

# Non asymptotic sharp oracle inequalities for high dimensional ergodic diffusion models. \*

L.I. Galtchouk <sup>†</sup>      S.M. Pergamenshchikov <sup>‡</sup>

## Abstract

In this paper we consider high dimensional ergodic diffusion models in nonparametric setting on the basis of discrete data, when the diffusion coefficients are unknown. For this problem, by using efficient sequential point-wise estimators we construct a model selection procedure and then we show sharp oracle inequalities, i.e. the inequalities in which the main term coefficient is closed to one. This means that the proposed sequential model selection procedure is optimal in this sense. Particularly, we show that the constructed procedure is the best in the class of weighted least square estimators with the Pinsker coefficients which provide the efficient estimation in the minimal asymptotical quadratic risk sense.

*Keywords:* High dimensional diffusion models; Sharp oracle inequality; Discrete data; Model selection; Quadratic risk; Sequential estimation.

*AMS 2000 Subject Classifications:* primary 62G08; secondary 62G05, 62G20.

---

\*This work is supported by the RSF Grant no. 17-11-01049, National Russian Research Tomsk State University, Russia

<sup>†</sup>International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, 36, Lenin str., 634041, Tomsk, Russia, e-mail: leonid.galtchouk0667@orange.fr

<sup>‡</sup>LMRS, UMR 6085 CNRS, Université de Rouen Normandie, Saint-Etienne du Rouvray Cedex, France International Laboratory SSP & QF, Tomsk State University, e-mail: Serge.Pergamenshchikov@univ-rouen.fr

# 1 Introduction

## 1.1 Problem and Motivations

In this paper we consider the following diffusion model

$$dy_t = \sum_{j=1}^q \theta_j \psi_j(y_t) dt + b(y_t) dW_t, 0 \leq t \leq T, \quad (1.1)$$

where the functions  $(\psi_j)_{1 \leq j \leq q}$  are known linear independent functions,  $(W_t)_{t \geq 0}$  is a standard Wiener process and the diffusion coefficient  $b(\cdot)$  is an unknown function. Moreover, it is assumed that the observations are accessible only at discrete times, i.e.

$$(y_{t_j})_{1 \leq j \leq N}, \quad t_j = j\delta, \quad (1.2)$$

where the frequency  $\delta = \delta_T \in (0, 1)$  is some function of  $T$  which will be specified later and the sample size  $N = N(T) = [T/\delta]$ . Here  $[a]$  denotes the integer part of  $a$ . It should be noted that in this case the diffusion coefficient is considered as a nuisance parameter. The problem is to estimate the unknown parameters  $(\theta_j)_{1 \leq j \leq q}$  in the "high dimensional" setting, i.e. when the number of parameters  $q > N$ . Usually, for such models one uses one of two methods: Lasso algorithm proposed in [37] for i.i.d. regression models and modified later for diffusion regression models (1.1) in [10] and the Dantzig selector method proposed in [3] and applied to diffusion models in [9]. But in all these papers the number of parameters  $q$  is known and, therefore, unfortunately, these methods can't be applied to estimate the number of parameters in regression models. In this paper we study this problem in the nonparametric estimation framework, i.e. we study the diffusion process defined as

$$dy_t = S(y_t) dt + b(y_t) dW_t \quad (1.3)$$

and the problem is to estimate the unknown function  $S(\cdot)$  on the basis of the observations (1.2). The nonparametric setting allows to consider the models (1.1) with unknown  $q$  or with  $q = +\infty$ . Note that, the case when the number of parameters  $q$  is unknown is one of challenging problems in the signal processing theory (see, for example, [2]). The models (1.3) are very important for various fields in the theory of stochastic processes such as optimal control, identification problem, filtration, financial markets, insurance etc (see, for example, [28, 29, 34, 35, 36]). Nonparametric estimation problems of the drift  $S$  were studied in a number of papers in the case of complete observations, i.e. when the whole trajectory  $(y_t)_{0 \leq t \leq T}$  is observed. In [12, 14, 15] a sequential approach was proposed for the point-wise efficient estimation. In [5, 6, 13, 19]

this problem was studied for integral risks when the diffusion coefficient is known. In practice, usually only discrete time observations are accessible. A natural question arises about proprieties and the behavior of estimates based on discrete time observations for such models. Nonparametric estimation based on discrete time observations for models (1.3) was considered firstly for estimating the squared unknown diffusion coefficient  $b^2(\cdot)$  on a fixed interval  $[0, T]$  (see, for example, [8, 25, 26, 27] and the references therein). Later, in [24] kernel estimates for the drift and diffusion coefficients were studied for reflecting ergodic (1.3), i.e. for the processes with values in  $[0, 1]$ . Concerning the estimation in the ergodic case, a sequential procedure was proposed in [26] for nonparametric estimating the drift coefficient in an integral metric and in [22] a special sequential method was developed to provide the efficient estimation for the point-wise risks, i.e. sequential kernel procedures for which normalized minimax risks attain a minimal value. This property is very important for nonparametric estimation, since usually the convergence rate is very slow, therefore, the influence of the value of the normalized risks in the estimation accuracy is much more important than in the parametric cases.

## 1.2 Main tool

In this paper for the problem (1.3) we develop a new model selection method which is an adaptive selection rule  $\hat{\lambda}$  of an estimator  $S^* = \hat{S}_{\hat{\lambda}}$  in the family of weighted least square estimators  $(\hat{S}_{\lambda})_{\lambda \in \Lambda}$ . Our goal is to show that the constructed procedure is optimal in the sense of the following oracle inequality: for any small  $\rho > 0$  and any  $T > 0$

$$\mathbf{E}\|S^* - S\|^2 \leq (1 + \rho) \min_{\lambda \in \Lambda} \mathbf{E}\|\hat{S}_{\lambda} - S\|^2 + \frac{\mathcal{B}_T}{\rho T}, \quad (1.4)$$

where  $\|f\|^2 = \int_a^b f^2(x)dx$  for some  $a < b$  and the term  $\mathcal{B}_T$  is some slowly varying function of  $T$ , i.e. for any  $\delta > 0$

$$\lim_{T \rightarrow \infty} \frac{\mathcal{B}_T}{T^\delta} = 0. \quad (1.5)$$

It should be noted that for the first time, such inequalities for non-Gaussian regression models in discrete time were obtained in [17], and for continuous processes in [30]. Later, such inequalities were called sharp (since the main coefficient  $1 + \rho$  arbitrarily close to one). Moreover, it should be noted that such inequalities allow us to prove asymptotic efficiency without knowing the regularity of the estimated function (see, for example, [19]).

Our approach is based on the sequential estimation method developed in [22] for the efficient nonparametric point-wise estimation in a special

functional class. Similarly to [19] we will use these estimators to pass to a heteroscedastic regression model and, then, using the methods developed in [17, 18] for such regression we construct the sequential model selection procedure. To obtain the oracle inequalities we use a special uniform concentration inequality from [20] obtained for discrete time observations.

It should be noted also that in this paper we construct the model selection procedure for arbitrary orthonormal basis. We recall that in [19] the model selection procedure was constructed only for the trigonometric basis. In practice, this basis is not well appropriate sometimes to signal processing problems (see, for example, [4, 7] and the references therein). So in the paper, we develop a new analytical tool to provide sharp oracle inequalities for any orthonormal basis on the grid.

### 1.3 Plan of the paper

The paper is organized as follows. In Section 2 we describe functional classes that will be used. Sequential estimators of the drift coefficient are constructed in Section 3. In Section 4 we introduce a regression model based on the sequential estimators. In Section 5 we construct the model selection procedure based on the sequential pointwise estimators. Main results are announced in Section 6 and their proofs are given in Section 8. In Section 7 we study some properties of the model selection. In Appendix we give all auxiliary results.

## 2 Main Conditions

In the paper we consider the estimation problem for the drift  $S$  on the interval  $[\mathbf{x}_0, \mathbf{x}_1]$ , where  $\mathbf{x}_0 < \mathbf{x}_1$  are some arbitrary fixed points. In order to obtain a reliable estimator of  $S$ , it is necessary to impose some conditions on this function which are similar to the periodicity of the deterministic signal in the white noise model. One of conditions which is sufficient for this purpose is the assumption that the process  $(y_t)_{t \geq 0}$  in (1.3) returns to any vicinity of each point  $x \in [\mathbf{x}_0, \mathbf{x}_1]$  infinite times. The ergodicity provides this property, when the coefficients of equation are known. In the case of unknown coefficients, one needs to impose the uniform ergodicity property. To obtain the uniform ergodicity property for the process (1.3) we use the functional class introduced in [22], i.e. for any fixed  $\mathbf{L} \geq 1$ ,  $\mathbf{M} > 0$  and  $\mathbf{x}_* > |\mathbf{x}_0| + |\mathbf{x}_1|$  we

set

$$\Sigma_{\mathbf{L},\mathbf{M}} = \left\{ S \in \mathbf{C}^1(\mathbb{R}) : \sup_{|x| \leq \mathbf{x}_*} (|S(x)| + |\dot{S}(x)|) \leq \mathbf{M}, \right. \\ \left. -\mathbf{L} \leq \inf_{|x| \geq \mathbf{x}_*} \dot{S}(x) \leq \sup_{|x| \geq \mathbf{x}_*} \dot{S}(x) \leq -1/\mathbf{L} \right\}. \quad (2.1)$$

Here and in the sequel we denote by  $\dot{f}$  and  $\ddot{f}$  the corresponding derivatives. Moreover, for some fixed parameters  $0 < \mathbf{b}_{\min} \leq \mathbf{b}_{\max} < \infty$  we denote by  $\mathcal{B}$  the class of functions  $b$  from  $\mathbf{C}^2(\mathbb{R})$  such that

$$\mathbf{b}_{\min} \leq \inf_{x \in \mathbb{R}} |b(x)| \leq \sup_{x \in \mathbb{R}} \max (|b(x)|, |\dot{b}(x)|, |\ddot{b}(x)|) \leq \mathbf{b}_{\max}. \quad (2.2)$$

Now we set

$$\Theta = \Sigma_{\mathbf{L},\mathbf{M}} \times \mathcal{B} = \{(S, b) : S \in \Sigma_{\mathbf{L},\mathbf{M}} \text{ and } b \in \mathcal{B}\}. \quad (2.3)$$

It is easy to see that the functions from  $\Sigma_{\mathbf{L},\mathbf{M}}$  are uniformly bounded on  $[\mathbf{x}_0, \mathbf{x}_1]$ , i.e.

$$s^* = \sup_{\mathbf{x}_0 \leq x \leq \mathbf{x}_1} \sup_{S \in \Sigma_{\mathbf{L},\mathbf{M}}} S^2(x) < \infty. \quad (2.4)$$

It should be noted that, for any  $\vartheta \in \Theta$ , there exists an invariant density for the process (1.3) which is defined as

$$\mathbf{q}_{\vartheta}(x) = \left( \int_{\mathbb{R}} b^{-2}(z) e^{\tilde{S}(z)} dz \right)^{-1} b^{-2}(x) e^{\tilde{S}(x)}, \quad (2.5)$$

where  $\tilde{S}(x) = 2 \int_0^x b^{-2}(v) S(v) dv$  (see, e.g., [23], Ch.4, 18, Th2). It is easy to see that this density is uniformly bounded on the class (2.3), i.e.

$$\mathbf{q}^* = \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \mathbf{q}_{\vartheta}(x) < +\infty \quad (2.6)$$

and bounded away from zero on the interval  $[-\mathbf{x}_*, \mathbf{x}_*]$ , i.e.

$$\mathbf{q}_* = \inf_{|x| \leq \mathbf{x}_*} \inf_{\vartheta \in \Theta} \mathbf{q}_{\vartheta}(x) > 0. \quad (2.7)$$

For any  $\mathbf{q}_{\vartheta}$ -integrable  $\mathbb{R} \rightarrow \mathbb{R}$  function  $f$ , i.e.  $\int_{\mathbb{R}} |f(x)| \mathbf{q}_{\vartheta}(x) dx < \infty$ , we set

$$\mathbf{m}_{\vartheta}(f) = \int_{\mathbb{R}} f(x) \mathbf{q}_{\vartheta}(x) dx. \quad (2.8)$$

We need the following condition for the observation frequency.

**A<sub>1</sub>)** The frequency  $\delta$  in the observations (1.2) has the following form

$$\delta = \delta_T = \frac{1}{(T+1)l_T}, \quad (2.9)$$

where the function  $l_T$  is such that,

$$\lim_{T \rightarrow \infty} \frac{l_T}{\ln T} = +\infty. \quad (2.10)$$

For example, one can take  $l_T = (\ln T)^{1+\iota}$  for some  $\iota > 0$ .

In this paper we consider the quadratic risk defined for any estimator  $\widehat{S}$  as

$$\mathcal{R}(\widehat{S}, S) = \mathbf{E}_\vartheta \int_{\mathbf{x}_0}^{\mathbf{x}_1} |\widehat{S}(x) - S(x)|^2 dx, \quad (2.11)$$

where  $\mathbf{E}_\vartheta$  is the expectation with respect to the distribution of the process (1.3) under the parameter  $\vartheta \in \Theta$ .

**Remark 2.1.** Note that we consider the estimation problem only for the drift function  $S$ , i.e. in this case the diffusion coefficient  $b$  is considered as a nuisance parameter.

### 3 Sequential point-wise estimation

In order to obtain a reliable estimator of the function  $S$  on the interval  $[\mathbf{x}_0, \mathbf{x}_1]$  we need some efficient point-wise estimators of this function. To give such estimators, we begin with the partition of the interval  $[\mathbf{x}_0, \mathbf{x}_1]$  by points  $(z_k)_{0 \leq k \leq n}$  defined as

$$z_k = \mathbf{x}_0 + \frac{k(\mathbf{x}_1 - \mathbf{x}_0)}{n}, \quad (3.1)$$

where  $n = n(T)$  is an integer-valued function of  $T$  such that

$$\lim_{T \rightarrow \infty} \frac{n(T)}{\sqrt{T}} = 1. \quad (3.2)$$

To construct an efficient procedure for  $S(z_k)$ , at any point  $z_k$ , we need to use some estimators of the invariant density  $\mathbf{q}_\vartheta(\cdot)$  and the squared diffusion coefficient  $b^2(\cdot)$ . To this end, we will estimate these functions by making use of the first  $N_0$  observations, i.e.  $(y_{t_j})_{1 \leq j \leq N_0}$ .

We begin with estimating the function  $S$ . We set

$$N_0 = [N^\gamma(T)] \quad \text{with} \quad 3/4 < \gamma < 1. \quad (3.3)$$

Now we estimate  $S(z_k)$ , at every point  $z_k$ , by making use of the observations  $(y_{t_j})_{N_0+1 \leq j \leq N}$ . To this end we use the sequential kernel estimator from [15] with the indicator kernel function  $\chi(y) = \mathbf{1}_{\{|y| \leq 1\}}$  and the sample size given by the stopping time

$$\tau_k = \inf \left\{ l \geq N_0 + 1 : \sum_{j=N_0+1}^l \chi_{j,k}(h) \geq H_k \right\}, \quad (3.4)$$

where

$$\chi_{j,k}(h) = \chi \left( \frac{y_{t_{j-1}} - z_k}{h} \right), \quad h = \frac{\mathbf{x}_1 - \mathbf{x}_0}{2n}$$

and  $H_k$  is some positive random threshold which will be specified later. On the set

$$\Gamma_k = \{\tau_k \leq N\} \quad (3.5)$$

we define the correction coefficient  $0 < \varkappa_k \leq 1$  from the equation

$$\sum_{j=N_0+1}^{\tau_k-1} \chi_{j,k}(h) + \varkappa_k \chi_{\tau_k,k}(h) = H_k. \quad (3.6)$$

Moreover, on the  $\Gamma_k^c$  we set  $\varkappa_k = 0$ . Using this definition we introduce the weight sequence

$$\tilde{\varkappa}_{j,k} = \mathbf{1}_{\{j < \tau_k\}} + \varkappa_k \mathbf{1}_{\{j = \tau_k\}}. \quad (3.7)$$

One can check directly that, for any  $j \geq 1$ , the coefficients  $\tilde{\varkappa}_{k,j}$  are  $\mathcal{G}_{j-1}$  measurable, where  $\mathcal{G}_j = \sigma(y_{t_l}, 0 \leq l \leq j)$ . Now we define the sequential estimator for  $S(z_k)$  as

$$S_k^* = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \Delta y_{t_j} \mathbf{1}_{\Gamma_k}, \quad (3.8)$$

where  $\Delta y_{t_j} = y_{t_j} - y_{t_{j-1}}$ .

To specify the threshold  $H_k$  we need a truncated estimator  $\tilde{q}(z_k)$  for the invariant density  $\mathbf{q}_\vartheta(z_k)$  given in (3.12) which must be greater than some positive function  $v_T$ . We impose the next condition on this function.

**A<sub>2</sub>)** Assume, that

$$\lim_{T \rightarrow \infty} \left( v_T + \frac{\ln T}{T(v_T)^2} + \frac{\ln T}{l_T(v_T)^5} \right) = 0. \quad (3.9)$$

For example, one can take  $v_T = \ln^{-\iota}(T+1)$  and  $l_T \geq \ln^{1+6\iota} T$ , for some  $\iota > 0$ . Finally, as in [22], we set

$$H_k = h(N - N_0)(2\tilde{q}(z_k) - v_T). \quad (3.10)$$

Note that in [12] it has been shown that such form of the threshold  $H_k$  provides the optimal convergence rate.

**Remark 3.1.** *It should be noted that it is not necessary to have a good estimator for the invariant density. As we will see later it suffices to construct some rough estimators, just to correct corresponding coefficients.*

**Remark 3.2.** *In fact, our procedure uses only the observations belonging to the interval  $[z_k - h, z_k + h]$ . This implies that the sample size equals asymptotically to  $2Nh\hat{q}(z_k)$ . This is related with the choice of the estimator kernel that is an indicator function. It is easy to verify (see [14]) that this kernel minimizes the variance of stochastic term in the kernel estimator. Ultimately, the last result provides efficiency of the procedure.*

### 3.1 Estimation of the invariant density

To estimate the density  $\mathbf{q}_\vartheta$  we will make use of the following kernel estimator

$$\hat{q}(z_k) = \frac{1}{2N_0 h_0} \sum_{j=2}^{N_0} \chi_{j,k}(h_0), \quad (3.11)$$

where

$$h_0(T) = \frac{1}{\sqrt{T_0}} \quad \text{and} \quad T_0 = \delta N_0.$$

We set

$$\tilde{q}(z_k) = \begin{cases} (v_T)^{1/2}, & \text{if } \hat{q}(z_k) < (v_T)^{1/2}; \\ \hat{q}(z_k), & \text{if } (v_T)^{1/2} \leq \hat{q}(z_k) \leq (v_T)^{-1/2}; \\ (v_T)^{-1/2}, & \text{if } \hat{q}(z_k) > (v_T)^{-1/2}. \end{cases} \quad (3.12)$$

The properties of this estimator are studied in [22].



### 3.2 Estimation of the squared diffusion coefficient

To estimate the squared diffusion coefficient  $b^2(z_k)$  we use the following sequential procedure. First we define the sample size for this procedure. To do this we set

$$\tau_{*,k} = \inf\{j \geq 1 : \sum_{l=1}^j \chi_{l,k}(h_0) \geq H_*\} \wedge N_0, \quad (3.13)$$

where

$$h_0 = \frac{1}{\sqrt{T_0}} \quad \text{and} \quad H_* = \frac{1}{\ln(T+1)} h_0 N_0.$$

Then we set

$$\widehat{b}_k = \frac{\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) (y_{t_j} - y_{t_{j-1}})^2}{\delta H_*} \mathbf{1}_{\Gamma_{*,k}}, \quad (3.14)$$

where  $\Gamma_{*,k} = \{\sum_{j=1}^{N_0} \chi_{j,k}(h_0) \geq H_*\}$ .

This estimator satisfies the following property.

**Proposition 3.1.** *For any  $a > 0$ ,*

$$\limsup_{T \rightarrow \infty} \max_{\vartheta \in \Theta} T^{\gamma-1/2-a} \mathbf{E}_{\vartheta} |\widehat{b}_k - b^2(z_k)| = 0. \quad (3.15)$$

**Remark 3.3.** *Note that in the case of known diffusion coefficient  $b(\cdot)$ , we can take  $\widehat{b}_k = b^2(z_k)$ .*

## 4 Regression model

In order to obtain an oracle inequality for discrete time data, we shall pass to a regression model by the same way as in [19]. From (1.3) and (3.8) it follows that, for any  $k$  on the set  $\Gamma_k$ ,

$$S_k^* = S(z_k) + B_k + \zeta_k, \quad (4.1)$$

where the approximation term

$$B_k = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\widetilde{\mathcal{L}}_{j,k}} \chi_{j,k}(h) \int_{t_{j-1}}^{t_j} S(y_u) du - S(z_k) \quad (4.2)$$

and the stochastic term

$$\zeta_k = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k} \chi_{j,k}(h)} \int_{t_{j-1}}^{t_j} b(y_u) dW_u.$$

Therefore, if we set

$$\xi_k = \frac{1}{\sqrt{\delta H_k}} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k} \chi_{j,k}(h)} \Delta W_{t_j} \quad \text{and} \quad \sigma_k = \frac{b(z_k)}{\sqrt{\delta H_k}}, \quad (4.3)$$

then we can rewrite the stochastic term as

$$\zeta_k = \tilde{B}_k + \sigma_k \xi_k,$$

where

$$\tilde{B}_k = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k} \chi_{j,k}(h)} \int_{t_{j-1}}^{t_j} (b(y_s) - b(z_k)) dW_s. \quad (4.4)$$

So, setting

$$\mathbf{G}_* = \bigcap_{k=1}^n \Gamma_k \quad \text{and} \quad Y_k = S_k^* \mathbf{1}_{\mathbf{G}_*}, \quad (4.5)$$

we can represent the equality (4.1) on the set  $\mathbf{G}_*$  as nonparametric regression model, i.e. for any  $1 \leq k \leq n$ ,

$$Y_k = S(z_k) + \mathbf{g}_k + \sigma_k \xi_k, \quad \mathbf{g}_k = B_k + \tilde{B}_k. \quad (4.6)$$

Note that the coefficients  $(\sigma_l)_{1 \leq l \leq n}$  are random variables and using their definitions one can obtain the following bounds

$$\sigma_{0,*} \leq \min_{1 \leq l \leq n} \sigma_l^2 \leq \max_{1 \leq l \leq n} \sigma_l^2 \leq \sigma_{1,*}, \quad (4.7)$$

where

$$\sigma_{0,*} = \frac{v_T \mathbf{b}_{\min}}{\delta N h} \quad \text{and} \quad \sigma_{1,*} = \frac{\mathbf{b}_{\max}}{v_T \delta (N - N_0) h}.$$

Now we set

$$\mathbf{g}_T^* = T \max_{1 \leq k \leq n} \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \mathbf{g}_k^2 \mathbf{1}_{\mathbf{G}_*} \quad (4.8)$$

**Proposition 4.1.** *Assume that the observation frequency  $\delta$  and the bandwidth  $h$  are defined in (2.9) and (3.4), respectively. Then, for any  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} \frac{\mathbf{g}_T^*}{T^a} = 0. \quad (4.9)$$

Concerning the random variables  $(\xi_k)_{1 \leq k \leq n}$ , we can show the following property.

**Proposition 4.2.** *The random variables  $(\xi_k)_{1 \leq k \leq n}$  are  $\mathcal{N}(0, 1)$  i.i.d. conditionally to  $\mathcal{G}_{N_0}$ .*

Similarly to the proof of Proposition 4.5 in [22] and using Theorem A.1 we can obtain the following property for the set  $\mathbf{G}_*$  in (4.5).

**Proposition 4.3.** *Assume that the conditions  $\mathbf{A}_1)$  –  $\mathbf{A}_2)$  hold. Then, for any  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^a \sup_{\vartheta \in \Theta} \mathbf{P}_{\vartheta}(\mathbf{G}_*^c) = 0. \quad (4.10)$$

We estimate the parameter  $\sigma_l^2$  as follows:

$$\hat{\sigma}_l = \frac{\hat{b}_l}{\delta H_l}, \quad (4.11)$$

where  $\hat{b}_l$  is the estimator of the squared diffusion coefficient  $b^2(z_l)$  defined in (3.14).

In order to write the oracle inequality, we need to study the properties of the last estimator. To this end we set

$$\varpi_T^* = n \max_{1 \leq l \leq n} \mathbf{E}_{\vartheta} |\hat{\sigma}_l - \sigma_l^2|. \quad (4.12)$$

**Proposition 4.4.** *Assume that the conditions  $\mathbf{A}_1)$  –  $\mathbf{A}_2)$  hold. Then, for any  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^{\gamma-1/2-a} \varpi_T^* = 0. \quad (4.13)$$

**Remark 4.1.** *Note, that the main technical difference of the considering case from the estimation problem on the complete observations considered in [19] is such that in the heteroscedastic regression model (4.6) the noise variances tend to zero as  $n \rightarrow \infty$  since the number of the point is proportionally to  $\sqrt{T}$ , but not to  $T$  as in [19]. By this reason we need to normalize the difference in (4.12) by  $n$ .*

## 5 Model selection

First we choose a basis  $(\phi_j)_{j \geq 1}$  in  $\mathcal{L}_2([\mathbf{x}_0, \mathbf{x}_1])$  such that, for any  $1 \leq i, j \leq n$ ,

$$(\phi_i, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \phi_i(z_l) \phi_j(z_l) = \mathbf{1}_{\{i=j\}}. \quad (5.1)$$

For example, one can take the trigonometric basis defined as  $\text{Tr}_1(x) \equiv 1/\sqrt{\mathbf{x}_1 - \mathbf{x}_0}$  and, for  $j \geq 2$ ,

$$\text{Tr}_j(x) = \sqrt{\frac{2}{\mathbf{x}_1 - \mathbf{x}_0}} \begin{cases} \cos(2\pi[j/2]l_0(x)) & \text{for even } j; \\ \sin(2\pi[j/2]l_0(x)) & \text{for odd } j, \end{cases} \quad (5.2)$$

where  $[a]$  denotes the integer part of  $a$  and  $l_0(x) = (x - \mathbf{x}_0)/(\mathbf{x}_1 - \mathbf{x}_0)$ . Note that if  $n$  is odd, then this basis is orthonormal for the empirical inner product, i.e. satisfies the property (5.1). By making use of this property we define the discrete Fourier representation for  $S$  on the sieve (3.1), i.e.,

$$S(z_k) = \sum_{j=1}^n \theta_{j,n} \phi_j(z_k), \quad 1 \leq k \leq n, \quad (5.3)$$

where

$$\theta_{j,n} = (S, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n S(z_l) \phi_j(z_l).$$

Moreover, using the regression model (4.6) we estimate these coefficients as

$$\widehat{\theta}_{j,n} = (Y, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n Y_l \phi_j(z_l). \quad (5.4)$$

By the model (4.6), we obtain on the set  $\mathbf{G}_*$

$$\widehat{\theta}_{j,n} = \theta_{j,n} + \zeta_{j,n}, \quad \zeta_{j,n} = \mathbf{g}_{j,n} + \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \xi_{j,n}, \quad (5.5)$$

where

$$\xi_{j,n} = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{l=1}^n \sigma_l \xi_l \phi_j(z_l), \quad \mathbf{g}_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \mathbf{g}_l \phi_j(z_l).$$

We estimate the values  $S(z_k)$ ,  $1 \leq k \leq n$ , by the weighted least squares estimators

$$\widehat{S}_\lambda(z_k) = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \phi_j(z_k), \quad 1 \leq k \leq n, \quad (5.6)$$

where the weight vector  $\lambda = (\lambda(1), \dots, \lambda(n))'$  belongs to some finite set  $\Lambda$  from  $[0, 1]^n$ . In the sequel, we denote by  $\nu$  the cardinal number of the set

$\Lambda$ ,  $\nu = \text{card}(\Lambda)$ , which is a function of  $T$ , i.e.  $\nu = \nu_T$ . Moreover, we set the norm for  $\Lambda$  as

$$\Lambda_* = \max_{\lambda \in \Lambda} \sum_{j=1}^n \lambda(j) \quad (5.7)$$

which can be a function of  $T$ , i.e.  $\Lambda_* = \Lambda_*(T)$ . We assume that the basis functions and the weights set  $\Lambda$  satisfy the following condition.

**A<sub>3</sub>)** For any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\phi_T^* + \nu_T}{T^a} = 0 \quad \text{and} \quad \lim_{T \rightarrow \infty} \frac{\Lambda_*(T)}{T^{1/6+a}} = 0, \quad (5.8)$$

where  $\phi_T^* = \max_{1 \leq j \leq n} \max_{\mathbf{x}_0 \leq x \leq \mathbf{x}_1} |\phi_j(x)|$ .

To estimate the function  $S$  on the interval  $x \in [\mathbf{x}_0, \mathbf{x}_1]$ , we use the step-function approximation, i.e.,

$$\widehat{S}_\lambda(x) = \sum_{l=1}^n \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < x \leq z_l\}}. \quad (5.9)$$

Now one needs to choose a cost function in order to define an optimal weight  $\lambda \in \Lambda$ . A best candidate for the cost function should be the empirical squared error given by the relation

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

In our case, the empirical squared error is equal to

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \theta_{j,n} + \sum_{j=1}^n \theta_{j,n}^2. \quad (5.10)$$

Since coefficients  $\theta_{j,n}$  are unknown, we need to replace the term  $\widehat{\theta}_{j,n} \theta_{j,n}$  by some estimator which we choose as

$$\widetilde{\theta}_{j,n} = \widehat{\theta}_{j,n} - \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \widehat{\sigma}_{j,n} \quad \text{and} \quad \widehat{\sigma}_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \widehat{\sigma}_l \phi_j^2(z_l), \quad (5.11)$$

where  $\widehat{\sigma}_l$  is the estimator for  $\sigma_l^2$  defined in(4.11). Note that if the diffusion is known, then we take in (5.11)  $\widehat{\sigma}_{j,n} = \sigma_{j,n}$  and

$$\sigma_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(z_l). \quad (5.12)$$

It is clear that the inequalities (4.7) imply

$$\sigma_{0,*} \leq \min_{1 \leq l \leq n} \sigma_{l,n} \leq \max_{1 \leq l \leq n} \sigma_{l,n} \leq \sigma_{1,*}. \quad (5.13)$$

Now, for using the estimator (5.11) instead of  $\theta_{j,n} \widehat{\theta}_{j,n}$  one needs to add to the cost function a corresponding penalty term that we take as

$$\widehat{P}_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \widehat{\sigma}_{j,n} \quad (5.14)$$

if the diffusion is unknown and as

$$P_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \sigma_{j,n} \quad (5.15)$$

when the diffusion is known. Finally, we use the following cost function

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widetilde{\theta}_{j,n} + \rho \widehat{P}_n(\lambda), \quad (5.16)$$

where the positive coefficient  $0 < \rho < 1$  will be specified later.

We define the model selection procedure as

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}. \quad (5.17)$$

**Remark 5.1.** *It should be emphasized that if in the model (1.3) the diffusion coefficient  $b(\cdot)$  is known, then the model selection procedure (5.17) is defined through the minimizing the cost function  $J_n(\lambda)$  with the penalty term (5.15).*

## 6 Main results

First we set

$$\mathbf{U}_T = \frac{\nu}{\nu_T^3} (1 + g_T^* + \Lambda_* \varpi_T^*) + \sqrt{T/\nu_T} \sup_{\vartheta \in \Theta} \sqrt{\mathbf{P}_\vartheta(\mathbf{G}_*^c)}, \quad (6.1)$$

where the function  $\nu_T$  is given in (3.10). Now we give the oracle inequality for the discrete norm (5.1).

**Theorem 6.1.** *Assume that the conditions  $\mathbf{A}_1) - \mathbf{A}_3)$  hold. Then there exists a positive constant  $\check{\mathbf{l}} > 0$  such that for any  $T \geq 1$ ,  $0 < \rho \leq 1/8$  and*

$\vartheta \in \Theta$ , the following oracle inequality holds for the model selection procedure (5.17) :

$$\mathbf{E}_\vartheta \|\widehat{S}_* - S\|_n^2 \leq \frac{1+5\rho}{1-6\rho} \min_{\lambda \in \Lambda} \mathbf{E}_\vartheta \|\widehat{S}_\lambda - S\|_n^2 + \check{\mathbf{I}} \frac{\mathbf{U}_T}{\rho T}. \quad (6.2)$$

Moreover, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\mathbf{U}_T}{T^a} = 0. \quad (6.3)$$

Note that, if the diffusion coefficient  $b(\cdot)$  is known, then the term  $\varpi_T^* = 0$  in the function  $\mathbf{U}_T$ .

Using Lemma A.3 with  $\tilde{\varepsilon} = \rho$  we obtain now the oracle inequality for the risk (2.11).

**Theorem 6.2.** *Assume that the conditions  $\mathbf{A}_1) - \mathbf{A}_3)$  hold. Then, for any  $T \geq 1$ ,  $0 < \rho \leq 1/8$  and  $\vartheta \in \Theta$ , the estimation procedure  $\widehat{S}_*$  defined in (5.17) satisfies the following inequality*

$$\mathcal{R}(\widehat{S}_*, S) \leq \frac{(1+\rho)^2(1+5\rho)}{1-6\rho} \min_{\lambda \in \Lambda} R_n(\widehat{S}_\lambda, S) + \frac{\mathcal{B}_T}{\rho T}, \quad (6.4)$$

where the rest term  $\mathcal{B}_T$  satisfies the limit the property (6.3) for any  $a > 0$ .

In the sequel to obtain the efficient properties for this procedure we will use the special weight coefficients introduced in [17]. To this end consider a numerical grid of the form

$$\mathcal{A} = \{1, \dots, \mathbf{k}^*\} \times \{r_1, \dots, r_{m^*}\},$$

where  $r_i = i\varepsilon$  and  $m^* = \lceil 1/\varepsilon^2 \rceil$ . The both parameters  $\mathbf{k}^* \geq 1$  and  $0 < \varepsilon \leq 1$  are some functions of  $T$ , i.e.  $\mathbf{k}^* = \mathbf{k}_T^*$  and  $\varepsilon = \varepsilon_T$ , such that, for any  $\gamma > 0$ ,

$$\lim_{T \rightarrow \infty} \left( \varepsilon_T + \frac{1}{T^\gamma \varepsilon_T} + \frac{1}{\mathbf{k}_T^*} + \frac{\mathbf{k}_T^*}{\ln T} \right) = 0. \quad (6.5)$$

One can take, for example,

$$\varepsilon_T = \frac{1}{\ln(T+1)} \quad \text{and} \quad \mathbf{k}^* = \bar{k} + \sqrt{\ln(T+1)},$$

for some fixed  $\bar{k} \geq 0$ . For each  $\alpha = (\beta, r) \in \mathcal{A}$  we introduce the weight sequence  $\lambda_\alpha = (\lambda_j(\alpha))_{j \geq 1}$  as

$$\lambda_j(\alpha) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + (1 - (j/\omega_\alpha)^\beta) \mathbf{1}_{\{j_0 < j \leq \omega_\alpha\}}, \quad (6.6)$$

where  $j_0 = j_0(\alpha) = \lceil \omega_\alpha / \ln(n+1) \rceil$ ,  $\omega_\alpha = \omega_* (r n)^{1/(2\beta+1)}$  and

$$\omega_* = (\mathbf{x}_1 - \mathbf{x}_0) \left( \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \right)^{1/(2\beta+1)}.$$

We set

$$\Lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}. \quad (6.7)$$

Note that, in this case, the cardinal  $\nu$  of the set  $\Lambda$  is the function of  $T$ , i.e.  $\nu = \nu_T = \mathbf{k}^* m^*$  and the conditions (6.5) imply that for any  $a > 0$

$$\lim_{T \rightarrow \infty} \frac{\nu_T}{T^a} = 0. \quad (6.8)$$

Moreover, from the definition (6.6) we can obtain that, for any  $\alpha \in \mathcal{A}$ ,

$$\sum_{j=1}^n \lambda_j(\alpha) \leq \omega_\alpha \leq \omega_* \left( \frac{n}{\varepsilon_T} \right)^{1/3}.$$

Therefore, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\Lambda_*}{T^{1/6+a}} = 0. \quad (6.9)$$

Hence, the condition  $\mathbf{A}_3$ ) holds and we obtain the following theorem.

**Theorem 6.3.** *Assume that the conditions  $\mathbf{A}_1$ ) –  $\mathbf{A}_2$ ) hold. Then, the model selection procedure (5.16) with the weights (6.7) satisfies the oracle inequality (6.4) with the rest term satisfying the property (6.3) for any  $a > 0$ .*

**Remark 6.1.** *It should be note that similarly to [19], we will use the inequality (6.4) to provide the efficiency property in the adaptive setting. This means that without using the regularity properties of the unknown function  $S$  we can estimate from above the risk for the model selection procedure by the risk for the efficient estimation procedure constructed on the regularity properties of the function  $S$ . The upper bound for the risk of the model selection procedure follows from the sharp oracle inequality.*

## 7 Properties of the model (5.5)

In order to prove the oracle inequalities, we need to study the conditions introduced in [30] for the general semimartingale model. To this end, for any  $\lambda \in \mathbb{R}^n$ , we define the functions

$$\Xi(\lambda) = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \lambda(j) \xi_{j,n} \quad \text{and} \quad \mathbf{B}(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{\sqrt{n}} \sum_{j=1}^n \lambda(j) \tilde{\xi}_{j,n}, \quad (7.1)$$

where  $\tilde{\xi}_{j,n} = \xi_{j,n}^2 - \sigma_{j,n}$ , the variables  $\xi_{j,n}$  are from (5.5),  $\sigma_{j,n} = \mathbf{E}\xi_{j,n}^2$ .



**Proposition 7.1.** For any  $n \geq 1$  and any  $\lambda = (\lambda_1, \dots, \lambda_n) \in [0, 1]^n$ ,

$$\mathbf{E}_\vartheta \Xi^2(\lambda) \leq \sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)}{n} |\lambda|^2. \quad (7.2)$$

**Proof.** From Proposition 4.2, (5.1), (5.5) and (4.7) we can obtain directly that

$$\begin{aligned} \mathbf{E}_\vartheta \Xi^2(\lambda) &= \frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} \mathbf{E}_\vartheta \sum_{l=1}^n \sigma_l^2 \left( \sum_{j=1}^n \lambda(j) \phi_j(z_l) \right)^2 \\ &\leq \sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} \sum_{j,k=1}^n \lambda(j) \lambda(k) \sum_{l=1}^n \phi_j(z_l) \phi_k(z_l) \\ &= \sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)}{n} \sum_{j=1}^n \lambda^2(j) \frac{(\mathbf{x}_1 - \mathbf{x}_0)}{n} \sum_{l=1}^n \phi_j^2(z_l). \end{aligned}$$

Hence Proposition 7.1.  $\square$

**Proposition 7.2.** For any  $n \geq 1$  and any  $\lambda = (\lambda_1, \dots, \lambda_n) \in [0, 1]^n$ ,

$$\mathbf{E}_\vartheta [\mathbf{B}^2(\lambda) | \mathcal{G}_{N_0}] \leq 6\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0) P_n(\lambda). \quad (7.3)$$

**Proof.** We begin with presentation of the variables  $\tilde{\xi}_{j,n}$  and  $\mathbf{B}(\lambda)$  from (7.1) as follows

$$\tilde{\xi}_{j,n} = \frac{(\mathbf{x}_1 - \mathbf{x}_0)}{n} \sum_{l=1}^n \left( \sigma_l^2 \phi_j^2(z_l) \eta_l + 2\mathbf{1}_{l \geq 2} \xi_l R_{j,l} \right),$$

where  $\eta_l = \xi_l^2 - 1$ ,  $R_{j,l} = \sigma_l \phi_j(z_l) \sum_{r=1}^{l-1} \sigma_r \phi_j(z_r) \xi_r$ , and

$$\mathbf{B}(\lambda) = \frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{n^{3/2}} \sum_{l=1}^n (\eta_l \gamma_{1,l} + 2\xi_l \gamma_{2,l})$$

with  $\gamma_{1,l} = \sigma_l^2 \sum_{j=1}^n \lambda(j) \phi_j^2(z_l)$ ,  $\gamma_{2,l} = \sum_{j=1}^n \lambda(j) R_{j,l} \mathbf{1}_{l \geq 2}$ .

In view of Proposition 4.2, one has

$$\mathbf{E} (\mathbf{B}^2(\lambda) | \mathcal{G}_{N_0}) = \frac{(\mathbf{x}_1 - \mathbf{x}_0)^4}{n^3} \sum_{l=1}^n \left( 2\gamma_{1,l}^2 + 4\mathbf{E} [\gamma_{2,l}^2 | \mathcal{G}_{N_0}] \right) := M_{1,1} + M_{1,2}.$$

Due to the Buniakovski-Cauchy-Schwartz inequality and the property (5.1), one has

$$\begin{aligned}\gamma_{1,l}^2 &= \sigma_l^4 \left( \sum_{j=1}^n \lambda(j) \phi_j^2(z_l) \right)^2 \leq \sigma_l^4 \left( \sum_{j=1}^n \lambda^2(j) \phi_j^2(z_l) \right) \left( \sum_{j=1}^n \phi_j^2(z_l) \right) \\ &\leq \sigma_{1,*} \frac{n}{(\mathbf{x}_1 - \mathbf{x}_0)} \sigma_l^2 \sum_{j=1}^n \lambda^2(j) \phi_j^2(z_l).\end{aligned}$$

This implies for the term  $M_{1,1}$ :

$$\begin{aligned}M_{1,1} &\leq 2\sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)^3}{n^2} \sum_{l=1}^n \sigma_l^2 \sum_{j=1}^n \lambda^2(j) \phi_j^2(z_l) \\ &= 2\sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)^3}{n^2} \sum_{j=1}^n \lambda^2(j) \sum_{l=1}^n \sigma_l^2 \phi_j^2(z_l) = 2\sigma_{1,*} (\mathbf{x}_1 - \mathbf{x}_0) P_n(\lambda).\end{aligned}$$

In order to bound the terme  $M_{1,2}$  we have

$$\begin{aligned}\mathbf{E} [\gamma_{2,l}^2 | \mathcal{G}_{N_0}] &= \sigma_l^2 \sum_{r=1}^{l-1} \sigma_r^2 \left( \sum_{j=1}^n \lambda(j) \phi_j(z_l) \phi_j(z_r) \right)^2 \\ &\leq \sigma_{1,*} \sigma_l^2 \sum_{r=1}^n \sum_{j,k=1}^n \lambda(j) \lambda(k) \phi_j(z_l) \phi_k(z_l) \phi_j(z_r) \phi_k(z_r) \\ &= \sigma_{1,*} \sigma_l^2 \sum_{j,k=1}^n \lambda(j) \lambda(k) \phi_j(z_l) \phi_k(z_l) \sum_{r=1}^n \phi_j(z_r) \phi_k(z_r) \\ &= \sigma_{1,*} \frac{n}{(\mathbf{x}_1 - \mathbf{x}_0)} \sigma_l^2 \sum_{j=1}^n \lambda^2(j) \phi_j^2(z_l),\end{aligned}$$

where the last equality follows from the property (5.1). Therefore, we obtain that the term  $M_{1,2}$  can be estimated as

$$M_{1,2} \leq 4\sigma_{1,*} \frac{(\mathbf{x}_1 - \mathbf{x}_0)^3}{n^2} \sum_{l=1}^n \sigma_l^2 \sum_{j=1}^n \lambda^2(j) \phi_j^2(z_l) = 4\sigma_{1,*} (\mathbf{x}_1 - \mathbf{x}_0) P_n(\lambda).$$

The upper bounds for  $M_{1,1}$  and  $M_{1,2}$  imply

$$\mathbf{E} (\mathbf{B}^2(\lambda) | \mathcal{G}_{N_0}) = M_{1,1} + M_{1,2} \leq 6\sigma_{1,*} (\mathbf{x}_1 - \mathbf{x}_0) P_n(\lambda).$$

Hence Proposition 7.2.  $\square$

## 8 Proofs

### 8.1 Proof of Theorem 6.1

First of all, note that on the set  $\mathbf{G}_*$  we can represent the empirical squared error  $\text{Err}_n(\lambda)$  in the form

$$\text{Err}_n(\lambda) = J_n(\lambda) + 2 \sum_{j=1}^n \lambda(j) \check{\theta}_{j,n} + \|S\|_n^2 - \rho \widehat{P}_n(\lambda) \quad (8.1)$$

with  $\check{\theta}_{j,n} = \widetilde{\theta}_{j,n} - \theta_{j,n} \widehat{\theta}_{j,n}$ . From (5.5) and (5.11) one obtains

$$\check{\theta}_{j,n} = \theta_{j,n} \zeta_{j,n} + \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} (\widetilde{\xi}_{j,n} - \widetilde{\sigma}_{j,n}) + 2 \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \xi_{j,n} \mathbf{g}_{j,n} + \mathbf{g}_{j,n}^2,$$

where  $\widetilde{\xi}_{j,n} = \xi_{j,n}^2 - \sigma_{j,n}$  and  $\widetilde{\sigma}_{j,n} = \widehat{\sigma}_{j,n} - \sigma_{j,n}$ . Setting now

$$M(\lambda) = \sum_{j=1}^n \lambda(j) \theta_{j,n} \zeta_{j,n} \quad \text{and} \quad \mathbf{D}(\lambda) = \sum_{j=1}^n \lambda(j) \widetilde{\sigma}_{j,n}, \quad (8.2)$$

we have from (8.1)

$$\begin{aligned} \text{Err}_n(\lambda) &= J_n(\lambda) + 2M(\lambda) + 2M_1(\lambda) - \frac{2(\mathbf{x}_1 - \mathbf{x}_0)}{n} \mathbf{D}(\lambda) \\ &\quad + \|S\|_n^2 - \rho P_n(\lambda) - \rho \check{P}_n(\lambda), \end{aligned} \quad (8.3)$$

where  $\check{P}_n(\lambda) = \widehat{P}_n(\lambda) - P_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \widetilde{\sigma}_{j,n}$ ,

$$M_1(\lambda) = \frac{1}{\sqrt{n}} \mathbf{B}(\lambda) + \Delta(\lambda) \quad \text{with} \quad \Delta(\lambda) = \Delta_1(\lambda) + \Delta_2(\lambda). \quad (8.4)$$

The function  $\mathbf{B}(\lambda)$  is given in (7.1) and

$$\Delta_1(\lambda) = \sum_{j=1}^n \lambda(j) \mathbf{g}_{j,n}^2 \quad \text{and} \quad \Delta_2(\lambda) = 2 \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \lambda(j) \xi_{j,n} \mathbf{g}_{j,n}.$$

In view of Proposition 7.2, for any  $\lambda \in [0, 1]^n$ ,

$$\mathbf{E}_\vartheta (\mathbf{B}^2(\lambda) | \mathcal{G}_{N_0}) \leq 6\sigma_*(\mathbf{x}_1 - \mathbf{x}_0) P_n(\lambda). \quad (8.5)$$

To estimate the second term in (8.4) note that

$$\sup_{\lambda \in [0, 1]^n} \Delta_1(\lambda) \leq \sum_{j=1}^n \mathbf{g}_{j,n}^2 = \|\mathbf{g}\|_n^2. \quad (8.6)$$

To study the function  $\Delta_2(\lambda)$  we recall that, for any  $\varepsilon > 0$  and any  $x, y \in \mathbb{R}$ ,

$$2xy \leq \varepsilon x^2 + \varepsilon^{-1}y^2. \quad (8.7)$$

Therefore, for any  $0 < \varepsilon < 1$ ,

$$\begin{aligned} |\Delta_2(\lambda)| &\leq \varepsilon \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 + \frac{\|\mathbf{g}\|_n^2}{\varepsilon} \\ &= \varepsilon P_n(\lambda) + \varepsilon \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} + \frac{\|\mathbf{g}\|_n^2}{\varepsilon}, \end{aligned}$$

where the vector  $\lambda^2 = (\lambda^2(j))_{1 \leq j \leq n}$ . Thus, for any  $\lambda \in [0, 1]^n$ ,

$$|\Delta(\lambda)| \leq \varepsilon P_n(\lambda) + \varepsilon \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} + 2\varepsilon^{-1} \|\mathbf{g}\|_n^2.$$

This implies

$$2|M_1(\lambda)| \leq 2 \frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2 \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} + 2\varepsilon P_n(\lambda) + 4\varepsilon^{-1} \|\mathbf{g}\|_n^2. \quad (8.8)$$

Then, taking into account here that  $P_n(\lambda^2) \leq P_n(\lambda)$ , one gets, for any  $0 < \varepsilon < 1$  and any  $\lambda \in \Lambda$ ,

$$2 \frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2 \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} \leq \varepsilon P_n(\lambda) + \frac{2}{\varepsilon n} \left( \frac{\mathbf{B}^2(\lambda)}{P_n(\lambda)} + \frac{\mathbf{B}^2(\lambda^2)}{P_n(\lambda^2)} \right) \quad (8.9)$$

Note that the inequalities (4.7) imply that

$$P_{0,n}(\lambda) \leq P_n(\lambda) \leq P_{1,n}(\lambda), \quad (8.10)$$

where

$$P_{0,n}(\lambda) = \frac{\sigma_{0,*}(\mathbf{x}_1 - \mathbf{x}_0)|\lambda|^2}{n} \quad \text{and} \quad P_{1,n}(\lambda) = \frac{\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0)|\lambda|^2}{n}.$$

From the inequalities (8.9)-(8.10) it follows

$$2 \frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2 \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} \leq \varepsilon P_n(\lambda) + \frac{2}{\varepsilon \sigma_{0,*}(\mathbf{x}_1 - \mathbf{x}_0)} \mathbf{B}^*(\lambda), \quad (8.11)$$

where

$$\mathbf{B}^*(\lambda) = \left( \frac{\mathbf{B}^2(\lambda)}{|\lambda|^2} + \frac{\mathbf{B}^2(\lambda^2)}{|\lambda^2|^2} \right).$$

By choosing  $\varepsilon = \rho/3$  one has from (8.8)

$$2|M_1(\lambda)| \leq \rho P_n(\lambda) + \frac{6}{\rho} \Upsilon_n(\lambda), \quad (8.12)$$

where

$$\Upsilon_n(\lambda) = \frac{\mathbf{B}^*(\lambda)}{\sigma_{0,*}(\mathbf{x}_1 - \mathbf{x}_0)} + 2\|\mathbf{g}\|_n^2.$$

To obtain an upper bound for the empirical risk we evaluate the maximal value of the term  $\mathbf{D}(\cdot)$ , i.e.

$$\mathbf{D}^* = \max_{\lambda \in \Lambda} |\mathbf{D}(\lambda - \lambda_0)|.$$

One can check directly that

$$\mathbf{E}_\vartheta \mathbf{D}^* \leq 2 \sum_{\lambda \in \Lambda} \mathbf{E}_\vartheta |\mathbf{D}(\lambda)| \leq \frac{2\nu \Lambda_*}{n} \varpi_T^*, \quad (8.13)$$

where  $\varpi_T^*$  is defined in (4.12). Moreover, we need to study the term

$$\check{P}_n(\lambda) = \hat{P}_n(\lambda) - P_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \tilde{\sigma}_{j,n}. \quad (8.14)$$

To this end, denoting

$$P^* = \sup_{\lambda \in \Lambda} |\check{P}_n(\lambda)|,$$

we obtain that

$$\begin{aligned} \mathbf{E}_\vartheta P^* &\leq \sum_{\lambda \in \Lambda} \mathbf{E}_\vartheta |\check{P}_n(\lambda)| \leq \sum_{\lambda \in \Lambda} \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \mathbf{E}_\vartheta |\hat{\sigma}_{j,n} - \sigma_{j,n}| \\ &\leq \sum_{\lambda \in \Lambda} \frac{\mathbf{x}_1 - \mathbf{x}_0}{n^2} |\lambda|^2 \varpi_T^* \leq \frac{\mathbf{x}_1 - \mathbf{x}_0}{n^2} \Lambda_* \nu \varpi_T^*. \end{aligned} \quad (8.15)$$

From (8.3) we obtain that, for some fixed  $\lambda_0 \in \Lambda$ ,

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) - \text{Err}_n(\lambda_0) &= J_n(\hat{\lambda}) - J_n(\lambda_0) + 2M(\hat{\mu}) - \frac{2(\mathbf{x}_1 - \mathbf{x}_0)}{n} \mathbf{D}(\hat{\mu}) \\ &\quad + 2(M_1(\hat{\lambda}) - M_1(\lambda_0)) - \rho(P_n(\hat{\lambda}) - P_n(\lambda_0)) \\ &\quad - \rho(\check{P}_n(\hat{\lambda}) - \check{P}_n(\lambda_0)), \end{aligned}$$

where  $\widehat{\mu} = \widehat{\lambda} - \lambda_0$ . By the definition of  $\widehat{\lambda}$  in (5.17) we obtain on the set  $\mathbf{G}_*$

$$\begin{aligned} \text{Err}_n(\widehat{\lambda}) &\leq \text{Err}_n(\lambda_0) + 2M(\widehat{\mu}) + \frac{6}{\rho}\Upsilon_n(\widehat{\lambda}) \\ &\quad + \frac{2(\mathbf{x}_1 - \mathbf{x}_0)}{n}\mathbf{D}^* + 2\rho P_n(\lambda_0) - \rho(\check{P}_n(\widehat{\lambda}) - \check{P}_n(\lambda_0)). \end{aligned} \quad (8.16)$$

To study the terme  $\Upsilon_n(\lambda)$  in the previous inequality, we begin with the terme  $\mathbf{B}^*(\lambda)$ . From (8.5) and (8.10) it follows that

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \mathbf{B}^*(\lambda) &\leq \mathbf{E}_\vartheta \left( \frac{B^2(\lambda)}{|\lambda|^2} + \frac{B^2(\lambda^2)}{|\lambda^2|^2} \right) \\ &\leq 6\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0) \mathbf{E}_\vartheta \left( \frac{P_n(\lambda)}{|\lambda|^2} + \frac{P_n(\lambda^2)}{|\lambda^2|^2} \right) \\ &\leq 12 \frac{\sigma_{1,*}^2(\mathbf{x}_1 - \mathbf{x}_0)^2}{n}. \end{aligned}$$

To estimate the norm  $\|\mathbf{g}\|_n^2$  note that  $g_l g_k = 0$  for  $l \neq k$ . Therefore,

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \|\mathbf{g}\|_n^2 &= \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \sum_{j=1}^n g_{j,n}^2 \\ &\leq \frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} \mathbf{E}_\vartheta \sum_{j=1}^n \left( \sum_{l=1}^n g_l \phi_j(z_l) \right)^2 \\ &= \frac{(\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} \mathbf{E}_\vartheta \sum_{j=1}^n \sum_{l=1}^n g_l^2 \phi_j^2(z_l) \leq (\mathbf{x}_1 - \mathbf{x}_0) \frac{g_T^*}{T}, \end{aligned}$$

where  $\mathbf{g}_T^*$  is given by (4.8). This implies

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \Upsilon_n &= \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \left( \frac{\mathbf{B}^*(\lambda)}{\sigma_{0,*}(\mathbf{x}_1 - \mathbf{x}_0)} + 2\|\mathbf{g}\|_n^2 \right) \\ &\leq 2(\mathbf{x}_1 - \mathbf{x}_0) \left( \frac{6\sigma_{1,*}^2}{n\sigma_{0,*}} + \frac{\mathbf{g}_T^*}{T} \right). \end{aligned} \quad (8.17)$$

Let us study now the term  $M$  in (8.3). For any  $\lambda \in \Lambda$ , we represent it as

$$M(\mu) = Z(\mu) + V(\mu) \quad \text{and} \quad \mu = \lambda - \lambda_0, \quad (8.18)$$

where

$$Z(\mu) = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \mu(j) \theta_{j,n} \xi_{j,n} \quad \text{and} \quad V(\mu) = \sum_{j=1}^n \mu(j) \theta_{j,n} \mathbf{g}_{j,n}.$$

We begin with the weighted discrete Fourier transformation, i.e. we set

$$\check{S}_\mu = \sum_{j=1}^n \mu(j) \theta_{j,n} \phi_j. \quad (8.19)$$

Due to definition of  $\xi_{j,n}$  in (5.5), we can estimate the term  $Z(\mu)$  as

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} Z^2(\mu) \leq \frac{\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0)}{n} \|\check{S}_\mu\|_n^2. \quad (8.20)$$

Moreover, using the inequality (8.7) with  $\varepsilon = \rho$ , we obtain

$$2V(\mu) = 2 \sum_{j=1}^n \mu(j) \theta_{j,n} \mathbf{g}_{j,n} \leq \rho \|\check{S}_\mu\|_n^2 + \frac{\|\mathbf{g}\|_n^2}{\rho}. \quad (8.21)$$

Therefore, on the set  $\mathbf{G}_*$

$$2M(\mu) \leq 2\rho \|\check{S}_\mu\|_n^2 + \frac{Z^*}{n\rho} + \frac{\|\mathbf{g}\|_n^2}{\rho}, \quad (8.22)$$

where

$$Z^* = \sup_{\mu \in \Lambda - \lambda_0} \frac{nZ^2(\mu)}{\|\check{S}_\mu\|_n^2}.$$

It is clear, that the upper bound (8.20) yields

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} Z^* \leq \sum_{\mu \in \Lambda - \lambda_0} \frac{n \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} Z^2(\mu)}{\|\check{S}_\mu\|_n^2} \leq \nu \sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0). \quad (8.23)$$

To estimate the norm  $\|\check{S}_\mu\|_n^2$  note that in view of (5.5) on the set  $\mathbf{G}_*$

$$\begin{aligned} \|\check{S}_\mu\|_n^2 - \|\widehat{S}_\mu\|_n^2 &= \sum_{j=1}^n \mu^2(j) (\theta_{j,n}^2 - \widehat{\theta}_{j,n}^2) \leq -2 \sum_{j=1}^n \mu^2(j) \theta_{j,n} \zeta_{j,n} \\ &= -2Z_1(\mu) - 2V_1(\mu), \end{aligned} \quad (8.24)$$

where

$$Z_1(\mu) = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \mu^2(j) \theta_{j,n} \xi_{j,n} \quad \text{and} \quad V_1(\mu) = \sum_{j=1}^n \mu^2(j) \theta_{j,n} \mathbf{g}_{j,n}.$$

Taking into account that  $|\mu(j)| \leq 1$ , similarly to inequality (8.20), we find

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} Z_1^2(\mu) \leq \frac{\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0)}{n} \|\check{S}_\mu\|_n^2.$$

Moreover, for the random variable

$$Z_1^* = \sup_{\mu \in \Lambda - \lambda_0} \frac{nZ_1^2(\mu)}{\|\check{S}_\mu\|_n^2},$$

we obtain the same upper bound as in (8.23), i.e.

$$\mathbf{E}_\vartheta Z_1^* \mathbf{1}_{\mathbf{G}_*} \leq \nu \sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0). \quad (8.25)$$

Furthermore, similarly to (8.21) we estimate the second term in (8.24) as

$$2|V_1(\mu)| \leq \rho \|\check{S}_\mu\|_n^2 + \frac{\|\mathbf{g}\|_n^2}{\rho}.$$

Therefore, on the set  $\mathbf{G}_*$

$$\|\check{S}_\mu\|_n^2 \leq \|\widehat{S}_\mu\|_n^2 + 2\rho \|\check{S}_\mu\|_n^2 + \frac{Z_1^*}{n\rho} + \frac{\|\mathbf{g}\|_n^2}{\rho},$$

i.e.

$$\|\check{S}_\mu\|_n^2 \leq \frac{1}{1-2\rho} \|\widehat{S}_\mu\|_n^2 + \frac{1}{(1-2\rho)\rho} \left( \frac{Z_1^*}{n} + \|\mathbf{g}\|_n^2 \right). \quad (8.26)$$

Using this inequality in (8.22) and putting  $Z_2^* = Z_2^* + Z_1^*$  yield on the set  $\mathbf{G}_*$

$$\begin{aligned} 2M(\widehat{\mu}) &\leq \frac{2\rho}{1-2\rho} \|\widehat{S}_{\widehat{\mu}}\|_n^2 + \frac{1}{\rho(1-2\rho)} \left( \frac{Z_2^*}{n} + \|\mathbf{g}\|_n^2 \right) \\ &\leq \frac{4\rho(\text{Err}_n(\widehat{\lambda}) + \text{Err}_n(\lambda_0))}{1-2\rho} + \frac{1}{\rho(1-2\rho)} \left( \frac{Z_2^*}{n} + \|\mathbf{g}\|_n^2 \right). \end{aligned}$$

Using this bound in (8.16), we obtain that

$$\begin{aligned} \text{Err}_n(\widehat{\lambda}) &\leq \frac{1+2\rho}{1-6\rho} \text{Err}_n(\lambda_0) + \frac{2\rho(1-2\rho)}{1-6\rho} P_n(\lambda_0) + \frac{1}{\rho(1-6\rho)} \left( \frac{Z_2^*}{n} + \|\mathbf{g}\|_n^2 \right) \\ &\quad + \frac{1-2\rho}{1-6\rho} \left( \frac{6}{\rho} \Upsilon_n(\widehat{\lambda}) + \frac{4(\mathbf{x}_1 - \mathbf{x}_0)}{n} \mathbf{D}^* - \rho(\check{P}_n(\widehat{\lambda}) - \check{P}_n(\lambda_0)) \right). \end{aligned}$$



Using here (8.13), (8.15), (8.17), (8.23), (8.25), we get

$$\begin{aligned} \mathbf{E}_\vartheta \text{Err}_n(\widehat{\lambda}) \mathbf{1}_{\mathbf{G}_*} &\leq \frac{1+2\rho}{1-6\rho} \mathbf{E}_\vartheta \text{Err}_n(\lambda_0) \mathbf{1}_{\mathbf{G}_*} + \frac{2\rho(1-2\rho)}{1-6\rho} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} P_n(\lambda_0) \\ &\quad + \frac{(\mathbf{x}_1 - \mathbf{x}_0)}{\rho(1-6\rho)} \left( \frac{(74-144\rho)\sigma_{1,*}(\nu + \bar{\sigma}_*)}{n} + \frac{(13-24\rho)g_T^*}{T} \right) \\ &\quad + (\mathbf{x}_1 - \mathbf{x}_0) \frac{(1-2\rho)(8+2\rho)}{(1-6\rho)n^2} \nu \Lambda_* \varpi_T^*, \end{aligned}$$

where  $\bar{\sigma}_* = \sigma_{1,*}/\sigma_{0,*}$ . From Proposition A.4 with  $\varepsilon = 2\rho$  it follows

$$\begin{aligned} \mathbf{E}_\vartheta \text{Err}_n(\widehat{\lambda}) \mathbf{1}_{\mathbf{G}_*} &\leq K_0 \mathbf{E}_\vartheta \text{Err}_n(\lambda_0) \mathbf{1}_{\mathbf{G}_*} \\ &\quad + (\mathbf{x}_1 - \mathbf{x}_0) \left( K_1 \frac{\sigma_{1,*}}{\rho(1-6\rho)n} + K_2 \frac{g_T^*}{T} + K_3 \frac{\nu \Lambda_* \varpi_T^*}{n^2} \right) \\ &\quad + \frac{2\|S\|_n \sqrt{(\mathbf{x}_1 - \mathbf{x}_0)\sigma_{1,*}}}{n} \sqrt{\mathbf{P}_\vartheta(\mathbf{G}_*^c)}, \end{aligned} \quad (8.27)$$

where

$$K_0 = \frac{1-12\rho^2}{(1-4\rho)(1-6\rho)}, \quad K_1 = \rho(1-2\rho) + (74-144\rho)(\nu + \bar{\sigma}_*),$$

$$K_2 = \frac{13-75\rho+94\rho^2}{\rho(1-4\rho)(1-6\rho)} \quad \text{and} \quad K_3 = \frac{(1-2\rho)(8+2\rho)}{(1-6\rho)}.$$

It is easy to see that, for  $0 < \rho \leq 1/8$ ,

$$\frac{1+4\rho}{1-6\rho} < K_0 \leq \frac{1+5\rho}{1-6\rho}, \quad K_1 < 1 + 74(\nu + \bar{\sigma}_*),$$

$$K_2 < \frac{39}{\rho(1-6\rho)} \quad \text{and} \quad K_3 < \frac{8}{1-6\rho}.$$

Using these bounds and replacing in (8.27)

$$\mathbf{E}_\vartheta \text{Err}_n(\widehat{\lambda}) \mathbf{1}_{\mathbf{G}_*} \quad \text{and} \quad \mathbf{E}_\vartheta \text{Err}_n(\lambda_0) \mathbf{1}_{\mathbf{G}_*}$$

by  $\mathbf{E}_\vartheta \|\widehat{S}_* - S\|_n^2 - \|S\|_n^2 \mathbf{P}_\vartheta(\mathbf{G}_*^c)$  and  $\mathbf{E}_\vartheta \|\widehat{S}_{\lambda_0} - S\|_n^2 - \|S\|_n^2 \mathbf{P}_\vartheta(\mathbf{G}_*^c)$ , respectively, we come to the inequality (6.2):

$$\mathbf{E}_\vartheta \|\widehat{S}_* - S\|_n^2 \leq \frac{1+5\rho}{1-6\rho} \mathbf{E}_\vartheta \|\widehat{S}_{\lambda_0} - S\|_n^2 + \check{\mathbf{I}} \frac{\mathbf{U}_T}{\rho T}, \quad (8.28)$$

where

$$\mathbf{U}_T = \frac{\nu}{v_T^3} (1 + g_T^* + \Lambda_* \varpi_T^*) + \sqrt{(T/v_T)} \sup_{\vartheta \in \Theta} \sqrt{\mathbf{P}_\vartheta(\mathbf{G}_*^c)},$$

$$\check{\mathbf{I}} = \sup_{T>0} \max \left\{ 4(\mathbf{x}_1 - \mathbf{x}_0) \max \left[ \frac{\mathbf{b}_{\max}(v_T^2(1 + 74\nu) + 74\bar{\mathbf{b}})}{\nu}, \frac{39v_T^3}{\nu}, 1 \right], B \right\}$$

with  $\bar{\mathbf{b}} = \mathbf{b}_{\max}/\mathbf{b}_{\min}$ ,  $B = 2M\mathbf{b}_{\max}\sqrt{\mathbf{x}_1 - \mathbf{x}_0}$ .

The weight  $\lambda_0 \in \Lambda$  being arbitrary in (8.28), we come to the oracle inequality (6.2). The property (6.3) follows directly from asymptotic behaviour of the right-hand part components of the function  $\mathbf{U}_T$ . Hence Theorem 6.1.  $\square$

## 8.2 Proof of Proposition 4.1

First, note that from (4.2) we can represent the term  $B_k$  in the following form

$$B_k = B_{1,k} + B_{2,k} + B_{3,k}$$

where

$$B_{1,k} = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) \varrho_j, \quad \varrho_j = \int_{t_{j-1}}^{t_j} (S(y_u) - S(y_{t_{j-1}})) du,$$

$$B_{2,k} = \frac{1}{H_k} \sum_{j=N_0+1}^{\tau_k} \tilde{\chi}_{j,k} \chi_{j,k}(h) (S(y_{t_{j-1}}) - S(z_k)),$$

$$B_{3,k} = \frac{1}{H_k} \sum_{j=N_0+1}^{\tau_k} (1 - \sqrt{\tilde{\chi}_{j,k}}) \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) S(y_{t_{j-1}}).$$

In Appendix we show that there exists some constant  $C^* > 0$  such that

$$\max_{1 \leq k \leq n} \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta B_{1,k}^2 \leq C^* \delta. \quad (8.29)$$

Moreover, using the definition (2.1), we obtain that

$$|B_{2,k}| \leq \frac{Mh}{H_k} \quad \text{and} \quad |B_{3,k}| \leq \frac{M}{H_k}.$$

Taking into account here the property (2.9), the definition of the bandwidth  $h$  in (3.4) and that for sufficiently large  $T$  the threshold  $H_k \geq hN\sqrt{v_T}/2$ , we obtain that

$$\lim_{T \rightarrow \infty} T \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta B_k^2 = 0.$$

Moreover, the term  $\tilde{B}_k$  defined in (4.4) we can represent as

$$\tilde{B}_k = \tilde{B}_{1,k} + \tilde{B}_{2,k},$$

where

$$\begin{aligned} \tilde{B}_{1,k} &= \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) \varpi_j^*, \quad \varpi_j^* = \int_{t_{j-1}}^{t_j} (b(y_u) - b(y_{t_{j-1}})) dW_u, \\ \tilde{B}_{2,k} &= \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) (b(y_{t_{j-1}}) - b(z_k)) \Delta W_{t_j}. \end{aligned}$$

In Appendix we show that there exists some constant  $C^* > 0$  such that for sufficiently large  $T$

$$\max_{1 \leq k \leq n} \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \tilde{B}_{1,k}^2 \leq \frac{8C^* b_{\max}^2 \delta}{hT \sqrt{v_T}}. \quad (8.30)$$

and

$$\max_{1 \leq k \leq n} \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \tilde{B}_{2,k}^2 \leq \frac{8b_{\max}^2 h}{T \sqrt{v_T}}. \quad (8.31)$$

From here we obtain the limit equation (4.9). Hence Proposition 4.1.  $\square$

**Acknowledgements.** This work was partially supported by the research project no. 2.3208.2017/4.6 (the Ministry of Science and Higher Education of the Russian Federation). The work of the last author was partially supported by the Russian Federal Professor program, project no. 1.472.2016/1.4 (the Ministry of Science and Higher Education of the Russian Federation).

## A Appendix

### A.1 Concentration inequalities

In this section we remind the concentration inequalities from [20] which will be used to study non asymptotic properties of the sequential estimation procedures. We set

$$\mathbf{D}_{N,h}(z) = \sum_{j=1}^N \left( \frac{1}{h} \chi \left( \frac{y_{t_j} - z}{h} \right) - \bar{\mathbf{q}}_{\vartheta,h}(z) \right) \delta, \quad (\text{A.1})$$

where  $N = \lceil T/\delta \rceil$ , the frequency  $\delta$  is defined in (2.9) – (2.10) and  $\bar{\mathbf{q}}_{\vartheta,h}(z) = \int_{-1}^1 \mathbf{q}_{\vartheta}(z + uh) du$ .

**Theorem A.1.** *Assume that the condition (3.9) holds. Then, for any  $z_* > 0$  and  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^a \max_{1 \leq n \leq N} \sup_{h \geq T^{-1/2}} \sup_{|z| \leq z_*} \sup_{\vartheta \in \Theta} \mathbf{P}_{\vartheta} (|\mathbf{D}_{n,h}(z)| \geq v_T T) = 0. \quad (\text{A.2})$$

The proof of this result follows directly from Theorem 2.3 in [20].

If we consider the deviation problem for the first  $N_0$  observations with the  $N_0$  defined in (3.3), we have to check the conditions (3.9) for the time duration

$$T_0 = \delta N_0 \approx \frac{T^{2\gamma-1}}{l_T^{1-\gamma}} \quad \text{as } T \rightarrow \infty.$$

Therefore, the frequency has be represented as

$$\delta = \frac{1}{T_0 l_0(T)} \quad \text{and} \quad l_0(T) = \frac{T l_T}{T_0} \approx T^{2-2\gamma} l_T^{2-\gamma} \quad \text{as } T \rightarrow \infty.$$

So, we need to check the conditions (3.9) by replacing  $T$  with  $T_0$  and  $l_T$  with  $l_0(T)$ , respectively. It is clear that these conditions hold if we take , for example,  $v_T = \ln^{-\iota}(T + 1)$  and  $l_T = \ln^{1+6\iota} T$ , for some  $\iota > 0$ .

**Corollary A.2.** *Assume that the condition (3.9) holds. Then, for any  $z_* > 0$  and  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^a \sup_{h \geq T_0^{-1/2}} \sup_{|z| \leq z_*} \sup_{\vartheta \in \Theta} \mathbf{P}_{\vartheta} (|\mathbf{D}_{N_0,h}(z)| \geq v_T T_0) = 0, \quad (\text{A.3})$$

where  $N_0$  is defined in (3.3) and  $T_0 = \delta N_0$ .

## A.2 Proof of Proposition 3.1

First note that Corollary A.2 implies directly that, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} T^a \mathbf{P}(\Gamma_{*,k}^c) = 0.$$

Using the model (1.3) in (3.14), we obtain that

$$y_{t_j} - y_{t_{j-1}} = \sqrt{\delta} b(y_{t_{j-1}}) \eta_j + U_j,$$

where  $\eta_j = (W_{t_j} - W_{t_{j-1}})/\sqrt{\delta} \sim \mathcal{N}(0, 1)$  and

$$U_j = \int_{t_{j-1}}^{t_j} S(y_u) du + \int_{t_{j-1}}^{t_j} (b(y_u) - b(y_{t_{j-1}})) dW_u.$$

Due to the bound (A.6), there exists some positive constant  $C^*$  such that, for any  $j \geq 1$ ,

$$\sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \left( U_j^2 | \mathcal{F}_{t_{j-1}} \right) \leq C^* \delta^2 \left( 1 + y_{t_{j-1}}^2 \right). \quad (\text{A.4})$$

So, the squared difference can be represented as

$$(y_{t_j} - y_{t_{j-1}})^2 = \delta b^2(y_{t_{j-1}}) \eta_j^2 + \tilde{U}_j \quad \text{and} \quad \tilde{U}_j = U_j^2 + 2\sqrt{\delta} b(y_{t_{j-1}}) \eta_j U_j.$$

From (A.4) we obtain that, for some positive constant  $C^*$  and for any  $j \geq 1$ ,

$$\sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \left( |\tilde{U}_j| \mid \mathcal{F}_{t_{j-1}} \right) \leq C^* \delta \sqrt{\delta} \left( 1 + y_{t_{j-1}}^2 \right). \quad (\text{A.5})$$

Therefore, on the set  $\Gamma_{*,k}$  we obtain that

$$\hat{b}_k = \frac{\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) b^2(y_{t_{j-1}})}{H_*} + \Upsilon_{1,T} + \Upsilon_{2,T},$$

where

$$\Upsilon_{1,T} = \frac{\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) b^2(y_{t_{j-1}}) \tilde{\eta}_j}{H_*} \quad \text{and} \quad \Upsilon_{2,T} = \frac{\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) \tilde{U}_j}{\delta H_*},$$

and  $\tilde{\eta}_j = \eta_j^2 - 1$ . Note now, that on the set  $\Gamma_{*,k}$

$$0 \leq \sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) - H_* \leq \chi_{\tau_{*,k},k}(h_0) \leq 1$$

and on  $\Gamma_{*,k}^c$

$$\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) \leq H_*.$$

Hence,

$$\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) \leq H_* + 1 \quad \text{a.s.}$$

Therefore, for some constant  $C^* > 0$

$$\begin{aligned} \mathbf{E}_\vartheta |\widehat{b}_k - b^2(z_k)| &\leq C^* \left( h_0 + \frac{1}{H_*} \right) \\ &\quad + \mathbf{E}_\vartheta |\Upsilon_{1,T}| + \mathbf{E}_\vartheta |\Upsilon_{2,T}| + b^2(z_k) \mathbf{P}(\Gamma_{*,k}^c). \end{aligned}$$

Note here, that

$$\mathbf{E}_\vartheta \Upsilon_{1,T}^2 = \frac{1}{H_*^2} \mathbf{E}_\vartheta \sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) b^4(y_{t_{j-1}}) \widetilde{\eta}_j^2 \leq 2\mathbf{b}_{max}^4 \frac{H_* + 1}{H_*^2} \leq \frac{4}{H_*} \mathbf{b}_{max}^4$$

and

$$\begin{aligned} \mathbf{E}_\vartheta |\Upsilon_{2,T}| &\leq \frac{1}{\delta H_*} \sum_{j=1}^{N_0} \mathbf{E}_\vartheta \mathbf{1}_{\{j \leq \tau_{*,k}\}} \chi_{j,k}(h_0) \mathbf{E}_\vartheta \left( |\widetilde{U}_j| \mid \mathcal{F}_{t_{j-1}} \right) \\ &\leq C^* \frac{\sqrt{\delta}}{H_*} \sum_{j=1}^{N_0} \mathbf{E}_\vartheta \mathbf{1}_{\{j \leq \tau_{*,k}\}} \chi_{j,k}(h_0) \left( 1 + y_{t_{j-1}}^2 \right) \\ &\leq C^* \sqrt{\delta}. \end{aligned}$$

Finally, for some constant  $C^* > 0$  and for sufficiently large  $T \rightarrow \infty$ ,

$$\mathbf{E}_\vartheta |\widehat{b}_k - b^2(z_k)| \leq C^* \left( h_0 + \frac{1}{\sqrt{H_*}} + \sqrt{\delta} + \mathbf{P}(\Gamma_{*,k}^c) \right).$$

Hence Proposition 3.1.  $\square$

### A.3 Proof of Proposition 4.4

First of all note that

$$\mathbf{E}_\vartheta |\widehat{\sigma}_l - \sigma_l^2| \leq \frac{1}{v_T \delta (N - N_0) h} \mathbf{E}_\vartheta |\widehat{b}_l - b^2(z_l)|.$$

Taking into account that according to the definition  $N_0$  in (3.3) we obtain through Proposition 3.1 the equality (4.13). Hence Proposition 4.4.  $\square$

## A.4 Technical results

**Lemma A.3.** *Let  $f$  be an absolutely continuous  $[\mathbf{x}_0, \mathbf{x}_1] \rightarrow \mathbb{R}$  function with  $\|\dot{f}\| < \infty$  and  $g$  be  $[\mathbf{x}_0, \mathbf{x}_1] \rightarrow \mathbb{R}$  a step-wise function of the form*

$$g(t) = \sum_{j=1}^n c_j \chi_{(z_{j-1}, z_j]}(t),$$

where  $c_j$  are some constants and the sequence  $(z_j)_{0 \leq j \leq n}$  is given in (3.1). Then for all  $\tilde{\varepsilon} > 0$ , the function  $\Delta = f - g$  satisfies the following inequalities

$$\|\Delta\|^2 \leq (1 + \tilde{\varepsilon}) \|\Delta\|_n^2 + \left(1 + \frac{1}{\tilde{\varepsilon}}\right) \frac{\|\dot{f}\|^2}{n^2} (\mathbf{x}_1 - \mathbf{x}_0)^2,$$

and

$$\|\Delta\|_n^2 \leq (1 + \tilde{\varepsilon}) \|\Delta\|^2 + \left(1 + \frac{1}{\tilde{\varepsilon}}\right) \frac{\|\dot{f}\|^2}{n^2} (\mathbf{x}_1 - \mathbf{x}_0)^2.$$

The proof is given in Lemma A.2 from [31].

## A.5 Proof of the upper bound (8.29)

First, note that

$$\begin{aligned} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) &= \sum_{j=N_0+1}^{\tau_k} \tilde{\varkappa}_{j,k} \chi_{j,k}(h) + \sum_{j=N_0+1}^{\tau_k} \left( \sqrt{\tilde{\varkappa}_{j,k}} - \tilde{\varkappa}_{j,k} \right) \chi_{j,k}(h) \\ &= H_k + (\sqrt{\varkappa_k} - \varkappa_k) \chi_{\tau_k, k}(h) \leq H_k + 1. \end{aligned}$$

So, by the Bunyakovskii - Cauchy - Schwarz inequality,

$$\begin{aligned} \mathbf{E}_\vartheta B_{1,k}^2 &\leq \mathbf{E}_\vartheta \frac{1}{\delta^2 H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \mathbf{E}_\vartheta \frac{1}{H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \varrho_j^2 \\ &\leq \frac{2}{\delta^2} \mathbf{E}_\vartheta \frac{1}{H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \mathbf{E}_\vartheta \left( \varrho_j^2 | \mathcal{F}_{t_{j-1}} \right). \end{aligned}$$

Moreover, taking into account that the function  $S$  from the class (2.1) is lipschitzian with the constant  $L_1 = M \vee L$ , we obtain that

$$\mathbf{E}_\vartheta \left( \varrho_j^2 | \mathcal{F}_{t_{j-1}} \right) \leq \delta L_1^2 \int_{t_{j-1}}^{t_j} \mathbf{E}_\vartheta \left( (y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}} \right) du.$$

Note here that, for  $t_{j-1} \leq u \leq t_j$ ,

$$\begin{aligned} \mathbf{E}_\vartheta((y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}}) &\leq 2\delta \left( \int_{t_{j-1}}^u \mathbf{E}_\vartheta(S^2(y_v) | \mathcal{F}_{t_{j-1}}) dv + b_{max}^2 \right) \\ &\leq 2\delta \left( 2L_1^2 \int_{t_{j-1}}^u (1 + \mathbf{E}_\vartheta(y_v^2 | \mathcal{F}_{t_{j-1}})) dv + b_{max}^2 \right). \end{aligned}$$

Due to Proposition A.6 from [22], one has for  $t_{j-1} \leq v \leq t_j$ ,

$$\mathbf{E}_\vartheta(y_v^2 | \mathcal{F}_{t_{j-1}}) \leq D_* L + y_{t_{j-1}}^2,$$

where  $D_* = (M + L\mathbf{x}_* + 2\mathbf{x}_*)^2(L + M) + b_{max}^2$ . So, for  $t_{j-1} \leq u \leq t_j$  and  $0 < \delta < 1$ ,

$$\mathbf{E}_\vartheta \left( (y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}} \right) \leq 2\delta \left( \mathbf{c}_1 + \mathbf{c}_2 y_{t_{j-1}}^2 + b_{max}^2 \right), \quad (\text{A.6})$$

where  $\mathbf{c}_1 = 2L_1^2(D_*L + 1)$  and  $\mathbf{c}_2 = 2L_1^2$ . Taking into account that

$$\chi_{j,k}(h)y_{t_{j-1}}^2 \leq 2(h^2 + z_k^2),$$

we obtain that, for some constant  $C^* > 0$ ,

$$\sup_{t_{j-1} \leq u \leq t_j} \sup_{\vartheta \in \Theta} \chi_{j,k}(h) \mathbf{E}_\vartheta \left( (y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}} \right) \leq C^* \delta \quad (\text{A.7})$$

and, therefore,

$$\sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta B_{1,k}^2 \leq C^* \delta \sup_{\vartheta \in \Theta} \mathbf{E}_\vartheta \frac{1}{H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) \leq C^* \delta.$$

Hence the upper bound (8.29).  $\square$

## A.6 Proof of the upper bounds (8.30)–(8.31)

Due to orthogonality of stochastic integral increments, we obtain that

$$\begin{aligned} \mathbf{E}_\vartheta \tilde{B}_{1,k}^2 &= \frac{1}{\delta^2} \mathbf{E}_\vartheta \frac{1}{H_k^2} \sum_{j=N_0+1}^{\tau_k} \tilde{\chi}_{j,k} \chi_{j,k}(h) \mathbf{E}_\vartheta \left( (\varpi_j^*)^2 | \mathcal{F}_{t_{j-1}} \right) \\ &\leq \frac{2b_{max}^2}{\delta^2} \mathbf{E}_\vartheta \frac{1}{H_k^2} \sum_{j=N_0+1}^{\tau_k} \tilde{\chi}_{j,k} \chi_{j,k}(h) \int_{t_{j-1}}^{t_j} \mathbf{E}_\vartheta \left( (y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}} \right) du. \end{aligned}$$



By using the bound (A.7) for the last conditional expectation we get

$$\mathbf{E}_\vartheta \tilde{B}_{1,k}^2 \leq 4C^* b_{\max}^2 \mathbf{E}_\vartheta \frac{1}{H_k}.$$

This inequality and (3.12), (3.4) provide the inequality (8.30). Moreover, one has

$$\begin{aligned} \mathbf{E}_\vartheta \tilde{B}_{2,k}^2 &= \frac{1}{\delta^2} \mathbf{E}_\vartheta \frac{1}{H_k^2} \sum_{j=N_0+1}^{\tau_k} \tilde{\varkappa}_{j,k} \chi_{j,k}(h) \left( b(y_{t_{j-1}}) - b(z_k) \right)^2 (\Delta W_{t_j})^2 \\ &\leq \frac{2b_{\max}^2}{\delta} \mathbf{E}_\vartheta \frac{1}{H_k^2} \sum_{j=N_0+1}^{\tau_k} \tilde{\varkappa}_{j,k} \chi_{j,k}(h) (y_{t_{j-1}} - z_k)^2 \leq \frac{4b_{\max}^2 h^2}{\delta} \mathbf{E}_\vartheta \frac{1}{H_k}. \end{aligned}$$

Hence the upper bound (8.31).  $\square$

## A.7 Proof of Proposition A.4

**Proposition A.4.** *For any  $0 < \varepsilon < 1/2$ ,*

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} P_n(\lambda_0) &\leq \frac{1}{1-2\varepsilon} \mathbf{E}_\vartheta \text{Err}_n(\lambda_0) \mathbf{1}_{\mathbf{G}_*} + \frac{(\mathbf{x}_1 - \mathbf{x}_0) \mathbf{g}_T^*}{\varepsilon(1-2\varepsilon)T} \\ &\quad + \frac{2(\mathbf{x}_1 - \mathbf{x}_0) \sigma_{1,*}}{n\varepsilon} + \frac{2\|S\|_n \sqrt{(\mathbf{x}_1 - \mathbf{x}_0) \sigma_{1,*}}}{\sqrt{n}} \sqrt{\mathbf{P}_\vartheta(\mathbf{G}_*^c)}, \end{aligned}$$

where the term  $\mathbf{g}_T^*$  is given in (4.6).

**Proof.** Note that on the set  $\mathbf{G}_*$

$$\begin{aligned} \text{Err}_n(\lambda) &= \sum_{j=1}^n (\lambda(j) \hat{\theta}_{j,n} - \theta_{j,n})^2 = \sum_{j=1}^n \lambda^2(j) \zeta_{j,n}^2 \\ &\quad - 2 \sum_{j=1}^n (1 - \lambda(j)) \lambda(j) \theta_{j,n} \zeta_{j,n} + \sum_{j=1}^n (1 - \lambda(j))^2 \theta_{j,n}^2. \end{aligned}$$

Taking into account here that

$$\zeta_{j,n}^2 = g_{j,n}^2 + \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \xi_{j,n}^2 + 2 \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} g_{j,n} \xi_{j,n},$$

we obtain

$$\text{Err}_n(\lambda) \geq \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 + 2\sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} I_1 - 2\sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} I_2,$$

where  $I_1 = \sum_{j=1}^n \lambda^2(j) g_{j,n} \xi_{j,n}$  and  $I_2 = \sum_{j=1}^n (1 - \lambda(j)) \lambda(j) \theta_{j,n} \xi_{j,n}$ . Moreover, note that, for any  $0 < \varepsilon < 1$ ,

$$2\sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} I_1 \leq \frac{1}{\varepsilon} \|g\|_n^2 + \frac{\varepsilon(\mathbf{x}_1 - \mathbf{x}_0)}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2.$$

Therefore

$$\text{Err}_n(\lambda_0) \geq \frac{(1 - \varepsilon)(\mathbf{x}_1 - \mathbf{x}_0)}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 - \frac{2\sqrt{\mathbf{x}_1 - \mathbf{x}_0}}{\sqrt{n}} I_2 - \frac{1}{\varepsilon} \|g\|_n^2$$

and

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \text{Err}_n(\lambda_0) &\geq \frac{(1 - \varepsilon)(\mathbf{x}_1 - \mathbf{x}_0)}{n} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 \\ &\quad - \frac{2\sqrt{\mathbf{x}_1 - \mathbf{x}_0}}{\sqrt{n}} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} I_2 - \frac{1}{\varepsilon} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \|g\|_n^2. \end{aligned}$$

Taking into account here the definition of  $\mathbf{B}(\cdot)$  in (7.1) and that  $\mathbf{E}_\vartheta I_2 = 0$  we can rewrite the last inequality as

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \text{Err}_n(\lambda_0) &\geq (1 - \varepsilon) \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} P_n(\lambda) + \frac{(1 - \varepsilon)}{\sqrt{n}} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \mathbf{B}(\lambda^2) \\ &\quad + \frac{2\sqrt{\mathbf{x}_1 - \mathbf{x}_0}}{\sqrt{n}} \mathbf{E}_\vartheta \mathbf{1}_{(\mathbf{G}_*)^c} I_2 - \frac{1}{\varepsilon} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \|g\|_n^2. \end{aligned}$$

Now Propositions 7.1 – 7.2 imply that

$$\begin{aligned} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \text{Err}_n(\lambda_0) &\geq (1 - 2\varepsilon) \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} P_n(\lambda) - \frac{2(\mathbf{x}_1 - \mathbf{x}_0) \sigma_{1,*}}{n\varepsilon} \\ &\quad - \frac{2\|S\|_n \sqrt{(\mathbf{x}_1 - \mathbf{x}_0) \sigma_{1,*}}}{\sqrt{n}} \sqrt{\mathbf{P}_\vartheta(\mathbf{G}^c)} - \frac{1}{\varepsilon} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \|g\|_n^2. \end{aligned}$$

Hence Proposition A.4.  $\square$

## References

- [1] Banon, G. (1978) Nonparametric identification for diffusion processes. *SIAM J. Control and Optimization* **16**, 380-390.
- [2] Beltaief, S., Chernoyarov, O. and Pergamenchtchikov, S. (2019) Model selection for the robust efficient signal processing observed with small Levy noise. - *Ann Inst Stat Math.*, <https://doi.org/10.1007/s10463-019-00726-2>
- [3] Candés E, Tao T (2007) The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$ . - *Ann. Statist*, **35** (6), p. 2313 – 2351.
- [4] Chernoyarov, O.V., Breznan, M. and Terekhov, A. V. (2013) Restoration of deterministic and interference distorted signals and images with use of the generalized spectra based on orthogonal polynomials and functions. - *Communications*, **2A**, 71 – 77.
- [5] Dalalyan, A.S. (2005) Sharp Adaptive Estimation of the Trend Coefficient for Ergodic Diffusion. - *Annals of Statistics*, **33** (6), 2507 - 2528.
- [6] Dalalyan, A.S. and Kutoyants, Yu.A. (2002) Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.* **11** (4), p. 402-427.
- [7] Debnath, L and Shah, F. (2015) Wavelet Transforms and Their Applications. - Basel: Birkhäuser.
- [8] Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discret observations. *J. Appl. Probab.* **30**(4), p. 790–804.
- [9] Fujimori, K. (2019) The Danzing selector for a linear model of diffusion processes. - *Statistical Inference for Stochastic Processes*, **22**, p. 475 – 498.
- [10] De Gregorio, A., Iacus, SM (2012) Adaptive LASSO-type estimation for multivariate diffusion processes. *Econom. Theory*, **28** (4), p. 838 – 860.
- [11] Galtchuk, L.I. (1978) Existence and uniqueness of a solution for stochastic equations with respect to semimartingales. *Theory of Probability Appl.*, **23**, 751-763.

- [12] Galtchouk, L. and Pergamenshchikov, S. (2001) Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Mathematical Methods of Statistics*, **10**, 3, p. 316-330.
- [13] Galtchouk, L. and Pergamenshchikov, S. (2004) Nonparametric sequential estimation of the drift in diffusion via model selection. *Mathematical Methods of Statistics*, **13**, 25-49.
- [14] Galtchouk, L. and Pergamenshchikov, S. (2005) Nonparametric sequential minimax estimation of the drift coefficient in diffusion processes. *Sequential Analysis*, **24**(3), 303-330.
- [15] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **9**, 1-16.
- [16] Galtchouk, L. and Pergamenshchikov, S. (2007) Uniform concentration inequality for ergodic diffusion processes. *Stochastic Processes and Applications*, **117**, 830-839.
- [17] Galtchouk, L. and Pergamenshchikov, S. (2009a) Sharp non-asymptotic Oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21**, 1-16.
- [18] Galtchouk, L. and Pergamenshchikov, S. (2009) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *Journal of the Korean Statistical Society*, **38**(4), 305-322.
- [19] Galtchouk, L. and Pergamenshchikov, S. (2011) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *Journal of Nonparametric Statistics*, **23** (2), 255-285.
- [20] Galtchouk, L. and Pergamenshchikov, S. (2013) Uniform concentration inequality for ergodic diffusion processes observed at discrete times. *Stochastic Processes and Applications*, **123**(1), 91-109.
- [21] Galtchouk, L. and Pergamenshchikov, S. (2014) Geometric ergodicity for classes of homogeneous Markov chains. - *Stochastic Processes and their Applications*, 2014, v. 124, 10, p. 3362-3391
- [22] Galtchouk, L. and Pergamenshchikov, S. (2015) Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions. *Bernoulli*, **21**(4), 2569 – 2594.

- [23] Gihkman, I.I. and Skorohod, A.V. (1968) Stochastic differential equations. *Naukova Dumka*, Kiev.
- [24] Gobet, E., Hoffmann, M. and Reiss, M. (2004) Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, **32**(5), p. 2223-2253.
- [25] Genon-Catalot, V. and Laredo, C. and Picard, D. (1992) Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.* **19** 317–335.
- [26] Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, p. 135 - 163.
- [27] Jacod, J. (2000) Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* **27**(1) 83–96.
- [28] Kabanov, Yu. M. and Pergamenshchikov, S. M. *Two Scale Stochastic Systems: Asymptotic Analysis and Control*. - Applications of Mathematics, Stochastic Modelling and Applied Probability, 49, Springer-Verlag, Berlin, New York, 2003.
- [29] Karatzas, I. and Shreve, S.E. (1998) *Methods of Mathematical Finance*. Springer, New York.
- [30] Konev, V.V. and S.M. Pergamenshchikov. Efficient robust nonparametric estimation in a semimartingale regression model. *Ann. Inst. Henri Poincaré Probab. Stat.*, **48** (4), 2012, 1217–1244.
- [31] Konev, V.V. and Pergamenshchikov, S.M. Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic processes and their applications*, **125**, 2015, 294 – 326.
- [32] Kutoyants, Yu.A. (1984a) *Parameter Estimation for Stochastic Processes*. Heldeman-Verlag, Berlin.
- [33] Kutoyants, Yu.A. (1984b) On nonparametric estimation of trend coefficients in a diffusion process. *In Statistics and Control of Stochastic Processes*. Steklov Seminar Proceedings, Moscow, p. 230-250.
- [34] Kutoyants, Yu.A. (2003) *Statistical Inferences for Ergodic Diffusion Processes*. Springer, Berlin.
- [35] Lamberton, D. and Lapeyre, B. (1996) *Introduction to stochastic calculus applied to finance*. Chapman & Hall, London.

- [36] Liptser, R.Sh. and Shiryaev, A.N. (1978) *Statistics of a random process*. Springer-Verlag, Berlin, Heidelberg, New York, v.I, II.
- [37] Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso.  
- *J. R. Statist. Soc.*, **Ser B 58** (1), p. 267 – 288.