



HAL
open science

Sequential Estimation for Nonparametric Autoregressive Models

Ouerdia Arkoun, Jean-Yves Brua, Sergei Pergamenshchikov

► **To cite this version:**

Ouerdia Arkoun, Jean-Yves Brua, Sergei Pergamenshchikov. Sequential Estimation for Nonparametric Autoregressive Models. *Statistical Topics and Stochastic Models for Dependent Data - Applications in Reliability, Survival Analysis and Related Fields*, In press. hal-02366193

HAL Id: hal-02366193

<https://hal.science/hal-02366193>

Submitted on 15 Nov 2019

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.

Sequential Estimation for Nonparametric Autoregressive Models

1.1. Introduction

The tradition of considering the problem of statistical estimation as that of estimating a finite number of parameters goes back to Fisher. Statistical models that explain the data more deeply are usually more complex: the unknowns of these models are, in general, some functions with certain properties of regularity. The problem of nonparametric estimation consists in estimating, from the observations, an unknown function belonging to a certain rather large functional class.

The theory of nonparametric estimation has been developed considerably in the last three decades, focusing on a some key themes, in particular the study of optimality of estimators and adaptive estimation. We are interested in particular in the optimality of the estimators when the sample size tends to infinity, such estimators are called asymptotically efficient. Numerous problems of asymptotic efficiency have been studied over the last forty years, both in a parametric and nonparametric framework and for different models.

We have focused here on showing the asymptotic efficiency of some kernel estimators for the following autoregressive model:

$$y_k = S(x_k)y_{k-1} + \xi_k, \quad 1 \leq k \leq n, \quad [1.1.1]$$

where $S(\cdot)$ is an unknown function, $x_k = a + \frac{k}{n}(b - a)$, the initial value y_0 is a constant and the noise $(\xi_k)_{k \geq 1}$ is i.i.d. sequence of unobservable random variables with $\mathbb{E}\xi_1 = 0$ and $\mathbb{E}\xi_1^2 = 1$.

The model [1.1.1] is a generalization of an autoregressive processes of the first order. It should be noted that the varying coefficient principle is well known in the regression analysis. It permits the use of more complex forms for regression coefficients and, therefore, the models constructed via this method are more adequate for applications (see, for example, Fan and Zhang (2008), Luo and Zhou (2009)). There is a number of papers which study these varying coefficient autoregressive models. In Dahlhaus (1996b), the process [1.1.1] is considered with the function S having a parametric form. Moreover, Dahlhaus (1996a) studied spectral properties of the stationary process [1.1.1] with the nonparametric function S . Belitser (2000) considered this model with Lipschitz conditions and proposed a recursive estimator. (The author established the convergence rate for quadratic risk). In all these papers, the authors propose some asymptotic (as $n \rightarrow \infty$) methods for different identification studies without considering optimal issues (in the minimax sense).

To our knowledge, for the first time, the minimax estimation problem for model [1.1.1] has been treated in Moulines *et al.* (2005) in the non adaptive case, i.e. when the regularity of S is known. More specifically, it was shown that the convergence rate is optimal for the quadratic risk by using a recursive method for autoregressive model of order d .

This chapter talks about the development of an adaptive sequential procedure for the estimation of the unknown functional coefficient in model [1.1.1]. The minimax nonadaptive and adaptive convergence rates are established in a Hölderian case and an asymptotically efficient estimator is constructed with the absolute error risk. Then, in a general framework, we propose a new sequential model selection procedure for this model which provides a sharp oracle inequality.

We assume that in the model [1.1.1] the i.i.d. random variables $(\xi_k)_{1 \leq k \leq n}$ have a density p (with respect to the Lebesgue measure) from the functional class \mathcal{P}_ς defined as

$$\mathcal{P}_\varsigma := \left\{ p \geq 0 : \int_{-\infty}^{+\infty} p(x) dx = 1, \int_{-\infty}^{+\infty} x p(x) dx = 0, \right. \\ \left. \int_{-\infty}^{+\infty} x^2 p(x) dx = 1, \sup_{k \geq 1} \frac{1}{\varsigma^k (2k-1)!!} \int_{-\infty}^{+\infty} |x|^{2k} p(x) dx \leq 1 \right\} \quad [1.1.2]$$

where $\varsigma \geq 1$ is some fixed parameter.

To obtain the stable (uniformly with respect to the function S) model [1.1.1], we assume that for some fixed $0 < \varepsilon < 1$ and $L > 0$ the unknown function S belongs to the ε -stability set

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

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

1.2. Pointwise estimation with absolute error risk

This section is devoted to nonparametric estimation for autoregressive models. We consider the problem of estimating an unknown function at a fixed point using data governed by autoregressive models defined in [1.1.1]. To define the risk associated with the use of an estimator and thus measure the quality of it, we use the loss function related to the absolute error. This work follows the minimax approach for which the goal is to find a lower bound of the asymptotic minimax risk and then to construct an estimator, said asymptotically efficient, for which the maximum risk reaches asymptotically this bound.

For a nonparametric autoregressive model where the autoregressive function is supposed to belong to a weak Hölder class with known regularity, we show that a kernel estimator is asymptotically efficient. When the regularity of the autoregressive function is unknown, we get the minimax adaptive convergence rate of estimators on a family of Hölderian classes. Moreover in the same case, we can construct through a sequential procedure an asymptotically efficient estimator over the same weak Hölder class.

For the model [1.1.1], we propose to estimate the unknown function S belonging to a certain Hölder class $\mathcal{H}^{(\beta)}$ at a fixed point z_0 , where β is the regularity parameter of this functional class. In terms of asymptotic efficiency, we are now able to define the maximum risk of an estimator and to describe the minimax approach used for this purpose.

1.2.1. Minimax approach

Let us begin by giving precisely the definition of an estimator for the considered model [1.1.1].

DEFINITION 1.1. – *For the autoregressive model [1.1.1], an estimator of S at point z_0 is a random variable $\omega \mapsto \tilde{S}_n = \tilde{S}_n(z_0, y_1, \dots, y_n)$ measurable with respect to the σ -algebra engendered by y_1, \dots, y_n .*

We define the risk of an estimator \tilde{S}_n of function S belonging to a functional class $\mathcal{H}^{(\beta)}$ of regularity β for a fixed z_0 by $\mathbf{E}_S |\tilde{S}_n - S(z_0)|$, where \mathbf{E}_S is the expectation when the randomness is determined by the model [1.1.1].

In our framework we consider the case for which the maximum risk of an estimator \tilde{S}_n is defined by

$$\mathcal{R}_n(\tilde{S}_n) := \sup_{S \in \mathcal{H}^{(\beta)}} \sup_{p \in \mathcal{P}} \mathbf{E}_{S,p} |\tilde{S}_n - S(z_0)|.$$

We are therefore led to consider the asymptotic behavior of minimax risk defined by

$$\mathcal{R}_n^* = \inf_{\tilde{S}_n} \mathcal{R}_n(\tilde{S}_n),$$

the infimum being taken on all the estimators.

The primary goal of the minimax approach is to find an estimator \hat{S}_n whose maximum risk is equal to the minimax risk. Such an estimator is called minimax. An estimator \hat{S}_n is said asymptotically efficient if

$$\mathcal{R}_n(\hat{S}_n) \underset{n \rightarrow \infty}{\sim} \mathcal{R}_n^*.$$

The goal of the approach is to find an estimator S_n^* , a sequence $(\varphi_n)_{n \in \mathbb{N}^*}$ strictly positive reals such as $\varphi_n \rightarrow +\infty$, when $n \rightarrow +\infty$ and constants $c > 0$ and $C < \infty$ such as

$$\limsup_{n \rightarrow \infty} \varphi_n \mathcal{R}_n(S_n^*) \leq C \text{ and } \liminf_{n \rightarrow \infty} \varphi_n \mathcal{R}_n^* \geq c. \quad [1.2.1]$$

DEFINITION 1.2.– *The family $(\varphi_n)_{n \in \mathbb{N}^*}$ is called minimax convergence rate of estimators over $\mathcal{H}^{(\beta)}$ if [1.2.1] is verified.*

DEFINITION 1.3.– *An estimator S_n^* satisfying*

$$c \leq \liminf_{n \rightarrow \infty} \varphi_n \mathcal{R}_n^* \leq \limsup_{n \rightarrow \infty} \varphi_n \mathcal{R}_n(S_n^*) \leq C$$

where $(\varphi_n)_{n \in \mathbb{N}^}$ is the minimax convergence rate and $c > 0$ and $C < \infty$ are constants, is said optimal estimator in convergence rate over $\mathcal{H}^{(\beta)}$.*

REMARK 1.1.– *To show the asymptotic efficiency of an estimator, it suffices to obtain the same lower and upper bounds ($C = c$ in Definition 1.3).*

1.2.2. Adaptive minimax approach

The minimax approach is called adaptive when one of the parameters defining the considered functional class is assumed to be unknown, for example the regularity of the autoregressive function S in the model [1.1.1]. For the Höderian class $\mathcal{H}^{(\beta)}$, we

assume that $\beta \in \mathcal{B}$, where \mathcal{B} is an interval and define the adaptive risk of an estimator \tilde{S}_n of $S(z_0)$ as

$$\mathcal{R}_{n,\beta}(\tilde{S}_n) = \sup_{S \in \mathcal{H}^{(\beta)}} \sup_{p \in \mathcal{P}} \mathbf{E}_S |\tilde{S}_n - S(z_0)|.$$

Then, as for the nonadaptive minimax approach, we look for an adaptive (independent of $\beta \in \mathcal{B}$) and asymptotically efficient estimator, as well as the exact asymptotic bound of the adaptive minimax risk

$$\inf_{\tilde{S}_n} \sup_{\beta \in \mathcal{B}} \mathcal{R}_{n,\beta}(\tilde{S}_n).$$

The question we are also asking is the existence of an optimal adaptive estimator in convergence rate, i.e an estimator, independent of $\beta \in \mathcal{B}$, which converges at this rate on each class $\mathcal{H}^{(\beta)}$. More precisely:

DEFINITION 1.4.– *The family $(\phi_n(\beta))_{n \in \mathbb{N}^*}$ is called adaptive minimax rate of convergence of estimators on the family of classes $(\mathcal{H}^{(\beta)})_{\beta \in \mathcal{B}}$ if*

– *for a certain estimator S_n^* and a constant $C > 0$, we have:*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \phi_n(\beta) \mathcal{R}_{n,\beta}(S_n^*) \leq C;$$

– *and there exists a constant $c > 0$ such that:*

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{S}_n} \sup_{\beta \in \mathcal{B}} \phi_n(\beta) \mathcal{R}_{n,\beta}(\tilde{S}_n) \geq c.$$

An estimator S_n^* checking the first previous point, with $\phi_n(\beta)$ the adaptive minimax convergence rate is said adaptive optimal in rate of convergence on the family $(\mathcal{H}^{(\beta)})_{\beta \in \mathcal{B}}$.

DEFINITION 1.5.– *An adaptive optimal estimator S_n^* in convergence rate is called adaptive asymptotically efficient on the family $\{\mathcal{H}^{(\beta)}\}_{\beta \in \mathcal{B}}$ if it satisfies:*

$$\lim_{n \rightarrow \infty} \inf_{\tilde{S}_n} \sup_{\beta \in \mathcal{B}} \phi_n(\beta) \mathcal{R}_{n,\beta}(\tilde{S}_n) = \lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \phi_n(\beta) \mathcal{R}_{n,\beta}(S_n^*).$$

However, optimal adaptive estimators of convergence do not always exist. Indeed, Lepskii (1990) shows that it does not exist for estimation at a fixed point, in a Gaussian white noise model, of a Hölderian function belonging to the class $\Sigma(L, \beta)$, $\beta \in \mathcal{B} \subset \mathbb{R}_+^*$ described in the definition 1.6 and \mathcal{B} containing at least two elements. Nevertheless, we can get a relationship of the type

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \phi_n(\beta) \mathcal{R}_{n,\beta}(S_n^*) \leq C,$$

for a certain estimator S_n^* , whereas $\phi_n(\beta)$ is not the minimax convergence rate on $\mathcal{H}^{(\beta)}$.

DEFINITION 1.6.— Let $L > 0$ and $\beta > 0$. The Hölder class $\Sigma(L, \beta)$ is defined by

$$\Sigma(L, \beta) = \left\{ S : \mathbb{R} \rightarrow \mathbb{R} : |S^{(m)}(x) - f^{(m)}(y)| \leq L|x - y|^{\beta-m}, \forall x, y \in \mathbb{R} \right\},$$

where $m = \lfloor \beta \rfloor$ is the largest integer strictly smaller than real β .

1.2.3. Nonadaptive procedure

This section deals with the nonparametric estimation of the autoregression coefficient function S in the model [1.1.1] at a given point z_0 , when the smoothness of S is known. For this problem we make use of the following modified kernel estimator

$$\hat{S}_n(z_0) = \frac{1}{A_n} \sum_{k=1}^n Q(u_k) y_{k-1} y_k \mathbf{1}_{(A_n \geq d)}, \quad [1.2.2]$$

where $Q(\cdot)$ is a kernel function,

$$A_n = \sum_{k=1}^n Q(u_k) y_{k-1}^2 \quad \text{with} \quad u_k = \frac{x_k - z_0}{h};$$

d and h are some positive parameters.

Our first goal is to find the convergence rate of the minimax risk taken over the *stable local Hölder class* at the point z_0

$$\begin{aligned} & \mathcal{H}^{(\beta)}(z_0, K, \varepsilon) \\ &= \left\{ S \in \Theta_{\varepsilon, L}, |\dot{S}(x) - \dot{S}(z_0)| \leq K|x - z_0|^{\beta-1}, \forall x \in [a, b] \right\}, \end{aligned} \quad [1.2.3]$$

with a known regularity $1 \leq \beta \leq 2$.

We assume that the regularity $\underline{\beta} \leq \beta \leq \bar{\beta}$, where $\underline{\beta} = 1 + \underline{\alpha}$ and $\bar{\beta} = 1 + \bar{\alpha}$ for some fixed parameters $0 \leq \underline{\alpha} < \bar{\alpha} \leq 1$.

We will find an asymptotical (as $n \rightarrow \infty$) positive lower bound for the minimax risk with the normalizing coefficient

$$\varphi_n = n^{\frac{\beta}{2\beta+1}}. \quad [1.2.4]$$

To obtain this convergence rate we set in [1.2.2]

$$h = n^{-\frac{1}{2\beta+1}} \quad \text{and} \quad d = \kappa_n n h, \quad [1.2.5]$$

where $\kappa_n \geq 0$,

$$\lim_{n \rightarrow \infty} \kappa_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{h}{\kappa_n^2} = 0. \quad [1.2.6]$$

As to the kernel function we assume that

$$\int_{-1}^1 Q(z) dz > 0 \quad \text{and} \quad \int_{-1}^1 z Q(z) dz = 0. \quad [1.2.7]$$

The two following theorems assert that the estimator [1.2.2] with the parameters [1.2.5]–[1.2.7] is asymptotically optimal with respect to the convergence rate [1.2.4] over the stable local Hölder class $\mathcal{H}^{(\beta)}(z_0, K, \varepsilon)$.

THEOREM 1.1.– For any $K > 0$ and $0 < \varepsilon < 1$

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{S}_n} \sup_{S \in \mathcal{H}^{(\beta)}(z_0, K, \varepsilon)} \varphi_n \mathcal{R}_n(\tilde{S}_n, S) > 0,$$

where the infimum is taken over all estimators.

Now we obtain an upper bound for the kernel estimator [1.2.2].

THEOREM 1.2.– For any $K > 0$ and $0 < \varepsilon < 1$ the kernel estimator [1.2.2] with the parameters [1.2.5]–[1.2.7] satisfies the following inequality

$$\limsup_{n \rightarrow \infty} \sup_{S \in \mathcal{H}^{(\beta)}(z_0, K, \varepsilon)} \varphi_n \mathcal{R}_n(\hat{S}_n, S) < \infty.$$

At the next step we study sharp asymptotic properties for the minimax estimators [1.2.2].

To this end similarly to Arkoun and Pergamenchtchikov (2008) we make use of the family of the *weak stable local Hölder classes* at the point z_0 , i.e. for any $\delta > 0$ we set

$$\mathcal{U}_{\delta, n}^{(\beta)}(z_0, \varepsilon) = \left\{ S \in \Gamma_\varepsilon : \|\dot{S}\| \leq \delta^{-1} \quad \text{and} \quad |\Omega_h(z_0, S)| \leq \delta h^\beta \right\}, \quad [1.2.8]$$

where

$$\Omega_h(z_0, S) = \int_{-1}^1 (S(z_0 + uh) - S(z_0)) du$$

and h is given in [1.2.5].

In this case we find in Theorem 1.3 a positive constant giving the exact asymptotic lower bound for the minimax risk with the normalizing coefficient [1.2.4]. Moreover, Theorem 1.4 shows that for the estimator [1.2.2] with the parameters [1.2.5]–[1.2.6] and the indicator kernel $Q = \mathbf{1}_{[-1, 1]}$ the asymptotic upper bound of the minimax risk coincides with this constant, i.e. in this case such estimators are asymptotically efficient.

THEOREM 1.3.— For any $\delta > 0$ and $0 < \varepsilon < 1$

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{S}_n} \sup_{S \in \mathcal{U}_{\delta, n}^{(\beta)}(z_0, \varepsilon)} \tau^{-1/2}(S) \varphi_n \mathcal{R}_n(\tilde{S}_n, S) \geq \mathbf{E}|\eta|,$$

where $\tau(S) = 1 - S^2(z_0)$ and η is a gaussian random variable with the parameters $(0, 1/2)$.

THEOREM 1.4.— The estimator [1.2.2] with the parameters [1.2.5]–[1.2.6] and $Q(z) = 1_{[-1, 1]}$ satisfies the following inequality

$$\limsup_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{S \in \mathcal{U}_{\delta, n}^{(\beta)}(z_0, \varepsilon)} \tau^{-1/2}(S) \varphi_n \mathcal{R}_n(\hat{S}_n, S) \leq \mathbf{E}|\eta|,$$

where $\tau(S) = 1 - S^2(z_0)$ and η is a gaussian random variable with the parameters $(0, 1/2)$.

REMARK 1.2.— One can show (see Galtchouk and Pergamenshchikov (2006a)) that for any $0 < \delta < 1$ and $n \geq 1$

$$\mathcal{H}^{(\beta)}(z_0, \delta, \varepsilon) \subset \mathcal{U}_{\delta, n}^{(\beta)}(z_0, \varepsilon).$$

This means that the “natural” normalizing coefficient for the functional class [1.2.8] is the sequence [1.2.4]. Theorem 1.3 and Theorem 1.4 extend usual the Hölder approach for the point estimation by keeping the minimax convergence rate [1.2.4].

1.2.4. Sequential kernel estimator

In the perspective of obtaining an optimal estimator in an adaptive framework (when the regularity β is unknown), similarly to Galtchouk and Pergamenshchikov (2001), we will apply in the next section the Lepskiĭ procedure to the model [1.1.1] based on the sequential kernel estimates.

Here we construct the sequential kernel estimator using the method proposed in Borisov and Konev (1977) for the parametric case. It should be noted that to apply the Lepskiĭ procedure the kernel estimators must have the tail distribution of the Gaussian type. To obtain this property one needs to use the sequential approach.

First of all, we need the first ν observations for the auxiliary estimation of $S(z_0)$:

$$\hat{S}_\nu = \frac{1}{A_\nu} \sum_{j=1}^{\nu} Q(u_j) y_{j-1} y_j, \quad A_\nu = \sum_{j=1}^{\nu} Q(u_j) y_{j-1}^2, \quad [1.2.9]$$

where the kernel $Q(\cdot)$ is the indicator function of the interval $[-1; 1]$; $u_j = (x_j - z_0)/h$ and h is some positive bandwidth. In the sequel for any $0 \leq k < m \leq n$ we set

$$A_{k,m} = \sum_{j=k+1}^m Q(u_j) y_{j-1}^2,$$

i.e. $A_\nu = A_{0,\nu}$. It is clear that to estimate $S(z_0)$ on the basis of the kernel estimate with the kernel Q we can use the observations $(y_j)_{k_* \leq j \leq k^*}$, where

$$k_* = [nz_0 - nh] + 1 \quad \text{and} \quad k^* = [nz_0 + nh].$$

Here $[a]$ is the integral part of a number a . So for the first estimation we chose ν as

$$\nu = \nu(h, \alpha) = k_* + \iota, \quad [1.2.10]$$

where

$$\iota = \iota(h, \alpha) = [\tilde{\epsilon}nh] + 1 \quad \text{and} \quad \tilde{\epsilon} = \tilde{\epsilon}(h, \alpha) = h^\alpha / \ln n.$$

Next, similarly to Arkoun (2011), we use a some kernel sequential procedure based on the observations $(y_j)_{\nu \leq j \leq n}$. The key idea from the sequential procedure proposed by Borisov and Konev (1977) is to replace the random denominator by some threshold $H > 0$ in the kernel estimator [1.2.2], which transform this estimator to a linear function of observations replacing the number of observations n by the following stopping time

$$\tau_H = \inf\{k \geq \nu + 1 : A_{\nu,k} \geq H\},$$

where $\inf\{\emptyset\} = n$ and the positive threshold H will be chosen as a positive random variable measurable with respect to the σ -field $\{y_1, \dots, y_\nu\}$. Therefore, we define on the set $\{\tau_H < n\}$

$$S_h^* = \frac{1}{H} \left(\sum_{j=\nu+1}^{\tau_H-1} Q(u_j) y_{j-1} y_j + \varkappa_H Q(u_{\tau_H}) y_{\tau_H-1} y_{\tau_H} \right), \quad [1.2.11]$$

where the correcting coefficient \varkappa_H on the set $\{A_{\nu,\tau_H} \geq H\}$ is defined as

$$A_{\nu,\tau_H-1} + \varkappa_H Q(u_{\tau_H}) y_{\tau_H-1}^2 = H$$

and $\varkappa_H = 1$ on the set $\{A_{\nu,\tau_H} < H\}$.

Now, to obtain an efficient estimate we need to use the all n observations, i.e. asymptotically for sufficiently large n the stopping time $\tau_H \approx n$. Similarly to Konev and Pergamenshchikov (1984), one can show that $\tau_H \approx \gamma(S) H$ as $H \rightarrow \infty$, where

$$\gamma(S) = 1 - S^2(z_0).$$

Therefore, to use asymptotically all observations we have to chose H as the number observations divided by $\gamma(S)$. But in our case we use $k^* - k_*$ observations to estimate $S(z_0)$. Therefore, to obtain optimal estimate we need to define H as $(k^* - k_*)/\gamma(S)$. taking into account that $k^* - k_* \approx 2nh$ and that $\gamma(S)$ is unknown we define the threshold H as

$$H = H(h, \alpha) = \rho nh, \quad \rho = \rho(h, \alpha) = \frac{2(1 - \tilde{\epsilon})}{\gamma(\tilde{S}_\nu)}, \quad [1.2.12]$$

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

$$\tilde{S}_\nu = \min(\max(\hat{S}_\nu, -1 + \varepsilon), 1 - \varepsilon).$$

In this paper we chose the bandwidth h in the following form

$$h = h(\beta) = (\kappa_n)^{\frac{1}{2\beta+1}}, \quad [1.2.13]$$

where the sequence $\kappa_n > 0$ such that

$$\kappa_* = \liminf_{n \rightarrow \infty} n\kappa_n > 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^\delta \kappa_n = 0 \quad [1.2.14]$$

for any $0 < \delta < 1$.

1.2.4.1. Property of the stopping time τ_H

PROPOSITION 1.1.– Assume that the threshold H is chosen in the form [1.2.12] and the bandwidth h satisfies the conditions [1.2.13] - [1.2.14]. Then for any $m \geq 1$

$$\limsup_{n \rightarrow \infty} \sup_{\underline{\alpha} \leq \alpha \leq \bar{\alpha}} h^{-m} \sup_{S \in \Theta_{\varepsilon, L}} \sup_{p \in \mathcal{P}_\zeta} \mathbf{P}_{S, p}(\tau_H = n) < \infty.$$

1.2.4.2. Decomposition of the estimate error

In the sequel we present the following decomposition of the estimate error on which the proof of the optimality of the sequential kernel estimator is based.

First we set

$$\check{\varkappa}_j = \mathbf{1}_{\{\tau_H \neq j\}} + \varkappa_H \mathbf{1}_{\{\tau_H = j\}}.$$

Then taking this into account we can represent the estimate error as

$$S_h^* - S(z_0) = -S(z_0) \mathbf{1}_{(\tau_H = n)} + \left(B_n(h) + \frac{1}{\sqrt{H}} \zeta_n(h) \right) \mathbf{1}_{(\tau_H < n)},$$

whith the approximative terms

$$B_n(h) = \frac{\sum_{j=\nu+1}^{\tau_H} \check{\varkappa}_j Q(u_j) (S(x_j) - S(z_0)) y_{j-1}^2}{H}$$

and

$$\zeta_n(h) = \frac{\sum_{j=\nu+1}^{\tau_H} \check{x}_j Q(u_j) y_{j-1} \xi_j}{\sqrt{H}}.$$

Recall that the nonsequential kernel estimator [1.2.2] does not have the above property in the case of the model [1.1.1]. Thus, in this case, the adaptive pointwise estimation is possible only in the sequential framework. In the following lemma, we obtain that the sequential kernel estimator has the the same form for the tail distribution as a Gaussian random variable.

LEMMA 1.1.– *The sequence $\zeta_n(h)$ satisfies the following properties:*

- *the sequence $\zeta_n(h)\mathbf{1}_{(\tau_H < n)}$ converges in law to $\mathcal{N}(0, 1)$ as $n \rightarrow \infty$ uniformly in $p \in \mathcal{P}_\zeta$ and $S \in \Theta_{\varepsilon, L}$,*
- *and for all $z \geq 2$*

$$\sup_{n \geq 1} \sup_{h > 0} \sup_{S \in \mathbf{C}[0,1]} \mathbf{P}_{S,p} \left(\zeta_n(h) \mathbf{1}_{(\tau_H < n)} \geq z \right) \leq 2e^{-z^2/8}.$$

1.2.5. Adaptive sequential procedure

We construct a sequential adaptive procedure for estimating the autoregressive function at a given point in nonparametric autoregression models with Gaussian noise. We make use of the sequential kernel estimators. The optimal adaptive convergence rate is given as well as the upper bound for the minimax risk.

We will construct an adaptive minimax sequential estimation for the function S from the functional class [1.2.3] of the unknown regularity β . To this end we will use the modification of the adaptive Lepskiĭ method proposed in Arkoun (2011) based on the sequential estimators [1.2.11]. We set

$$d_n = \frac{n}{\ln n} \quad \text{and} \quad \phi_n(\beta) = (d_n)^{\frac{\beta}{2\beta+1}}. \quad [1.2.15]$$

Moreover, we chose the bandwidth h in the form [1.2.13] with $\kappa_n = 1/d_n$, i.e. we set

$$\check{h} = \check{h}(\beta) = \left(\frac{1}{d_n} \right)^{\frac{1}{2\beta+1}}.$$

We define the grids on the intervals $[\underline{\beta}, \bar{\beta}]$ and $[\underline{\alpha}, \bar{\alpha}]$ as

$$\beta_k = \underline{\beta} + \frac{k}{m}(\bar{\beta} - \underline{\beta}) \quad \text{and} \quad \alpha_k = \underline{\alpha} + \frac{k}{m}(\bar{\alpha} - \underline{\alpha})$$

for $0 \leq k \leq m$ with $m = \lfloor \ln d_n \rfloor + 1$, and we set

$$\phi_{n,k} = \phi_n(\beta_k) \quad \text{and} \quad \check{h}_k = \check{h}(\beta_k).$$

Replacing in [1.2.10] and [1.2.12] the parameters h and α we define

$$\check{\nu}_k = \nu(\check{h}_k, \alpha_k) \quad \text{and} \quad \check{H}_k = H(\check{h}_k, \alpha_k).$$

Now using these parameters in the estimators [1.2.9] and [1.2.11] we set $\check{S}_k = S_{\check{h}_k}^*(z_0)$ and

$$\check{\omega}_k = \max_{0 \leq j \leq k} \left(|\check{S}_j - \check{S}_k| - \frac{\check{\lambda}}{N_j} \right), \quad [1.2.16]$$

where

$$\check{\lambda} > \check{\lambda}_* = 4\sqrt{2} \left(\frac{\bar{\beta} - \underline{\beta}}{(2\underline{\beta} + 1)(2\bar{\beta} + 1)} \right)^{1/2}.$$

In particular, if $\underline{\beta} = 1$ and $\bar{\beta} = 2$ we get $\check{\lambda}_* = 4(2/15)^{1/2}$. We also define the optimal index as

$$\check{k} = \max \left\{ 0 \leq k \leq m : \check{\omega}_k \leq \frac{\check{\lambda}}{N_k} \right\}.$$

The adaptive estimator is now defined as

$$\hat{S}_{a,n} = S_{\check{h}_k}^* \quad \text{and} \quad \check{h}_k = \check{h}_{\check{k}}. \quad [1.2.17]$$

REMARK 1.3.— *It should be noted that the usual adaptive pointwise estimation (see, for example, Lepskii (1990), Galtchouk and Pergamenschikov (2001)) differs from the threshold $\check{\lambda}$ by the fact that [1.2.16] does not depend on the parameters $L > 0$ and $K > 0$ of the Hölder class [1.2.3].*

Now we are able to give the main results for the adaptive estimation of the function S at a fixed point $z_0 \in]0, 1[$, in the Gaussian model [1.1.1], i.e. when $p = p_0$ where p_0 is $(0, 1)$ –Gaussian density.

For any estimate \check{S}_n of $S(z_0)$, we define the adaptive risk over the class $\mathcal{H}^{(\beta)}(\varepsilon, L, K)$ as

$$\mathcal{R}_{a,n}(\check{S}_n) = \sup_{\beta \in [\underline{\beta}; \bar{\beta}]} \sup_{S \in \mathcal{H}^{(\beta)}(\varepsilon, L, K)} \phi_n(\beta) \mathbf{E}_{S, p_0} |\check{S}_n - S(z_0)|, \quad [1.2.18]$$

where $\phi_n(\beta)$ is defined in [1.2.15].

First we give the lower bound for the minimax risk. More precisely, with the convergence rate $\phi_n(\beta)$, the lower bound for the minimax risk is strictly positive.

THEOREM 1.5.— *There exists $K_0 > 0$ such that for all $K > K_0$, the risk [1.2.18] admits the following lower bound:*

$$\liminf_{n \rightarrow \infty} \inf_{\check{S}_n} \mathcal{R}_{a,n}(\check{S}_n) > 0,$$

where the infimum is taken over all estimators \check{S}_n .

To obtain an upper bound for the adaptive risk [1.2.18] of the procedure [1.2.17] we need to study the family $(S_h^*)_{\alpha \leq \alpha \leq \bar{\alpha}}$.

THEOREM 1.6.– *The sequential procedure [1.2.11] with the bandwidth h defined in [1.2.13] for $\kappa_n = \ln n/n$ satisfies the following property*

$$\limsup_{n \rightarrow \infty} \sup_{\alpha \leq \alpha \leq \bar{\alpha}} (\Upsilon_n(h))^{-1} \sup_{S \in \mathcal{H}^{(\beta)}(\varepsilon, L, K)} \sup_{p \in \mathcal{P}_\zeta} \mathbf{E}_{S,p} |S_h^* - S(z_0)| < \infty$$

where $\Upsilon_n(h) = h^\beta + (nh)^{-1/2}$.

Using this theorem we can establish the minimax property for the procedure [1.2.11].

THEOREM 1.7.– *The estimation procedure [1.2.11] satisfies the following asymptotic property*

$$\limsup_{n \rightarrow \infty} \mathcal{R}_{a,n}(\hat{S}_{a,n}) < \infty.$$

REMARK 1.4.– *Theorem 1.5 gives the lower bound for the adaptive risk, i.e. the convergence rate $\phi_n(\beta)$ is best for the adapted risk. Moreover, by Theorem 1.7 the adaptive estimator [1.2.11] possesses this convergence rate. In this case, this estimator is called optimal in sense of the adaptive risk [1.2.18]*

1.3. Estimation with quadratic integral risk

In this section, we still consider the problem of estimating the autoregressive function of the model [1.1.1] at any point in $[a, b]$ but we will use the quadratic integral risk to measure the mean error of an estimator \hat{S}_n defined as

$$\mathcal{R}_p(\hat{S}_n, S) = \mathbf{E}_{p,S} \|\hat{S}_n - S\|^2, \quad \|S\|^2 = \int_a^b S^2(x) dx, \quad [1.3.1]$$

Recalling that the distribution p of the noises ξ_k in model [1.1.1] is unknown, so we still use the robust nonparametric estimation approach :

$$\mathcal{R}^*(\hat{S}_n, S) = \sup_{p \in \mathcal{P}} \mathcal{R}_p(\hat{S}_n, S), \quad [1.3.2]$$

where \mathcal{P} is defined in [1.1.2]

In this context, we focus here on obtaining an oracle inequality which is the first step in getting asymptotic efficiency. More precisely, in order to estimate the function S in model [1.1.1], we make use of the estimator family $(\hat{S}_\lambda, \lambda \in \Lambda)$, where \hat{S}_λ is a weighted least square estimator with the Pinsker weights. For this family, similarly to Galtchouk and Pergamenschikov (2009a), we construct a special selection rule, i.e.

a random variable $\widehat{\lambda}$ with values in Λ , for which we define the selection estimator as $\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}$. Our goal is to show the non asymptotic sharp oracle inequality for the robust risks [1.3.2], i.e. to show that for any $\check{\varrho} > 0$ and $n \geq 1$

$$\mathcal{R}^*(\widehat{S}_*, S) \leq (1 + \check{\varrho}) \min_{\lambda \in \Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathcal{B}_n}{\check{\varrho}n}, \quad [1.3.3]$$

where \mathcal{B}_n is a rest term such that for any $\check{\delta} > 0$,

$$\lim_{n \rightarrow \infty} \frac{\mathcal{B}_n}{n^{\check{\delta}}} = 0.$$

In this case the estimator \widehat{S}_* is called optimal in the oracle inequality sense.

In order to obtain this inequality for model [1.1.1], we develop a new model selection method based on the truncated sequential procedure constructed in the previous section for the pointwise efficient estimation. To this purpose, we use the non-asymptotic analysis tool proposed for regression models in Galtchouk and Pergamenschikov (2009a) based on the non-asymptotic studies from Baron *et al.* (1999) for a family of least-squares estimators and extended in Fourdrinier and Pergamenschikov (2007) to some others estimators families. Here we follow the approach proposed by Galtchouk and Pergamenschikov (2011), i.e. we pass to a discrete time regression model by making use of the truncated sequential procedure introduced previously. To this end, at any point $(z_l)_{1 \leq l \leq d}$ of a partition of the interval $[a, b]$, we define a sequential procedure (τ_l, S_l^*) with a stopping rule τ_l and an estimator S_l^* . For $Y_l = S_l^*$ with $1 \leq l \leq d$, we come to the regression equation on some set $\Gamma \subseteq \Omega$:

$$Y_l = S(z_l) + \zeta_l, \quad 1 \leq l \leq d. \quad [1.3.4]$$

Here, in contrast with the classical regression model, the noise sequence $(\zeta_l)_{1 \leq l \leq d}$ has a complex structure, namely,

$$\zeta_l = \xi_l^* + \varpi_l,$$

where $(\xi_l^*)_{1 \leq l \leq d}$ is a "main noise" sequence of uncorrelated random variables and $(\varpi_l)_{1 \leq l \leq n}$ is a sequence of bounded random variables.

It should be emphasized that only sharp oracle inequalities of type [1.3.3] allow to synthesis efficiency property in the adaptive setting.

1.3.1. Main Conditions

As in Arkoun and Pergamenschikov (2016) we assume that in the model [1.1.1] the i.i.d. random variables $(\xi_k)_{k \geq 1}$ have a density p (with respect to Lebesgue measure) from the functional class \mathcal{P}_ς defined previously in [1.1.2], where $\varsigma \geq 1$ is

some parameter, which may be a function of the number observation n , i.e. $\varsigma = \varsigma(n)$, such that for any $\tilde{\delta} > 0$

$$\lim_{n \rightarrow \infty} \frac{\varsigma(n)}{n^{\tilde{\delta}}} = 0. \quad [1.3.5]$$

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

$$\mathbf{m}_q^* = \sup_{p \in \mathcal{P}} \mathbf{E}_p |\xi_1|^q < \infty,$$

where \mathbf{E}_p is the expectation with respect to the density p from \mathcal{P}_ς . To obtain the stable (uniformly with respect to the function S) model [1.1.1], we assume that for some fixed $0 < \varepsilon < 1$ and $L > 0$ the unknown function S belongs to the ε -stability set introduced in Arkoun and Pergamenchtchikov (2016) as

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

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

1.3.2. Passage to a discrete time regression model

We will use as a basic procedure the pointwise procedure from Arkoun and Pergamenchtchikov (2016) at the points $(z_l)_{1 \leq l \leq d}$ defined as

$$z_l = a + \frac{l}{d}(b - a) \quad \text{and} \quad d = [\sqrt{n}], \quad [1.3.6]$$

where $[a]$ is the integer part of a number a . So we propose to use the first ι_l observations for the auxiliary estimation of $S(z_l)$. We set

$$\hat{S}_{\iota_l} = \frac{1}{A_{\iota_l}} \sum_{j=1}^{\iota_l} Q_{l,j} y_{j-1} y_j, \quad A_{\iota_l} = \sum_{j=1}^{\iota_l} Q_{l,j} y_{j-1}^2, \quad [1.3.7]$$

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

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

Note that to estimate $S(z_l)$ on the basis of kernel estimator 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.

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

where $\tilde{z}_l = (z_l - a)/(b - a)$ and $\tilde{h} = h/(b - a)$. Note that, only for the last point $z_d = b$, we take $k_{2,d} = n$. We choose ι_l in [1.3.7] as

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

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

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

Next, similarly to Arkoun (2011), we use a kernel sequential procedure based on the observations $(y_j)_{\iota_l \leq j \leq n}$. To transform the kernel estimator in a linear function of observations and we replace the number of observations n by the following stopping time

$$\tau_l = \inf\{\iota_l + 1 \leq k \leq k_{2,l} : A_{\iota_l,k} \geq H_l\}, \quad [1.3.8]$$

where $\inf\{\emptyset\} = k_{2,l}$ and the positive threshold H_l will be chosen as a positive random variable which is measurable with respect to the σ -algebra $\{y_1, \dots, y_{\iota_l}\}$.

Now we define the sequential estimator as

$$S_l^* = \frac{1}{H_l} \left(\sum_{j=\iota_l+1}^{\tau_l-1} Q_{l,j} y_{j-1} y_j + \varkappa_l Q_{l,\tau_l} y_{\tau_l-1} y_{\tau_l} \right) \mathbf{1}_{\Gamma_l}, \quad [1.3.9]$$

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

$$A_{\iota_l, \tau_l-1} + \varkappa_l^2 Q_{l,\tau_l} y_{\tau_l-1}^2 = H_l. \quad [1.3.10]$$

Note that, to obtain the efficient kernel estimator of $S(z_l)$ we need to use all $k_{2,l} - \iota_l - 1$ observations. Similarly to Konev and Pergamenschikov (1984), one can show that $\tau_l \approx \gamma_l H_l$ as $H_l \rightarrow \infty$, where

$$\gamma_l = 1 - S^2(z_l).$$

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

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

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

$$\tilde{S}_{\iota_l} = \min(\max(\hat{S}_{\iota_l}, -1 + \tilde{\varepsilon}), 1 - \tilde{\varepsilon}). \quad [1.3.11]$$

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

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

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

PROPOSITION 1.2.– *The convergence rate in probability of the estimator [1.3.11] is more rapid than any power function, i.e. for any $\mathbf{b} > 0$*

$$\lim_{n \rightarrow \infty} n^{\mathbf{b}} \max_{1 \leq l \leq d} \sup_{S \in \Theta_{\varepsilon, L}} \sup_{p \in \mathcal{P}} \mathbf{P}_{S, p} \left(|\widetilde{S}_{l_t} - S(z_l)| > \epsilon_0 \right) = 0, \quad [1.3.12]$$

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

Now we set

$$Y_l = S_l^* \mathbf{1}_\Gamma \quad \text{and} \quad \Gamma = \bigcap_{l=1}^d \Gamma_l. \quad [1.3.13]$$

Using the convergence [1.3.12], we can study the probability properties of the set Γ in the following theorem.

THEOREM 1.8.– *For any $\mathbf{b} > 0$, the probability of the set Γ^c satisfies the following asymptotic equality*

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

In view of this theorem we can shrink the set Γ^c . So, using the estimators [1.3.13] on the set Γ we obtain the discrete time regression model [1.3.4] in which

$$\xi_l^* = \frac{\sum_{j=l_t+1}^{\tau_l-1} Q_{l,j} y_{j-1} \xi_j + \varkappa_l Q(u_{l,\tau_l}) y_{\tau_l-1} \xi_{\tau_l}}{H_l} \quad [1.3.14]$$

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

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

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 [1.3.4] the random variables $(\xi_j^*)_{1 \leq j \leq d}$ are defined only on the set Γ . For technical reasons we need to define these variables on the set Γ^c as well. To this end, for any $j \geq 1$ we set

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

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

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

and $\check{A}_{\iota_l, k_{2,l}} \geq H_l$. So now we can modify the stopping time [1.3.8] as

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

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 [1.3.10] we define the correction coefficient as

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

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

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

Note that on the set Γ , for any $1 \leq l \leq d$, the random variables $\eta_l = \xi_l^*$. Moreover we can show that, for any $1 \leq l \leq d$ and $p \in \mathcal{P}_c$

$$\mathbf{E}_{S,p}(\eta_l | \mathcal{G}_l) = 0, \quad \mathbf{E}_{S,p}(\eta_l^2 | \mathcal{G}_l) = \sigma_l^2, \quad \mathbf{E}_{S,p}(\eta_l^4 | \mathcal{G}_l) \leq \check{\mathbf{m}} \sigma_l^4, \quad [1.3.15]$$

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

$$\sigma_{0,*} \leq \min_{1 \leq l \leq d} \sigma_l^2 \leq \max_{1 \leq l \leq d} \sigma_l^2 \leq \sigma_{1,*}, \quad [1.3.16]$$

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

$$\sup_{S \in \Theta_{\epsilon, L}} \mathbf{E}_{S,p} \mathbf{1}_\Gamma \varpi_l^2 \leq \left(L^2 h^2 + \frac{\check{v}_n}{(nh)^2} \right), \quad [1.3.17]$$

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

THEOREM 1.9.– For any $\mathbf{b} > 0$ the sequence $(\check{v}_n)_{n \geq 1}$ satisfies the following limiting equality

$$\lim_{n \rightarrow \infty} n^{-\mathbf{b}} \check{v}_n = 0. \quad [1.3.18]$$

REMARK 1.5.– It should be noted that the property [1.3.18] means that the asymptotic behavior of the upper bound [1.3.17] is approximately as h^{-2} when $n \rightarrow \infty$. We will use this in the oracle inequalities below.

REMARK 1.6.– *It should be emphasized that to estimate the function S in [1.1.1] we use the approach developed in Galtchouk and Pergamenshchikov (2011) for the sequential drift estimation problem in the stochastic differential equation. On the basis of the efficient sequential kernel procedure developed in Galtchouk and Pergamenshchikov (2005), Galtchouk and Pergamenshchikov (2006b) and Galtchouk and Pergamenshchikov (2015) with the kernel-indicator, the stochastic differential equation is replaced by regression model. It should be noted that to obtain the efficient estimator one needs to take the kernel-indicator estimator. By this reason, in this paper, we use the kernel-indicator in the sequential estimator [1.3.9]. It also should be noted that the sequential estimator [1.3.9] which has the same form as in Arkoun and Pergamenshchikov (2016), except the last term, in which the correction coefficient is replaced by the square root of the coefficient used in Konev (2016). We modify this procedure to calculate the variance of the stochastic term [1.3.14].*

1.3.3. Model selection

In this section we consider the nonparametric estimation problem in the non asymptotic setting for the regression model [1.3.4] for some set $\Gamma \subseteq \Omega$. The design points $(z_l)_{1 \leq l \leq d}$ are defined in [1.3.6]. The function $S(\cdot)$ is unknown and has to be estimated from observations the Y_1, \dots, Y_d . Moreover, we assume that the unobserved random variables $(\eta_l)_{1 \leq l \leq d}$ satisfy the properties [1.3.15] with some nonrandom constant $\mathfrak{m} > 1$ and the known random positive coefficients $(\sigma_l)_{1 \leq l \leq d}$ satisfy the inequality [1.3.16] for some nonrandom positive constants $\sigma_{0,*}$ and $\sigma_{1,*}$. Concerning the random sequence $\varpi = (\varpi_l)_{1 \leq l \leq n}$ we suppose that

$$\mathbf{u}_d^* = \mathbf{E}_{S,p} \mathbf{1}_\Gamma \|\varpi\|_d^2 < \infty. \quad [1.3.19]$$

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

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

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

$$(\phi_i, \phi_j)_d = \frac{b-a}{d} \sum_{l=1}^d \phi_i(z_l) \phi_j(z_l) = \mathbf{1}_{\{i=j\}}. \quad [1.3.20]$$

For example, we can take the trigonometric basis $(\phi_j)_{j \geq 1}$ in $\mathbf{L}_2[a, b]$ defined as

$$\phi_1 = 1, \quad \phi_j(x) = \sqrt{\frac{2}{b-a}} \text{Tr}_j(2\pi[j/2]l_0(x)), \quad j \geq 2,$$

where the function $\text{Tr}_j(x) = \cos(x)$ for even j and $\text{Tr}_j(x) = \sin(x)$ for odd j , $[x]$ denotes the integer part of x . and $l_0(x) = (x-a)/(b-a)$. Note that, using the

orthonormality property [1.3.20] we can represent for any $1 \leq l \leq d$ the function S as

$$S(z_l) = \sum_{j=1}^d \theta_{j,d} \phi_j(z_l) \quad \text{and} \quad \theta_{j,d} = (S, \phi_j)_d. \quad [1.3.21]$$

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

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

From [1.3.4] we obtain immediately the following regression scheme

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

where

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

Note that the upper bound [1.3.16] and the Bounyakovskii-Cauchy-Schwarz inequality imply that

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

We estimate the function S on the grid [1.3.6] by the weighted least-squares estimator

$$\hat{S}_\lambda(z_l) = \sum_{j=1}^d \lambda(j) \hat{\theta}_{j,d} \phi_j(z_l) \mathbf{1}_\Gamma, \quad 1 \leq l \leq d, \quad [1.3.22]$$

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 $a \leq t \leq b$

$$\hat{S}_\lambda(t) = \sum_{l=1}^d \hat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < t \leq z_l\}}. \quad [1.3.23]$$

Moreover, denoting $\lambda^2 = (\lambda^2(1), \dots, \lambda^2(n))'$ we define the following sets

$$\Lambda_1 = \{\lambda^2, \lambda \in \Lambda\} \quad \text{and} \quad \Lambda_2 = \Lambda \cup \Lambda_1.$$

Denote by ν the cardinal number of the set Λ and

$$\nu^* = \max_{\lambda \in \Lambda} \sum_{j=1}^d \mathbf{1}_{\{\lambda(j) > 0\}}.$$

In order to obtain a good estimator, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in [1.3.22]. We define the empirical squared risk as

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

Using [1.3.21] and [1.3.22] we can rewrite this risk as

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

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

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

Note that from [1.3.16] - [1.3.20] it follows that

$$\mathbf{s}_{j,d} \leq \sigma_{1,*}.$$

Similarly to Galtchouk and Pergamenshchikov (2011) one needs to introduce a penalty term in the cost function to compensate such modification of the empirical risk. We choose it as

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

Finally, we define the cost function in the following form

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

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

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_d(\lambda)$$

and define an estimator of $S(t)$ of the form [1.3.23]:

$$\widehat{S}_*(t) = \widehat{S}_{\widehat{\lambda}}(t) \quad \text{for} \quad a \leq t \leq b. \quad [1.3.24]$$

REMARK 1.7.— We use the procedure [1.3.24] to estimate the function S in the autoregressive model [1.1.1] through the regression scheme [1.3.4] generated by the sequential procedures [1.3.13].

1.3.4. Main results

In this section we formulate all main results. First we obtain the sharp oracle inequality for the selection model procedure [1.3.24] for the general regression model [1.3.4].

THEOREM 1.10.– *There exists some constant $l^* > 0$ such that for any weight vectors set Λ , any $p \in \mathcal{P}_{\text{varsigma}}$, any $n \geq 1$ and $0 < \delta \leq 1/12$, the procedure [1.3.24], satisfies the following oracle inequality*

$$\begin{aligned} \mathbf{E}_{S,p} \|\widehat{S}_* - S\|_d^2 &\leq \frac{1+4\delta}{1-6\delta} \min_{\lambda \in \Lambda} \mathbf{E}_{S,p} \|\widehat{S}_\lambda - S\|_d^2 \\ &\quad + l^* \frac{\nu \varsigma^2}{\delta} \left(\frac{\sigma_{1,*}^2}{\sigma_{0,*} d} + \mathbf{u}_d^* + \delta^2 \sqrt{\mathbf{P}_{S,p}(\Gamma^c)} \right). \end{aligned}$$

Using Lemma A.2 from Konev and Pergamenshchikov (2015) we obtain the oracle inequality for the quadratic risks [1.3.1].

THEOREM 1.11.– *There exists some constant $l^* > 0$ such that for any weight vectors set Λ , any continuously differentiable function S , any $p \in \mathcal{P}_\varsigma$, any $n \geq 1$ and $0 < \delta \leq 1/12$, the procedure [1.3.24] satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_p(\widehat{S}_*, S) &\leq \frac{(1+4\delta)(1+\delta)^2}{1-6\delta} \min_{\lambda \in \Lambda} \mathcal{R}_p(\widehat{S}_\lambda, S) \\ &\quad + l^* \frac{\varsigma^2 \nu}{\delta} \left(\frac{\|\dot{S}\|^2}{d^2} + \frac{\sigma_{1,*}^2}{\sigma_{0,*} d} + \mathbf{u}_d^* + \delta^2 \sqrt{\mathbf{P}_{S,p}(\Gamma^c)} \right). \end{aligned}$$

Now we assume that the cardinal ν of Λ and the parameter ς in the density family [1.1.2] are functions of the number observations n , i.e. $\nu = \nu(n)$ and $\varsigma = \varsigma(n)$ such that for any $\check{\delta} > 0$

$$\lim_{n \rightarrow \infty} \frac{\nu(n)}{n^{\check{\delta}}} = 0. \quad [1.3.25]$$

Using Theorems 1.8 – 1.9 and the bounds [1.3.16] - [1.3.17] we obtain the oracle inequality for the estimation problem for the model [1.1.1].

THEOREM 1.12.– *Assume that the conditions [1.3.5] and [1.3.25] hold. Then for any $p \in \mathcal{P}_\varsigma$, $S \in \Theta_{\varepsilon,L}$, $n \geq 3$ and $0 < \delta \leq 1/12$, the procedure [1.3.24] satisfies the following oracle inequality*

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

where the term $\check{\mathbf{B}}_n(p)$ is such that for any $\check{\delta} > 0$

$$\lim_{n \rightarrow \infty} \frac{\check{\mathbf{B}}_n(p)}{n^{\check{\delta}}} = 0.$$

We obtain the same inequality for the robust risks

THEOREM 1.13.— *Assume that the conditions [1.3.5] and [1.3.25] hold. Then for any $n \geq 3$, any $S \in \Theta_{\varepsilon, L}$ and any $0 < \delta \leq 1/12$, the procedure [1.3.24] satisfies the following oracle inequality*

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

where the term $\check{\mathbf{B}}_n$ is such that for any $\check{\delta} > 0$

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

REMARK 1.8.— *It is well known that to obtain the efficiency property we need to specify the weight coefficients $(\lambda(j))_{1 \leq j \leq n}$ (see, for example, Galtchouk and Pergamenschikov (2009b)). Consider for some fixed $0 < \varepsilon < 1$ a numerical grid of the form*

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

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

$$\begin{cases} \lim_{n \rightarrow \infty} k^*(n) = +\infty, & \lim_{n \rightarrow \infty} \frac{k^*(n)}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \varepsilon(n) = 0 & \text{and} \quad \lim_{n \rightarrow \infty} n^{\check{\delta}} \varepsilon(n) = +\infty \end{cases} \quad [1.3.26]$$

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

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

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

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

with the elements

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

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

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

Now we define the set Λ as

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

Note that these weight coefficients are used in Konev and Pergamenshchikov (2012, 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 properties [1.3.26] imply condition [1.3.25].

1.4. Bibliography

- Arkoun, O. (2011), Sequential adaptive estimators in nonparametric autoregressive models, *Sequential Analysis*, 44, 228–246.
- Arkoun, O., Pergamenchtchikov, S. (2008), Nonparametric estimation for an autoregressive model, *Journal of Mathematics and Mechanics of Tomsk State University*, (2), 20–30.
- Arkoun, O., Pergamenchtchikov, S. (2016), Sequential robust estimation for nonparametric autoregressive models, *Sequential Analysis*, 35, 489–515.
- Baron, A., Birgé, L., Massart, P. (1999), Risk bounds for model selection via penalization, *Probab. Theory Related Fields*, 113, 301–413.
- Belitser, E. (2000), Recursive estimation of a drifted autoregressive parameter, *The Annals of Statistics*, 26, 860–870.
- Borisov, V., Konev, V. (1977), Sequential estimation of parameters of discrete processes, *Automat. and Remote control*, 10, 58–64.
- Dahlhaus, R. (1996a), Maximum likelihood estimation and model selection for locally stationary processes, *Journal of Nonparametric Statistics*, 6, 171–191.
- Dahlhaus, R. (1996b), On the kullback-leibler information divergence of locally stationary processes, *Stochastic Processes and Their Applications*, 62, 139–168.
- Fan, J., Zhang, W. (2008), Statistical methods with varying coefficient models, *Statistics and Its Interface*, (1), 179–195.
- Fourdrinier, D., Pergamenshchikov, S. (2007), Improved model selection method for a regression function with dependent noise, *Annals of the Institute of Statistical Mathematics*, 59, 435–464.
- Galtchouk, L., Pergamenshchikov, S. (2001), Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes, *Math. Methods Statist.*, 10(3).
- Galtchouk, L., Pergamenshchikov, S. (2005), Nonparametric sequential minimax estimation of the drift coefficient in diffusion processes, *Sequential Analysis*, 24, 303–330.

- Galtchouk, L., Pergamenshchikov, S. (2006a), Asymptotically efficient estimates for nonparametric regression models, *Statistics and Probability Letters*, 76, 852–860.
- Galtchouk, L., Pergamenshchikov, S. (2006b), Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes, *Statistical Inference for Stochastic Processes*, 9, 1–16.
- Galtchouk, L., Pergamenshchikov, S. (2009a), Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models, *Journal of Nonparametric Statistics*, 21, 1–16.
- Galtchouk, L., Pergamenshchikov, S. (2009b), Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression, *Journal of Korean Statistical Society*, 38, 305–322.
- Galtchouk, L., Pergamenshchikov, S. (2011), Adaptive sequential estimation for ergodic diffusion processes in quadratic metric, *Journal of Nonparametric Statistics*, 23, 255–285.
- Galtchouk, L., Pergamenshchikov, S. (2015), Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions, *Bernoulli*, 21, 2569–2594.
- Konev, V. (2016), On one property of martingales with conditionally gaussian increments and its application in the theory of nonasymptotic inference, *Doklady Mathematics*, 94, 1–5.
- Konev, V., Pergamenshchikov, S. (1984), Estimate of the number of observations in sequential identification of the parameters of dynamical systems, *Avtomat. i Telemekh.*, 12, 56–63.
- Konev, V., Pergamenshchikov, S. (2012), Efficient robust nonparametric estimation in a semimartingale regression model, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48, 1244–1277.
- Konev, V., Pergamenshchikov, S. (2015), Robust model selection for a semimartingale continuous time regression from discrete data, *Stochastic Processes and their Applications*, 125, 294–326.
- Lepskii, O. V. (1990), A problem of adaptive estimation in gaussian white noise, *Theory of Probability and Its Applications*, 35, 454–466.
- Luo, X. H., Y. Z. H., Zhou, Y. (2009), Nonparametric estimation of the production function with time-varying elasticity coefficients, *Systems Engineering-Theory and Practice*, 29, 144–149.
- Moulines, E., Priouret, P., Roueff, F. (2005), On recursive estimation for time varying autoregressive processes, *Annals of Statistics*, 33, 2610–2654.

