepsilon_1)\sum_{j=1}^{d}(1-\widetilde{\lambda}(j))^2\theta_{j,d}^2 - 2M_n + \frac{\varpi_n^*}{\varepsilon_1 n} + \sum_{j=1}^{d}\widetilde{\lambda}^2(j)\zeta_{j,d}^2\,.$$

By the same way we get

$$\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\zeta_{j,d}^2 \le \frac{(1+\varepsilon_1)(b-a)}{d}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\,\eta_{j,d}^2 + (1+\varepsilon_1^{-1})\frac{\varpi_n^*}{n}$$

and, then, for any $0<\varepsilon_1<1$ on the set Γ

$$\|\widetilde{S}_n - S\|_d^2 \le (1+\varepsilon_1)\Upsilon_n(S) - 2M_n + (1+\varepsilon_1)\mathbf{U}_n + \frac{3\varpi_n^*}{\varepsilon_1 n}\,,$$

where $\Upsilon_n(S)$ is defined in (8.3),

$$\mathbf{U}_n = \frac{1}{d^2}\sum_{j=1}^{d}\widetilde{\lambda}^2(j)\left(d(b-a)\eta_{j,d}^2 - \varsigma_*\right) \tag{8.7}$$

and the variance ς_* is defined in (4.2). 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 $\sigma_{1,*}$ is given in (2.16). Moreover, using that $\mathbf{E}_{\mathbf{p},S}\,M_n = 0$, we get

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

Consequently, Proposition 2.2 yields

$$\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\Theta_{\varepsilon,L}} |\mathbf{E}_{\mathbf{p},S}\,M_n\,\mathbf{1}_\Gamma| = 0\,. \tag{8.8}$$

Now, the property (3.7), Proposition 8.1 and Lemma A.8 imply the inequality (8.6). □
It is clear that Theorem 3.1 and Theorem 8.1 imply Theorem 4.2.

Conclusion

In Conclusion, we focus on the methodological and theoretical investments of this paper in the statistical treatment of the big data model with dependent observations.

- First of all, it should be emphasized that providing an efficient estimation for dependent observations autoregressive models is possible only on the basis of the sequential analysis approach through the transition from the observations (1.2) to the regression scheme (2.14).

- In this paper for the first time for the model (1.2), the sharp lower bound (4.3) for the weighted robust quadratic risk is obtained. It turns out that through the Fisher information $\mathcal{I}(S)$ defined in Remark 3, this bound can be represented as the nonparametric version of the Rao-Cramer inequality (4.5). It should also be noted that we cannot use here the usual methods to obtain the lower bound since the functions (7.1) must be in the stable class (2.4), i.e. we cannot use Gaussian random variables in the representation (7.1), as it is usually done.

- To show the efficiency properties (4.4), we generally use the adaptive estimation approach proposed in the papers of Galtchouk and Pergamenchtchikov for the diffusion processes. It should be emphasized that in all these papers, the main results are based on the properties of geometric ergodicity, which cannot be used for the model (1.2). That is the reason why we develop, for this model, special efficient estimation methods. In Section 7, to obtain a lower bound for the asymptotic risk, we construct an analytical tool based on a special version of the van Trees inequality given in Proposition 6.1. In Section 8 we develop special asymptotic methods to obtain the sharp upper bound for known regularity parameters $\mathbf{k}$ and $\mathbf{t}$ of the Sobolev ellipse (4.1). Finally, through the oracle inequality from [Arkoun et al. (2019)], we obtain the asymptotic efficiency in the adaptive setting.

A Appendix

A.1 Properties of the prior distribution 7.3

In this section we study properties of the distribution used in (7.2).

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 $\int_{-\mathbf{N}}^{\mathbf{N}} z\rho_{\mathbf{N}}(z)\mathrm{d}z = 0$ and, moreover, $\int_{-\mathbf{N}}^{\mathbf{N}} z^2\rho_{\mathbf{N}}(z)\mathrm{d}z \to 1$ and $J_{\mathbf{N}} = \int_{\mathbb{R}} (\dot{\rho}_{\mathbf{N}}(z))^2/\rho_{\mathbf{N}}(z)\,\mathrm{d}z \to 1$ as $\mathbf{N} \to \infty$.*

Proof. First we set $V(z) = \left(\int_{-1}^{1} e^{-\frac{1}{1-t^2}}\mathrm{d}t\right)^{-1} e^{-\frac{1}{1-z^2}}\,\mathbf{1}_{\{|z|\leq 1\}}$. 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 $\int_{-1}^{1} V(z)\mathrm{d}z = 1$. Now for $\mathbf{N} \geq 2$ we set $\chi_{\mathbf{N}}(z) = \int_{-1}^{1} \mathbf{1}_{\{|z+u|\leq N-1\}} V(u)\mathrm{d}u = \int_{\mathbb{R}} \mathbf{1}_{\{|t|\leq N-1\}} V(t-z)\mathrm{d}t$. Using here the properties of the function V 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 $\dot{\chi}_{\mathbf{N}}(z) = -\int_{\mathbb{R}} \mathbf{1}_{\{|t|\leq N-1\}} \dot{V}(t-z)\mathrm{d}t = -\int_{-1}^{1} \mathbf{1}_{\{|u+z|\leq N-1\}} \dot{V}(u)\mathrm{d}u$. Note here that $|\dot{V}(z)| \leq \mathbf{c}_*\sqrt{V(z)}$ for some $\mathbf{c}_* > 0$. Now through the Bunyakovsky - Cauchy - Schwartz inequality we get that $\dot{\chi}^2_{\mathbf{N}}(z) \leq 2\mathbf{c}_*\chi_{\mathbf{N}}(z)$ for $|z| < \mathbf{N}$. Now we set $\rho_{\mathbf{N}}(z) = \left(\int_{-\mathbf{N}}^{N} \varphi(t)\chi_{\mathbf{N}}(t)\mathrm{d}t\right)^{-1} \varphi(z)\chi_{\mathbf{N}}(z)$, where $\varphi(z)$ the $(0,1)$-Gaussian density, It is clear that $\rho_{\mathbf{N}}(z)$ is the the continuously differentiable probability density with the support $[-\mathbf{N}, \mathbf{N}]$ such that for any $\mathbf{N}$ the integral $\int_{-\mathbf{N}}^{\mathbf{N}} z\rho_{\mathbf{N}}(z)\mathrm{d}z = 0$ and $\int_{-\mathbf{N}}^{N} \varphi(t)\chi_{\mathbf{N}}(t)\mathrm{d}t \to 1$, $\int_{-\mathbf{N}}^{\mathbf{N}} z^2\rho_{\mathbf{N}}(z)\mathrm{d}z \to 1$ for $\mathbf{N} \to \infty$. Moreover, the

Fisher information can be represented as

$$J_N = \left(\int_{-\mathbf{N}}^{\mathbf{N}} \varphi(t) \chi_{\mathbf{N}}(t) \mathrm{d}t \right)^{-1} \left(\int_{\mathbb{R}} \frac{\dot{\varphi}^2(z)}{\varphi(z)} \chi_{\mathbf{N}}(z) \mathrm{d}z + \Delta_{\mathbf{N}} \right) ,$$

where, taking into account that $\dot{\chi}_{\mathbf{N}}(z) = 0$ for $|z| \le \mathbf{N} - 2$,

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

Therefore, $\Delta_{\mathbf{N}} \to 0$ as $\mathbf{N} \to \infty$. Hence Lemma A.1. □

Lemma A.2. *The term* (7.8) *is such that* $\lim_{n\to\infty} \max_{1\le j\le d} |\overline{\Psi}_{n,j} - 1| = 0$.

Proof. First, note that $y_l = y_0 \prod_{i=1}^{l} S_\kappa(x_i) + \sum_{\iota=1}^{l} \prod_{i=\iota+1}^{l} S_\kappa(x_i) \xi_\iota$ for $l \ge 1$. Therefore, $\widetilde{\mathbf{E}}_0 \, y_l^2 = y_0^2 \, \widetilde{\mathbf{E}}_0 \prod_{i=1}^{l} S_\kappa^2(x_i) + \sum_{\iota=1}^{l-1} \widetilde{\mathbf{E}}_0 \prod_{i=\iota+1}^{l} S_\kappa^2(x_i) + 1$ and due to (7.5) we obtain that for any $n \ge 1$ for which $\delta_n^* < 1$

$$\sup_{l\ge 1} |\widetilde{\mathbf{E}}_0 \, y_l^2 - 1| \le (\delta_n^*)^2 \frac{(y_0^2 + 1)}{1 - (\delta_n^*)^2} \to 0 \quad \text{as} \quad n \to \infty \, .$$

Since $(b-a) \sum_{l=1}^{n} \phi_j^2(x_l) = n$, we get Lemma A.2. □

Lemma A.3. *The term* $\mathbf{R}_{0,n}$ *in* (7.7) *is such that* $\lim_{n\to\infty} n^{\mathbf{b}} \, \mathbf{R}_{0,n} = 0$ *for any* $\mathbf{b} > 0$ *and* $0 < \rho < 1$.

Proof. First note that taking into account in the definition of term $\mathbf{R}_{0,n}$ in (7.7), that $|\eta_j^*| \le \ln n$, we get that $\mathbf{R}_{0,n} \le \left(\mathbf{r} + \ln^2 n \, \sum_{j=1}^{d} s_j^2 \right) \mu_\kappa(\Gamma^c)$. Therefore, to show this lemma it suffices to check that $\lim_{n\to\infty} n^{\mathbf{b}} \, \mu_\kappa(\Gamma^c) = 0$ for any $\mathbf{b} > 0$. To do this note that the definition of d in (7.6) implies $\mu_\kappa(\Gamma^c) \le \mathbf{P}(\zeta_n > \mathbf{r})$ and $\zeta_n = \sum_{j=1}^{d} \mathbf{a}_j \kappa_j^2$. So, it suffices to show that

$$\lim_{n\to\infty} n^{\mathbf{b}} \, \mathbf{P}(\zeta_n > \mathbf{r}) = 0 \quad \text{for any} \quad \mathbf{b} > 0 \, . \tag{A.1}$$

Indeed, first note that the definition (7.3) through Lemma A.1 and the property (7.4) imply directly

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

Setting then $\widetilde{\zeta}_n = \zeta_n - \mathbf{E} \, \zeta_n = (b-a) \sum_{j=1}^{d} s_j^* \, \mathbf{a}_j \widetilde{\eta}_j / n$ and $\widetilde{\eta}_j = (\eta_j^*)^2 - \mathbf{E}(\eta_j^*)^2$, we get for large n that $\{\zeta_n > \mathbf{r}\} \subset \left\{ \widetilde{\zeta}_n > \mathbf{r}_1 \right\}$ for $\mathbf{r}_1 = \mathbf{r}(1-\rho)/2$. Now the correlation inequality from [Galtchouk and Pergamenshchikov (2013)] and the bound $|\widetilde{\eta}_j| \le 2 \ln^2 n$ imply that for any $p \ge 2$ there exists some constant $C_p > 0$ for which

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

i.e. the expectation $\mathbf{E} \, \widetilde{\zeta}_n^p \to 0$ as $n \to \infty$ and, therefore, $n^{\mathbf{b}} \mathbf{P}(\widetilde{\zeta}_n > \mathbf{r}_1) \to 0$ as $n \to \infty$ for any $\mathbf{b} > 0$. This implies (A.1) and hence Lemma A.3. □

A.2 Properties of the trigonometric basis.

First we need the following lemma from [Konev and Pergamenshchikov (2015)].

Lemma A.4. *Let f be an absolutely continuous function, $f:[a,b]\to\mathbb{R}$, with $\|\dot f\|<\infty$ and g be a piecewise constant function $[a,b]\to\mathbb{R}$ of a form $g(x)=\sum_{j=1}^{d} c_j\,\chi_{(z_{j-1},z_j]}(x)$ where c_j are some constants. Then for any $\varepsilon>0$, the function $\Delta=f-g$ satisfies the following inequalities*

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

Lemma A.5. *For any $1\le j\le d$ the trigonometric Fourier coefficients $(\theta_{j,d})_{1\le j\le d}$ for the functions S from the class $\mathcal{W}_{\mathbf{k},\mathbf{r}}$ with $\mathbf{k}\ge 1$ satisfy, for any $\widetilde\varepsilon>0$, the following inequality $\theta_{j,d}^2\le(1+\widetilde\varepsilon)\,\theta_j^2+4\mathbf{r}(1+\widetilde\varepsilon^{-1})(b-a)^{2\mathbf{k}}d^{-2k}$.*

Proof. First we represent the function S as $S(x)=\sum_{l=1}^{d}\theta_l\,\phi_l(x)+\Delta_d(x)$ and $\Delta_d(x)=\sum_{l>d}\theta_l\,\phi_l(x)$, i.e. $\theta_{j,d}=(S,\phi_j)_d=\theta_j+(\Delta_d,\phi_j)_d$ and, therefore, $\forall\widetilde\varepsilon>0$ we get $\theta_{j,d}^2\le(1+\widetilde\varepsilon)\theta_j^2+(1+\widetilde\varepsilon^{-1})\|\Delta_d\|_d^2$. Lemma A.4 with $g=0$ implies $\|\Delta_d\|_d^2\le 2\|\Delta_d\|^2+2(b-a)^2d^{-2}\|\dot\Delta_d\|^2$. Using here that $2\pi[l/2]\ge l$ for $l\ge 2$, we obtain that $\|\Delta_d\|^2=\sum_{l>d}\theta_l^2\le\mathbf{r}/\mathbf{a}_d\le(b-a)^{2\mathbf{k}}\mathbf{r}d^{-2\mathbf{k}}$ and

$$\|\dot\Delta_d\|^2=(2\pi/(b-a))^2\sum_{l>d}\theta_l^2\,[l/2]^2\le(b-a)^{2(\mathbf{k}-1)}d^{-2(\mathbf{k}-1)}\,. \tag{A.2}$$

Hence Lemma A.5 □

Lemma A.6. *For any $d\ge 2$ and $1\le N\le d$ the coefficients $(\theta_{j,d})_{1\le j\le d}$ of functions S from the class $\mathcal{W}_{\mathbf{r},\mathbf{k}}$ with $\mathbf{k}\ge 1$ satisfy, for any $\widetilde\varepsilon>0$, the inequality $\sum_{j=N}^{d}\theta_{j,d}^2\le(1+\widetilde\varepsilon)\sum_{j\ge N}\theta_j^2+(1+\widetilde\varepsilon^{-1})(b-a)^{2\mathbf{k}}\mathbf{r}\,d^{-2}N^{-2(\mathbf{k}-1)}$.*

Proof. Note that $\sum_{j=N}^{d}\theta_{j,d}^2=\min_{x_1,\dots,x_{N-1}}\|S-\sum_{j=1}^{N-1}x_j\phi_j\|_d^2\le\|\Delta_N\|_d^2$ and $\Delta_N(t)=\sum_{j\ge N}\theta_j\phi_j(t)$. Lemma A.4 and (A.2), imply Lemma A.6 □

A.3 Technical lemmas

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

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

where the coefficient $\sigma_{1,*}$ is given in (2.16).

Proof. Using the definition of $\eta_{j,d}$ in (3.5) and the bounds (2.16), we get

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

Now, the orthonormality property (3.2) implies this lemma. □

Lemma A.8. *For the sequence (8.7) the following limit property holds true*

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

Proof. First of all, note that, using the definition of $\mathbf{s}_{j,d}$ in (3.11), we obtain

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

where $\overline{\mathbf{s}}_{j,d} = \sum_{l=1}^{d} \sigma_l^2 \overline{\phi}_j(x_l)$ and $\overline{\phi}_j(z) = (b-a)\phi_j^2(z) - 1$. 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 $\|\widetilde{\lambda}\|^2 = \sum_{j=1}^{d} \widetilde{\lambda}^2(j)$, $\mathbf{U}_{1,n} = (b-a)\sum_{l=1}^{d} H_l^{-1} - \varsigma_*$ and $\mathbf{U}_{2,n} = \sum_{j=1}^{d} \widetilde{\lambda}^2(j)\overline{\mathbf{s}}_{j,d}$. Note now that using Proposition 2.12 and the dominated convergence theorem in the definition (2.6) we obtain that

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

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

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

i.e. $\lim_{n\to\infty} \sup_{S\in\Theta_{\varepsilon,L}} \sup_{\mathbf{p}\in\mathcal{P}} |\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_{1,n}| = 0$. Therefore, taking into account that

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

we obtain that

$$\overline{\lim}_{n\to\infty}\, n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\Theta_{\varepsilon,L}} \sup_{\mathbf{p}\in\mathcal{P}} |\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n| \le (b-a)\overline{\lim}_{n\to\infty}\, \mathbf{U}^*_{2,n}\,, \tag{A.3}$$

where

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

Now, using Lemma A.2 from [Galtchouk and Pergamenshchikov (2009a)] we get

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

From the definition of $\sigma_{1,*}$ in (2.16) it follows that $\limsup_{n\to\infty} d\,\sigma_{1,*} < \infty$, i.e.

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

Therefore, the bound (A.3) implies $\overline{\lim}_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\Theta_{\varepsilon,L}} \sup_{\mathbf{p}\in\mathcal{P}} |\mathbf{E}_{\mathbf{p},S}\,\mathbf{U}_n| = 0$. Using the inequality (A.4) from [Arkoun et al. (2019)] we get $\mathbf{E}_{\mathbf{p},S}\eta_{j,d}^4 \le 64\mathbf{v}^*\sigma_{1,*}^2$, where the coefficient $\mathbf{v}^*$ is given in (2.15). From this we obtain, that

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

Thus, Proposition 2.2 yields $\lim_{n\to\infty} n^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{k},\mathbf{r}}} \mathbf{E}_{\mathbf{p},S}|\mathbf{U}_n|\mathbf{1}_{\Gamma^c} = 0$. □

References

[Amderson1970] Anderson, T.W. : *The Statistical Analysis of Time Series,* New York : *Wiley.*, 1994.

[Arkoun and Pergamenchtchikov (2008)] Arkoun, O. and Pergamenchtchikov, S. (2008) : Nonparametric Estimation for an Autoregressive Model, *Vestn. Tomsk. Gos. Univ. Mat. Mech.*. 2: 20-30.

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

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

[Arkoun et al. (2019)] Arkoun, O. , Brua, J.-Y. and Pergamenchtchikov, S. (2019) Sequential Model Selection Method for Nonparametric Autoregression. - *Sequential Analysis*, **38** (4), 437 - 460.

[Bayisa et al. (2019)] Bayisa, F.L., Zhou, Z., Cronie, O. and Yu, J. (2019) Adaptive algorithm for sparse signal recovery. *Digital Signal Processing*, **87**, 10 – 18.

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

[Beltaief et al. (2020)] Beltaief, S., Chernoyarov, O. and Pergamenchtchikov, S. (2020) Model selection for the robust efficient signal processing observed with small Levy noise. - *Ann Inst Stat Math.*, **72**, 1205 - 1235.

[De Souza and al. (2019)] De Souza, D.P., Kuhn, E.V., Seara, R. (2019) A Time-Varying Autoregressive Model for Characterizing Nonstationary Processes. - *IEEE Signal Processing Letters*, **26** (1), 134 - 138.

[Fan and Zhang (2008)] Fan, J. and Zhang, W. (2008) Statistical Methods with Varying Coefficient Models. -*Statistics and Its Interface*, **1**, 179–195.

[Galtchouk and Pergamenshchikov (2009a)] 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.

[Galtchouk and Pergamenshchikov (2009b)] 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.

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

[Galtchouk and Pergamenshchikov (2013)] 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.

[Galtchouk and Pergamenshchikov (2022)] Galtchouk, L. and Pergamenshchikov, S. (2022) Adaptive efficient analysis for big data ergodic diffusion models. - *Statistical inference for stochastic processes*, **25**, 127 – 158.

[Hastie et al.(2008)] Hastie, T., Friedman, J. and Tibshirani, R. The Elements of Statistical Leaning. Data Mining, Inference and Prediction.// Second Edition, Springer, Springer series in Statistics, - 2008, 764 p.

[Konev and Pergamenshchikov (2009)] Konev, V.V. and Pergamenshchikov, S. M. (2009) Non-parametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. - *Vestn. Tomsk. Gos. Univ. Mat. Mech.*, 4(8), 31-45.

[Konev and Pergamenshchikov (2012)] 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.

[Konev and Pergamenshchikov (2015)] 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.

[Li and al. (2022)] Li, Z., Wu, X., Wang, X., Zhong, H., Chen, J. and Xu Ma, X. (2022) Measurement and Analysis of Contribution Rate for China Rice Input Factors via a Varying-Coefficient Production Function Model. - *Agriculture*, 1**12**, 1431.

[Luo et al. (2009)] Luo, X. H., Yang, Z. H. and Zhou, Y. (2009) Nonparametric Estimation of the Production Function with Time-Varying Elasticity Coefficients, *Systems Engineering Theory & Practice* 29: 144-149.

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

[Niedzwiecki (2000)] Niedzwiecki, M. Identification of Time-Varying Processes.// Wiley : New York, NY, USA, - 2000, 344 p.

[Orbe and al. (2006)] Orbe, S. Ferreira, E. and Rodriguez-Poo, J. (2006) On the estimation and testing of time varying constraints in econometric models. - *Statistica Sinica*, **16**, 1313 - 1333.

[Pchelintsev et al. (2021)] Pchelintsev, E., Pergamenshchikov, S.M. and Povzun, M. (2022) Efficient estimation methods for non-Gaussian regression models in continuous time. - *Annals of the Institute of Statistical Mathematics*, **74**, 113 - 142.

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