Statistical Inference for Stochastic Processes (2022) 25:127–158
https://doi.org/10.1007/s11203-021-09241-9

# Adaptive efficient analysis for big data ergodic diffusion models

**Leonid I. Galtchouk[1] · Serge M. Pergamenshchikov[2,3]**

Received: 25 May 2020 / Accepted: 10 March 2021 / Published online: 27 March 2021


**Abstract**

We consider drift estimation problems for high dimension ergodic diffusion processes in nonparametric setting based on observations at discrete fixed time moments in the case when diffusion coefficients are unknown. To this end on the basis of sequential analysis methods we develop model selection procedures, for which we show non asymptotic sharp oracle inequalities. Through the obtained inequalities we show that the constructed model selection procedures are asymptotically efficient in adaptive setting, i.e. in the case when the model regularity is unknown. For the first time for such problem, we found in the explicit form the celebrated Pinsker constant which provides the sharp lower bound for the minimax squared accuracy normalized with the optimal convergence rate. Then we show that the asymptotic quadratic risk for the model selection procedure asymptotically coincides with the obtained lower bound, i.e this means that the constructed procedure is efficient. Finally, on the basis of the constructed model selection procedures in the framework of the big data models we provide the efficient estimation without using the parameter dimension or any sparse conditions.

**Keywords** Adaptive nonparametric drift estimation · Asymptotic efficiency · Discrete time data · Nonasymptotic estimation · Model selection · Quadratic risk · Sharp oracle inequality

**Mathematics Subject Classification** Primary 62G08; Secondary 62G05 · 62G20

This research was supported by RSF, Grant No. 20-61-47043, National Russian Research Tomsk State University, Russia.

✉ Leonid I. Galtchouk
leonid.galtchouk0667@orange.fr

Serge M. Pergamenshchikov
Serge.Pergamenchtchikov@univ-rouen.fr

1 International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, 36, Lenin str., Tomsk, Russia 634041

2 Laboratoire de Mathématiques Raphael Salem, Université de Rouen, Avenue de l'Université, BP 12, 76801 Saint Etienne du Rouvray Cedex, France

3 International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, 36 Lenina prosp., Tomsk, Russia 634050

# 1 Introduction

## 1.1 Problem

In this paper we consider the high dimensional diffusion model introduced in Fujimori (2019), i.e. we study the diffusion process defined as

$$\mathrm{d}y_t = \left( \psi_0(y_t) + \sum_{j=1}^{q} \beta_j \psi_j(y_t) \right) \mathrm{d}t + b(y_t)\,\mathrm{d}W_t, \quad 0 \le t \le T, \tag{1.1}$$

where $(\psi_j)_{0\le j\le q}$ are known linearly independent functions, $(W_t)_{t\ge 0}$ is a standard Wiener process, $(\beta_j)_{1\le j\le q}$ are unknown parameters and $b(\cdot)$ is unknown diffusion coefficient. It is assumed that observations are accessible only at the discrete time moments

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

where the frequency $\delta = \delta_T \in (0, 1)$ and the sample size $N = N(T)$ are some functions of $T$ that will be specified later and such that $\delta_T \to 0$, $N(T) \to +\infty$ as $T \to \infty$. We study the model (1.1) in big data setting (see, for example, Fujimori 2019; De Gregorio and Iacus 2012; Galtchouk and Pergamenshchikov 2019), i.e. in the case when the parameter dimension is greater than the number of observations, i.e. $q > N$. We remind, that for such model usually one uses the LASSO algorithm or Dantzig selector (see, for example, Hastie et al. 2009). But these methods can not be used if the the parameter dimension $q$ is unknown or equals to $+\infty$. By this reason in this paper, similarly to Galtchouk and Pergamenshchikov (2019), we study the model (1.1) in nonparametric setting, i.e.

$$\mathrm{d}y_t = S(y_t)\,\mathrm{d}t + b(y_t)\,\mathrm{d}W_t, \quad 0 \le t \le T. \tag{1.3}$$

The problem is to estimate the function $S(\cdot)$ on the basis of the observations (1.2). Indeed, such problems are important at various applications such that signal processing (Kutoyants 1977, 1984a, b; Bayisa et al. 2019), stochastic optimal control (Kabanov and Pergamenshchikov 2003), finance (Lamberton and Lapeyre 1996; Karatzas and Shreve 1998b) and etc. We consider the quadratic risk defined, for any estimator $\widehat{S}$, as

$$\mathcal{R}_\vartheta(\widehat{S}) = \mathbf{E}_\vartheta\, \|\widehat{S} - S\|^2 \quad \text{and} \quad \|f\|^2 = \int_{\mathbf{x}_0}^{\mathbf{x}_1} |f(x)|^2 \mathrm{d}x, \tag{1.4}$$

where $\mathbf{E}_\vartheta$ is the expectation with respect to the distribution of the process (1.3) for the functions $\vartheta = \vartheta(\cdot) = (S(\cdot), b(\cdot))$ and $\mathbf{x}_0 < \mathbf{x}_1$ are some fixed points.

## 1.2 Motivations

Nonparametric estimation problems for $S$ were studied in a number of papers in the case of complete observations, that is when the whole trajectory $(y_t)_{0\le t\le T}$ was observed. A sufficiently complete survey one can find, for example, in Kutoyants (2003). It should be noted that, for the first time, the famous Pinsker constant representing the asymptotic efficiency property for nonparametric diffusion models was found by Dalalyan and Kutoyants (2002); Dalalyan (2005) for a special weighted integral risk using very nice local time tool. In non asymptotic setting Galtchouk and Pergamenshchikov (2001, 2004, 2005, 2006) developed nonparametric sequential estimation methods for the models (1.3) on the basis of which in Galtchouk and Pergamenshchikov (2011) they calculated the Pinsker constant for the risk

(1.4). It should be noted that in all the cited papers the estimation problems were studied in the case of complete observations. In practice, usually for the models (1.3), the observations are accessible only at the discrete time moments (1.2). A natural question arises about properties of estimators based on discrete observations. In such setting estimation problems for models of the form (1.3) were considered firstly for estimating the unknown diffusion coefficient $b(\cdot)$ on a fixed time interval, when the observation frequency goes to zero, (see, for example, Florens-Zmirou 1993; Jacod 2000 and the references therein). Later, in Gobet et al. (2004) kernel estimates for the drift and diffusion coefficients were studied for the reflected processes (1.3) with the values in the interval [0, 1]. Minimax optimal convergence rates are found as the sample size goes to infinity. As to the ergodic case, it should be noted that firstly sequential procedures were proposed in Hoffmann (1999) for nonparametric drift estimation problems of the process (1.3) in an integral metric. Some upper and lower asymptotic bounds were found for the $\mathbf{L}_p$ - risks. Later, in the paper Comte et al. (2009) a non-asymptotic oracle inequality was obtained for a special empiric quadratic risk defined as a function of the observations at the discrete time moments. In the asymptotic setting, when the observation frequency goes to zero and the length of the observation time interval tends to infinity, the constructed estimators reach the minimax optimal convergence rates. Unfortunately, in all these papers, the efficiency property is not studied for estimation procedures on incomplete observations.

### 1.3 Key ideas

Our approach is based on the sequential analysis methods developed in the papers Galtchouk and Pergamenshchikov (2001, 2005, 2006) for nonparametric estimation problems. This approach makes possible to replace the random denominator by a conditional constant in a sequential Nadaraya–Watson estimator. Let us recall that in the case of complete observations the sequential estimator efficiency was proved by making use of a uniform concentration inequality (see Galtchouk and Pergamenshchikov 2007), besides an indicator kernel estimator. As it turns out later in Galtchouk and Pergamenshchikov (2011), the efficient kernel estimate in the above given sense provides constructing a selection model adaptive procedure that appears efficient in the quadratic metric. Therefore, in order to realize this program (i.e. from efficient pointwise estimators to an efficient $\mathbf{L}_2$ - estimator) in the case of discrete time observations, one needs to use suitable concentration inequalities. Such concentration inequalities are obtained in Galtchouk and Pergamenshchikov (2013) through the uniform geometric ergodicity method for the process (1.3) developed in Galtchouk and Pergamenshchikov (2014) which provides uniformly over functions $S(\cdot)$ and $b(\cdot)$ non asymptotic uppers bounds for the convergence rate in the ergodic theorem. Using this tool we can show that the corresponding weighted least square estimator for $S$ in (1.3) setting on the regularity parameters is efficient for the risk (1.4) in the Pinsker sense (Pinsker 1981). Finally, using sharp oracle inequalities, we can estimate from above the risk of the model selection procedure with the risk of effective estimation and obtain the efficiency property in the adaptive sense, i.e. without using regularity properties of the function $S$.

### 1.4 Organization of the paper

The paper is organized as follows. In Sect. 2 we represent the truncated sequential point wise method and we announce main conditions. In Sect. 3 through the above sequential estimators we pass to a nonparametric regression model. In Sect. 4 we construct the model selection procedure. The main results are collected in Sect. 5. In Sect. 6 we study main properties of the basic regression model needed to obtain oracle inequalities. In Sect. 7 we find a sharp upper bound for the asymptotic risks, i.e. we calculate the Pinsker constant. In Sect. 8 we prove all main results. In Conclusion we summarize all main contributions of this paper. In "Appendix" we postpone all necessary technical results.

## 2 Truncated sequential estimation method

First of all we describe the sequential method for the model (1.3). Note that in the complete observations case the kernel estimation has the following form

$$\widehat{S}_T(z) = \frac{\int_0^T Q\left(\frac{y_t - z}{h}\right) \mathrm{d}y_t}{\int_0^T Q\left(\frac{y_t - z}{h}\right) \mathrm{d}t}, \tag{2.1}$$

where $Q$ is a kernel, i.e. a function such that $Q(x) = 0$ for $|x| > 1$, and $h$, $h > 0$, is a bandwidth. As we see in this case the estimator is non linear function of the observations $(y_t)_{0 \le t \le T}$ and, therefore, it cannot be studied in non asymptotic setting, i.e. for a fixed finite $T$. By these reasons Galtchouk and Pergamenshchikov (2004) proposed some sequential version for this estimator in which the main idea is to transform the random denominator into a non random constant $H > 0$, i.e.

$$\widehat{S}_{\tau_H}(z) = \frac{\int_0^{\tau_H} Q\left(\frac{y_t - z}{h}\right) \mathrm{d}y_t}{H}, \tag{2.2}$$

where the observations duration is defined by the following stopping time

$$\tau_H(z) = \inf\left\{ t \ge 0 : \int_0^t Q\left(\frac{y_u - z}{h}\right) \mathrm{d}u \ge H \right\}. \tag{2.3}$$

It is clear, that to apply this method the stopping time (2.3) must be finite almost surely. To do this, one needs to assume some conditions under which the process (1.3) returns to any vicinity of the point $z \in [\mathbf{x}_0, \mathbf{x}_1]$ infinite times. A natural condition which provides such properties is the ergodicity. Moreover, in order to develop minimax estimation methods, we need a uniform ergodicity property with respect to $S$ over some functional class. To do this we use the functional class introduced in Galtchouk and Pergamenshchikov (2015), i.e. for some fixed $\mathbf{L} \ge 1$, $\mathbf{M} > 0$ and $\mathbf{x}_* > |\mathbf{x}_0| + |\mathbf{x}_1|$ we set

$$\Sigma_{\mathbf{L},\mathbf{M}} = \Bigg\{ S \in \mathbf{C}^1(\mathbb{R}) : \sup_{|x| \le \mathbf{x}_*} \left( |S(x)| + |\dot{S}(x)| \right) \le \mathbf{M}, \\ -\mathbf{L} \le \inf_{|x| \ge \mathbf{x}_*} \dot{S}(x) \le \sup_{|x| \ge \mathbf{x}_*} \dot{S}(x) \le -1/\mathbf{L} \Bigg\}. \tag{2.4}$$

Here and in the sequel we denote by $\dot{f}$ and $\ddot{f}$ the corresponding derivatives. Note, that (see, for example, Galtchouk 1978) for any $S$ from $\Sigma_{\mathbf{L},\mathbf{M}}$ the Eq. (1.3) with a lipschitz diffusion function $b$ has a unique strong solution and, moreover, it is ergodic with the ergodic density defined as

$$\mathbf{q}_\vartheta(x) = \left( \int_{\mathbb{R}} b^{-2}(z)\, e^{\widetilde{S}(z)} \mathrm{d}z \right)^{-1} b^{-2}(x)\, e^{\widetilde{S}(x)}, \tag{2.5}$$

where $\widetilde{S}(x) = 2\int_0^x b^{-2}(v)S(v)\mathrm{d}v$ and $\vartheta = (S, b)$ (see, for example, Gihman and Skorohod 1968, Ch.4, 18, Th2). On the basis of this density in Galtchouk and Pergamenshchikov (2006) for the estimation problems of the functions $S$ from the class (2.4) it was proposed the truncated sequential procedure, in which the stopping time (2.3) is replaced by the $\tau_H \wedge T$. Obviously, that to obtain an asymptotic efficient estimation one needs to use all observations on the time interval $[0, T]$, i.e. asymptotically $\tau_H \approx T$ as $T \to \infty$. Moreover, as it is shown in Galtchouk and Pergamenshchikov (2011) to obtain the efficiency with respect to the quadratic risk (1.4) one needs to choose the kernel $Q$ as the indicator function, i.e.

$$Q(y) = \mathbf{1}_{\{|y|\le 1\}}. \tag{2.6}$$

Using the uniform geometric ergodicity property developed in Galtchouk and Pergamenshchikov (2014) one can show that asymptotically, as $H \to \infty$ and $h \to 0$,

$$\tau_H(z) \approx \frac{H}{h\mathbf{q}_\vartheta(z)\int_{-1}^{1} Q(u)\mathrm{d}u} = \frac{H}{2h\mathbf{q}_\vartheta(z)}, \tag{2.7}$$

i.e. to obtain the asymptotic coincidence of the rule (2.3) with $T$ one needs to choose the threshold $H$ as

$$H \approx 2Th\mathbf{q}_\vartheta(z). \tag{2.8}$$

It is clear, that this is impossible, because the ergodic density $\mathbf{q}_\vartheta$ is unknown. Therefore, firstly on the basis of observations $(y_u)_{0\le u\le T_0}$ with $T_0 < T$ we need to estimate the density $\mathbf{q}_\vartheta$ and then we use this estimator in (2.8) to choose the threshold $H$ and, finally, we estimate the function $S(z)$ on the basis of the observations $(y_u)_{T_0\le u\le T}$. In the case of complete observations, this program was realized in Galtchouk and Pergamenshchikov (2011). In the discrete data case, using the method developed in Galtchouk and Pergamenshchikov (2015) we transform the sequential procedure from Galtchouk and Pergamenshchikov (2011) by the following way. First, to estimate the ergodic density we will use the first $N_0$ observations defined as

$$N_0 = [N^\gamma(T)] \quad \text{and} \quad 5/6 < \gamma < 1, \tag{2.9}$$

where $[x]$ is the integer part of $x$. Later, in Remark 4.1 we will explain this choice.
To estimate the density $\mathbf{q}_\vartheta$ we will use the following kernel estimator

$$\widehat{q}(z) = \frac{1}{2\,N_0\,h_0}\sum_{j=1}^{N_0} \chi_j(z, h_0), \tag{2.10}$$

where $h_0 = h_0(T) = T_0^{-1/2}$, $T_0 = \delta N_0$ and

$$\chi_j(z, h) = Q\left(\frac{y_{t_{j-1}} - z}{h}\right).$$

We recall, that $Q$ is the indicator defined in (2.6). Furthermore, note that to study the stopping time we need to divide the threshold $H$ in (2.7) by the estimator of the density (2.10) which generally speaking may be very small. To avoid this situation we modify the estimator (2.10) as

$$\widetilde{q}(z) = \begin{cases} (\upsilon_T)^{1/2}, & \text{if } \widehat{q}(z) < (\upsilon_T)^{1/2};\\ \widehat{q}(z), & \text{if } (\upsilon_T)^{1/2} \le \widehat{q}(z) \le (\upsilon_T)^{-1/2};\\ (\upsilon_T)^{-1/2}, & \text{if } \widehat{q}(z) > (\upsilon_T)^{-1/2}, \end{cases} \tag{2.11}$$

where $\upsilon_T$ is a positive function of $T$ going to zero as $T \to \infty$.
Then to estimate the function $S(z)$ we use the observations $(y_{t_j})_{N_0+1\le j\le N}$. To use all this observations in the sequential procedure we will choose the threshold $H$ such that the observations duration in the sequential procedure will be asymptotically less than $N - N_0$, i.e. $\tau < N - N_0$ and, moreover, $\tau/(N - N_0) \to 1$ as $T \to \infty$. To do this we set

$$H(z) = h(N - N_0)(2\widetilde{q}(z) - \upsilon_T) \quad \text{and} \quad h = T^{-1/2}. \tag{2.12}$$

Note, that this threshold is $\mathcal{F}_{t_{N_0}}$ - measurable, where $\mathcal{F}_t = \sigma\{y_t,\ 0 \le u \le t\}$. Therefore, we can use it to define the following stopping time

$$\tau(z) = \inf\left\{ l \ge N_0 + 1 : \sum_{j=N_0+1}^{l} \widetilde{\chi}_j(z, h) \ \ge H(z) \right\}, \tag{2.13}$$

where $\widetilde{\chi}_j(z, h) = \chi_j(z, h)\mathbf{1}_{\{j<N\}} + H(z)\mathbf{1}_{\{j\ge N\}}$. Note, that this stopping time $\tau(z) \le N$ a.s. Now we need to define the correction coefficient $0 < \varkappa(z) \le 1$ as

$$\sum_{j=N_0+1}^{\tau(z)-1} \widetilde{\chi}_j(z, h) + \varkappa(z)\widetilde{\chi}_{\tau(z)}(z, h) = H(z). \tag{2.14}$$

Finally, we define the sequential estimator for $S(z)$ as

$$S^*(z) = \frac{1}{\delta H(z)} \left( \sum_{j=N_0+1}^{\tau(z)} \sqrt{\widetilde{\varkappa}_j(z)}\widetilde{\chi}_j(z, h)\,(y_{t_j} - y_{t_{j-1}}) \right) \mathbf{1}_{\Gamma(z)}, \tag{2.15}$$

where $\Gamma(z) = \{\tau(z) < N\}$ and $\widetilde{\varkappa}_j(z) = \mathbf{1}_{\{j<\tau(z)\}} + \sqrt{\varkappa(z)}\mathbf{1}_{\{j=\tau(z)\}}$. Using here the model (1.3) we can represent this estimator on $\Gamma(z)$ as

$$S^*(z) = S(z) + \mathbf{g}_1(z) + \mathbf{g}_2(z) + \frac{b(z)}{\sqrt{\delta H(z)}}\xi(z), \tag{2.16}$$

where

$$\mathbf{g}_1(z) = \frac{1}{\delta H(z)} \sum_{j=N_0+1}^{\tau(z)} \sqrt{\widetilde{\varkappa}_j(z)}\widetilde{\chi}_j(z, h) \int_{t_{j-1}}^{t_j} S(y_u)\,\mathrm{d}u - S(z),$$

$$\mathbf{g}_2(z) = \frac{1}{\delta H(z)} \sum_{j=N_0+1}^{\tau(z)} \sqrt{\widetilde{\varkappa}_j(z)}\widetilde{\chi}_j(z, h) \int_{t_{j-1}}^{t_j} (b(y_s) - b(z))\mathrm{d}W_s$$

and

$$\xi(z) = \frac{1}{\sqrt{\delta H(z)}} \sum_{j=N_0+1}^{\tau(z)} \sqrt{\widetilde{\varkappa}_j(z)}\widetilde{\chi}_j(z, h)\,(W_{t_j} - W_{t_{j-1}}). \tag{2.17}$$

To construct the model selection procedure we need to estimate the diffusion coefficient $b^2(z)$ for $\mathbf{x}_0 \le z \le \mathbf{x}_1$. For this we will use the following truncated sequential procedure. First, we define the corresponding stopping time as

$$\mathbf{t}_0(z) = \inf\left\{ j \ge 1 : \sum_{l=1}^{j} \chi_l(z, h_0) \ge H_0 \right\} \wedge N_0, \tag{2.18}$$

where $H_0 = h_0 N_0 / \ln(T+1)$. Then we set

$$\widehat{b}(z) = \frac{\sum_{j=1}^{\mathbf{t}_0(z)} \chi_j(z, h_0)(y_{t_j} - y_{t_{j-1}})^2}{\delta\, H_0} \mathbf{1}_{\Gamma_0(z)}, \tag{2.19}$$

where $\Gamma_0(z) = \{\mathbf{t}_0 < N_0\}$. By the similar reasons as in (2.12) we choose the threshold $H_0$ so that the stopping time (2.18) will be less that $N$ asymptotically as $T \to \infty$. If $b(z)$ is known, we set $\widehat{b}(z) = b^2(z)$.

Note that we consider the efficient estimation problem only for the drift function $S$, i.e. the diffusion coefficient $b$ is considered as a nuisance parameter for which we assume that it is two time continuously differentiable such that

$$0 < \mathbf{b}_{min} \leq \inf_{x \in \mathbb{R}} |b(x)| \leq \sup_{x \in \mathbb{R}} \max\left(|b(x)|, |\dot{b}(x)|, |\ddot{b}(x)|\right) \leq \mathbf{b}_{max}, \tag{2.20}$$

where $\mathbf{b}_{min}$ and $\mathbf{b}_{max}$ are some fixed constants. Denoting by $\mathcal{B}$ the class of such functions, we set

$$\Theta = \Sigma_{\mathbf{L},\mathbf{M}} \times \mathcal{B} = \left\{(S, b) : S \in \Sigma_{\mathbf{L},\mathbf{M}} \quad \text{and} \quad b \in \mathcal{B}\right\}. \tag{2.21}$$

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

$$s^* = \sup_{\mathbf{x}_0 \leq x \leq \mathbf{x}_1} \sup_{S \in \Sigma_{\mathbf{L},\mathbf{M}}} S^2(x) < \infty. \tag{2.22}$$

Moreover, it should be noted also, that

$$0 < \mathbf{q}_* = \inf_{|x| \leq \mathbf{x}_*} \inf_{\vartheta \in \Theta} \mathbf{q}_\vartheta(x) \leq \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \mathbf{q}_\vartheta(x) = \mathbf{q}^* < +\infty. \tag{2.23}$$

To use the concentration inequalities from Galtchouk and Pergamenshchikov (2013) we need the following conditions.

$\mathbf{A}_1$) *The frequency $\delta$ in the observations (1.2) has the following form*

$$\delta = \delta_T = \frac{1}{(T+1) l_T}, \tag{2.24}$$

*where the function $l_T$ is such that,*

$$\lim_{T \to \infty} \frac{l_T}{\ln T} = +\infty. \tag{2.25}$$

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

$\mathbf{A}_2$) *Assume, that*

$$\lim_{T \to \infty} \left(\upsilon_T + \frac{\ln T}{T(\upsilon_T)^2} + \frac{\ln T}{l_T\,(\upsilon_T)^5}\right) = 0. \tag{2.26}$$

For example, one can take $\upsilon_T = \ln^{-\mathbf{a}}(T+1)$ and $l_T \geq \ln^{1+6\mathbf{a}} T$, for some $\mathbf{a} > 0$.

***Remark 2.1*** It should be noted the choice of the bandwidth $h$ in (2.12) is due to the following reasons. According to the method developed in Galtchouk and Pergamenshchikov (2011) to provide an efficient sequential estimation one needs to choose the bandwidth $h$ as small as possible, but to use the concentration inequalities from Galtchouk and Pergamenshchikov (2013) the bandwidth must be greater than $T^{-1/2}$. Therefore, there is only one way $h = T^{-1/2}$.

***Remark 2.2*** It should be noted that the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$) provide an efficient point wise estimation developed in Galtchouk and Pergamenshchikov (2015) for $N(T) \approx T^2 l_T$ as $T \to \infty$. In this paper, similarly to the complete observations case considered in Galtchouk and Pergamenshchikov (2011), to construct model selection procedures on the observations (1.2) we will use efficient point wise sequential estimators (2.15).

## 3 Regression model

To obtain an efficient estimator of the function $S$ on the interval $[\mathbf{x}_0, \mathbf{x}_1]$, similarly to Galtchouk and Pergamenshchikov (2011), we will use the point wise sequential estimators (2.15) at the points $(z_k)_{0\le k\le n}$ defined as

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

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

$$\frac{\check{\mathbf{x}}}{n} \ge 2h \quad \text{and} \quad \lim_{T\to\infty} \frac{n(T)}{\sqrt{T}} = 1. \tag{3.2}$$

For example, we can take, $n(T) = 2[\sqrt{T}\check{\mathbf{x}}/4] - 1$.

To develop model selection methods we shall pass to a regression model by the same way as in Galtchouk and Pergamenshchikov (2011), i.e we set

$$\mathbf{G}_* = \cap_{k=1}^n \Gamma(z_k) \quad \text{and} \quad Y_k = S^*(z_k)\,\mathbf{1}_{\mathbf{G}_*}, \tag{3.3}$$

where the set $\Gamma(z)$ and the estimators $S^*(z)$ are defined in (2.12) and (2.15). Using the form (2.16) we obtain on the set $\mathbf{G}_*$ the following regression model

$$Y_k = S(z_k) + \mathbf{g}_k + \sigma_k \xi_k, \quad \sigma_k = b(z_k)/\sqrt{\delta H_k}, \tag{3.4}$$

where $\mathbf{g}_k = \mathbf{g}_1(z_k) + \mathbf{g}_2(z_k)$, $\xi_k = \xi(z_k)$ and $H_k = H(z_k)$ is defined in (2.12).
First note that, Proposition 4.5 from Galtchouk and Pergamenshchikov (2015) directly implies the following property.

**Proposition 3.1** *For any $a > 0$, under the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$)*

$$\lim_{T\to\infty} T^a \sup_{\vartheta\in\Theta} \mathbf{P}_\vartheta(\mathbf{G}_*^c) = 0. \tag{3.5}$$

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

**Proposition 3.2** *The random variables $(\xi_k)_{1\le k\le n}$ are $\mathcal{N}(0,1)$ i.i.d. conditionally to $\mathcal{G}_0 = \mathcal{F}_{t_{N_0}}$, where $\mathcal{F}_t = \sigma\{y_u,\ 0 \le u \le t\}$.*

***Proof*** Note, that $t_{\tau_k}$ is a stopping time for the filtration $(\mathcal{F}_u)_{0\le u\le T}$. Therefore, $\xi_k$ can be represented as

$$\xi_k = \int_{t_{N_0}}^{t_{\tau_k}} \Psi_k(u)\mathrm{d}W_u, \quad \Psi_k(u) = \sum_{j=N_0+1}^{N} \psi_{j,k}\,\mathbf{1}_{\{t_{j-1}\le u<t_j\}},$$

where $\psi_{j,k} = \sqrt{\widetilde{\varkappa}_j(z_k)}\widetilde{\chi}_j(z_k, h)/\sqrt{\delta H_k}$. Note here, that $\int_{t_{N_0}}^{t_{\tau_k}} \Psi_k^2(u)\mathrm{d}u = 1$ and, moreover, the first condition in (3.2) yields, that for $k \neq l$

$$\mathbf{E}_\vartheta \left(\xi_k\xi_l|\mathcal{G}_0\right) = \mathbf{E}_\vartheta \left(\int_{t_{N_0}}^{t_{\tau_k}} \Psi_k(u)\Psi_l(u)\mathrm{d}u|\mathcal{G}_0\right) = 0.$$

Thus, the time-changed Brownian motion property (see, for example, Karatzas and Shreve 1998a, p. 174) implies this Proposition. □

Now we set

$$\mathbf{g}_T^* = T \max_{1\le k\le n} \sup_{\vartheta\in\Theta} \mathbf{E}_\vartheta\, \mathbf{g}_k^2 \mathbf{1}_{\mathbf{G}_*}. \tag{3.6}$$

**Proposition 3.3** *For any $a > 0$, under the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$)*

$$\lim_{T\to\infty} T^{-a}\, \mathbf{g}_T^* = 0. \tag{3.7}$$

We estimate the parameter $\sigma_l^2$ as

$$\widehat{\sigma}_l = \widehat{b}_l/(\delta H_l), \tag{3.8}$$

where $\widehat{b}_l = \widehat{b}(z_l)$ is defined in (2.19). Note that the coefficients $(\sigma_l)_{1\le l\le n}$ are random variables such that

$$\sigma_{0,*} \le \min_{1\le l\le n} \sigma_l^2 \le \max_{1\le l\le n} \sigma_l^2 \le \sigma_{1,*}, \tag{3.9}$$

where

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

Now, we need to study the properties of the estimator (3.8). To this end we set

$$\varpi_T^* = n \max_{1\le l\le n} \mathbf{E}_\vartheta\, |\widehat{\sigma}_l - \sigma_l^2|. \tag{3.10}$$

**Proposition 3.4** *If the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$) hold, then for any $a > 0$,*

$$\lim_{T\to\infty} T^{\gamma-1/2-a}\, \varpi_T^* = 0. \tag{3.11}$$

Propositions 3.3 - 3.4 are shown in “Appendix A.2”.

***Remark 3.1*** It should be noted that the obtained regression model (3.4) differs from the models considered before in the papers Galtchouk and Pergamenshchikov (2009a, b, 2011). More precisely, in Galtchouk and Pergamenshchikov (2009a, b) it is studied the model (3.4) with the unknown variances $\sigma_k$, but without drift coefficients, i.e. $\mathbf{g}_k = 0$, in Galtchouk and Pergamenshchikov (2011) the coefficients $\sigma_l$ are known, but the there are non zero drift coefficients $\mathbf{g}_l$. In this paper we need to study the more general regression model (3.4) in which the variances are unknown and there are non zero drift coefficients $\mathbf{g}_l$, i.e. we can’t use the methods of those papers.

## 4 Model selection

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

$$(\phi_i, \phi_j)_n = \frac{\check{\mathbf{x}}}{n}\sum_{l=1}^{n} \phi_i(z_l)\phi_j(z_l) = \mathbf{1}_{\{i=j\}}. \tag{4.1}$$

One can take the trigonometric basis defined as $\phi_1(x) \equiv 1/\sqrt{\check{\mathbf{x}}}$ and, for $j \ge 2$,

$$\phi_j(x) = \sqrt{\frac{2}{\check{\mathbf{x}}}}\begin{cases} \cos(2\pi[j/2]\,\mathbf{l}_0(x)) & \text{for even } \quad j; \\ \sin(2\pi[j/2]\mathbf{l}_0(x)) & \text{for odd } \quad j, \end{cases} \tag{4.2}$$

where $\mathbf{l}_0(x) = (x - \mathbf{x}_0)/\check{\mathbf{x}}$. Note that if $n$ is odd, then this basis is orthonormal for the empirical inner product, i.e. satisfies the property (4.1). In the sequel, we will denote by $\|\cdot\|_n$ the norm corresponding to the scalar product (4.1). To estimate $S$ we use the discrete Fourier expansion on the sieve (3.1), i.e.

$$S(z_k) = \sum_{j=1}^{n} \theta_{j,n}\phi_j(z_k), \qquad 1 \le k \le n, \tag{4.3}$$

where

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

Moreover, using the regression model (3.4) we estimate these coefficients as

$$\widehat{\theta}_{j,n} = (Y, \phi_j)_n = \frac{\check{\mathbf{x}}}{n}\sum_{l=1}^{n} Y_l\phi_j(z_l). \tag{4.4}$$

By the model (3.4), we obtain on the set $\mathbf{G}_*$

$$\widehat{\theta}_{j,n} = \theta_{j,n} + \zeta_{j,n}, \quad \zeta_{j,n} = \mathbf{g}_{j,n} + \sqrt{\frac{\check{\mathbf{x}}}{n}}\xi_{j,n}, \tag{4.5}$$

where

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

According to the model selection approach proposed in Galtchouk and Pergamenshchikov (2011), we estimate the values $S(z_k)$ by the weighted least squares estimators

$$\widehat{S}_\lambda(z_k) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(z_k), \qquad 1 \le k \le n, \tag{4.6}$$

where the weight vector $\lambda = (\lambda(1), \ldots, \lambda(n))'$ belongs to some finite set $\Lambda$ from $[0, 1]^n$ and $\lambda'$ denotes the transpose of $\lambda$. 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 need the following norm of the set $\Lambda$

$$\Lambda_* = \max_{\lambda\in\Lambda}\sum_{j=1}^{n} \lambda(j), \tag{4.7}$$

which can be a function of $T$, i.e. $\Lambda_* = \Lambda_*(T)$.

We need the following condition.

$\mathbf{A}_3$) *For any* $a > 0$,

$$\lim_{T\to\infty} \frac{\nu_T}{T^a} = 0 \quad \text{and} \quad \lim_{T\to\infty} \frac{\Lambda_*(T)}{T^{1/3+a}} = 0. \tag{4.8}$$

***Remark 4.1*** Note that, the property (3.11) and the condition for $\gamma$ in (2.9) imply that under the condition $\mathbf{A}_3$) the term $\varpi_T^* \Lambda_* \to 0$ as $T \to \infty$. This is one of the basic properties used in the proof of Theorem 5.1.

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

$$\widehat{S}_\lambda(x) = \sum_{l=1}^{n} \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < x \le z_l\}}, \quad x \in [\mathbf{x}_0, \mathbf{x}_1]. \tag{4.9}$$

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

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

In our case, the empirical squared error is equal to

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^{n} \lambda(j) \widehat{\theta}_{j,n} \theta_{j,n} + \sum_{j=1}^{n} \theta_{j,n}^2. \tag{4.10}$$

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

$$\widetilde{\theta}_{j,n} = \widehat{\theta}_{j,n}^2 - \frac{\check{\mathbf{x}}}{n} \widehat{\sigma}_{j,n} \quad \text{and} \quad \widehat{\sigma}_{j,n} = \frac{\check{\mathbf{x}}}{n} \sum_{l=1}^{n} \widehat{\sigma}_l \phi_j^2(z_l), \tag{4.11}$$

where $\widehat{\sigma}_l$ is the estimator for $\sigma_l^2$ defined in (3.8). Note that if the diffusion is known, then we take in (4.11) $\widehat{\sigma}_{j,n} = \sigma_{j,n}$ and

$$\sigma_{j,n} = \frac{\check{\mathbf{x}}}{n} \sum_{l=1}^{n} \sigma_l^2 \phi_j^2(z_l). \tag{4.12}$$

It is clear that the inequalities (3.9) imply

$$\sigma_{0,*} \le \min_{1\le l\le n} \sigma_{l,n} \le \max_{1\le l\le n} \sigma_{l,n} \le \sigma_{1,*}. \tag{4.13}$$

Now, for using the estimator (4.11) instead of $\theta_{j,n} \widehat{\theta}_{j,n}$ one needs to add to the cost function a suitable penalty term that we take as

$$\widehat{P}_n(\lambda) = \frac{\check{\mathbf{x}}}{n} \sum_{j=1}^{n} \lambda^2(j) \widehat{\sigma}_{j,n} \tag{4.14}$$

if the diffusion is unknown and as

$$P_n(\lambda) = \frac{\check{\mathbf{x}}}{n} \sum_{j=1}^{n} \lambda^2(j)\, \sigma_{j,n} \tag{4.15}$$

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

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,n} + \rho\, \widehat{P}_n(\lambda), \tag{4.16}$$

where the positive coefficient $0 < \rho < 1$ will be specified later.
We define the model selection procedure as

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda\in\Lambda} J_n(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}. \tag{4.17}$$

***Remark 4.2*** It should be emphasized that if in the model (1.3) the diffusion coefficient $b(\cdot)$ in known, then we use the penalty term (4.15).

To obtain the efficient properties for this procedure we will use the special weight coefficients introduced in Galtchouk and Pergamenshchikov (2009a, b). To this end we consider a $2-$dimensional numerical grid of the form

$$\mathcal{A} = \{1, \ldots, \mathbf{k}^*\} \times \{l_1, \ldots, l_{m^*}\}, \tag{4.18}$$

where $l_i = i\varepsilon$ and $m^* = [1/\varepsilon^2]$. 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 $a > 0$,

$$\lim_{T\to\infty} \left( \varepsilon_T + \frac{1}{T^a \varepsilon_T} + \frac{1}{\mathbf{k}^*_T} + \frac{\mathbf{k}^*_T}{\ln T} \right) = 0. \tag{4.19}$$

One can take, for example, $\varepsilon_T = 1/\ln(T+1)$ and $\mathbf{k}^* = \overline{k} + \sqrt{\ln(T+1)}$ for some fixed $\overline{k} \geq 1$. Now, for $\alpha = (k,l) \in \mathcal{A}$, we define the vector $\lambda_\alpha = (\lambda_\alpha(j))_{j\geq 1}$ as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\leq j\leq j_0\}} + \left(1 - (j/\omega_\alpha)^k\right) \mathbf{1}_{\{j_0<j\leq\omega_\alpha\}}, \tag{4.20}$$

where $j_0 = j_0(\alpha) = \left[\omega_\alpha/\ln(T+1)\right]$, $\omega_\alpha = \check{\omega}_k\,(lT)^{1/(2k+1)}$ and

$$\check{\omega}_k = \check{\mathbf{x}} \left( \frac{(k+1)(2k+1)}{\pi^{2k} k} \right)^{1/(2k+1)}.$$

We set

$$\Lambda = \left(\lambda_\alpha\right)_{\alpha\in\mathcal{A}}. \tag{4.21}$$

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 (4.19) imply that, $\lim_{T\to\infty} T^{-a}\nu_T = 0$ for any $a > 0$.
Moreover, from (4.20) we can obtain that, for any $\alpha \in \mathcal{A}$,

$$\sum_{j=1}^{n} \lambda_\alpha(j) \leq \omega_\alpha \leq \check{\omega}_k \left(\frac{T}{\varepsilon_T}\right)^{1/3}.$$

Therefore, $\lim_{T\to\infty} T^{-1/3-a}\,\Lambda_* = 0$, for any $a > 0$, and the condition $\mathbf{A}_3$) holds.

***Remark 4.3*** Note that, in Galtchouk and Pergamenshchikov (2019) the weight vectors (4.20) are defined on the basis of number of the points $n \approx \sqrt{T}$ and the oracle inequality is shown under the condition which is slightly different from (4.8). Unfortunately, it turns out, that such weight vectors do not provide the efficient estimation. By this reason we replaced in (4.20) $n$ by $T$ and we modified the condition $\mathbf{A}_3$).

## 5 Main results

**Oracle inequalities.** First we study non asymptotic properties for the procedure (4.17).

**Theorem 5.1** *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* (4.17) *satisfies the inequality*

$$\mathcal{R}_\vartheta(\widehat{S}_*) \leq \frac{(1+\rho)^2(1+4\rho)}{1-6\rho} \min_{\lambda\in\Lambda} \mathcal{R}_\vartheta(\widehat{S}_\lambda) + \frac{\mathbf{U}_{\vartheta,T}}{\rho T}, \tag{5.1}$$

*where the remainder term* $\mathbf{U}_{\vartheta,T}$ *is such that, for any* $a > 0$,

$$\lim_{T\to\infty} T^{-a} \sup_{\vartheta\in\Theta} \mathbf{U}_{\vartheta,T} = 0. \tag{5.2}$$

**Theorem 5.2** *Assume that the conditions* $\mathbf{A}_1$)–$\mathbf{A}_2$) *hold. Then, the model selection procedure* (4.17) *with the weights* (4.21) *satisfies the oracle inequality* (5.1) *with the remainder term satisfying the property* (5.2), *for any* $a > 0$.

***Remark 5.1*** Note, that similarly to Galtchouk and Pergamenshchikov (2011), we will use the inequality (5.1) to provide the efficiency property in the adaptive setting, i.e. without using the regularity of the unknown function $S$. More precisely, through this inequality we can estimate from above the risk for the model selection procedure by the risk of the efficient estimator constructed on the basis of the regularity parameters of $S$ and, as a consequence we obtain the adaptive efficiency property for $\widehat{S}_*$.

**Adaptive efficiency property.** To study minimax properties for the procedure (4.17) we use the functional Sobolev ball defined as

$$W^0_{k,r} = \left\{ f \in \mathbf{C}^k_0([\mathbf{x}_0, \mathbf{x}_1]) : \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq r \right\}, \tag{5.3}$$

where $r > 0$ and the integer $k \geq 1$ are some parameters, $\mathbf{C}^k_0([\mathbf{x}_0, \mathbf{x}_1])$ is the space of $k$ times differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that $f^{(i)}(x) = 0$ for $0 \leq i \leq k-1$ and $x \notin [\mathbf{x}_0, \mathbf{x}_1]$. Moreover, let $S_0$ be a fixed continuously differentiable function from $\Sigma_{\mathbf{L},\mathbf{M}}$. We set

$$W_{k,r} = S_0 + W^0_{k,r} \quad \text{and} \quad \Theta_{k,r} = W_{k,r} \times \mathcal{B}. \tag{5.4}$$

Note that one can represent the class $W^0_{k,r}$ as an ellipse in $\mathcal{L}_2[\mathbf{x}_0, \mathbf{x}_1]$ with trigonometric basis (4.2), i.e.

$$W^0_{k,r} = \left\{ f \in \mathbf{C}^k_0([\mathbf{x}_0, \mathbf{x}_1]) : \sum_{j=0}^{\infty} a_j \theta_j^2 \leq r \right\}, \tag{5.5}$$

where $\theta_j = (f, \phi_j) = \int_{\mathbf{x}_0}^{\mathbf{x}_1} f(x)\phi_j(x)dx$ and $a_j = \sum_{i=0}^k \left(2\pi[j/2]/\check{\mathbf{x}}\right)^{2i}$.
To study the minimal value for the quadratic risks we set

$$l_* = \frac{(2k+1)r^{1-\iota_k} k^{\iota_k}}{(\pi(k+1)(2k+1))^{\iota_k}} \quad \text{and} \quad \mathbf{J}_\vartheta = \int_{\mathbf{x}_0}^{\mathbf{x}_1} \frac{b^2(x)}{\mathbf{q}_\vartheta(x)} \mathrm{d}x, \tag{5.6}$$

where $\iota_k = 2k/(2k+1)$. It is well known that, for any $S \in \Theta_{k,r}$, the optimal rate of convergence of estimators is $T^{-\iota_k}$ (see, for example, Galtchouk and Pergamenshchikov 2004).

Now we denote by $\Xi_T$ the set of all possible estimators of $S$ which are measurable with respect to the $\sigma$-field $\sigma\{y_t, 0 \le t \le T\}$, i.e. based on the observations $(y_t)_{0\le t\le T}$.

**Theorem 5.3** *For any integer $k \ge 1$ and $r > 0$, the quadratic risk $\mathcal{R}_\vartheta(\widehat{S})$ with the normalizing coefficient $\upsilon(\vartheta) = \mathbf{J}_\vartheta^{-\iota_k}$ admits the following lower bound*

$$\liminf_{T\to\infty} T^{\iota_k} \inf_{\widehat{S}\in\Xi_T} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\mathcal{R}_\vartheta(\widehat{S}) \ge l_*.$$

Assume now, that the penalty parameter $\rho$ in (4.16) is a function of $T$, i.e. $\rho = \rho_T$ such that, for any $a > 0$,

$$\lim_{T\to\infty} \rho_T = 0 \quad \text{and} \quad \lim_{T\to\infty} T^a \rho_T = \infty. \tag{5.7}$$

We can take, for example, $\rho_T = (6 + \ln(T+1))^{-1}$.

**Theorem 5.4** *Assume that the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$) hold. Then, for any integer $k \ge 2$ and $r > 0$, the quadratic risk for the model selection procedure $\widehat{S}_*$ defined in (4.17) with the parameter $\rho$ of the form (5.7) through the trigonometric basis (4.2) and the weight family (4.21) satisfies the following upper bound*

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\mathcal{R}_\vartheta(\widehat{S}_*) \le l_*. \tag{5.8}$$

Theorems 5.3– 5.4 imply immediately the efficiency property.

**Theorem 5.5** *Under the conditions of Theorem 5.4 the model selection procedure $\widehat{S}_*$ is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} \frac{\inf_{\widehat{S}\in\Xi_T} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\mathcal{R}_\vartheta(\widehat{S})}{\sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\mathcal{R}_\vartheta(\widehat{S}_*)} = 1.$$

***Remark 5.2*** It should be noted that from Theorems 5.3– 5.4 it follows that the coefficient $J_\vartheta^{\iota_k}\mathbf{l}_*$ is the well-known Pinsker constant, which is calculated for the first time for the model (1.3). In the particular case, for the model (1.3) with the diffusion coefficient $b(\cdot) \equiv 1$ the Pinsker constant was calculated in Galtchouk and Pergamenshchikov (2011). Note also that the parameter $l_*$ is the well-known Pinsker constant for the "signal plus white noise" model obtained in Pinsker (1981). Therefore, the Pinsker constant for the model (1.3) is obtained by multiplying the constant from Pinsker (1981) by the Pinsker variance (5.6) in the power $\iota_k$.

**Big data analysis.** Now we apply the developed methods to the model (1.1). We assume that the functions $(\psi_j)_{1\le j\le q}$ are orthonormal in $\mathcal{L}_2[\mathbf{x}_0, \mathbf{x}_1]$, i.e. $(\psi_i, \psi_j) = \int_{\mathbf{x}_0}^{\mathbf{x}_1} \psi_i(t)\psi_j(t)\mathrm{d}t = \mathbf{1}_{\{i=j\}}$. We use the estimators (4.9) to estimate the parameters $\beta = (\beta_j)_{1\le j\le q}$ as $\widehat{\beta}_\lambda = (\widehat{\beta}_{\lambda,j})_{1\le j\le q}$ and $\widehat{\beta}_{\lambda,j} = (\psi_j, (\widehat{S}_\lambda - \psi_0))$. Moreover, to estimate these parameters we use also the model selection procedure (4.17) setting $\widehat{\beta}_{*,j} = (\psi_j, (\widehat{S}_* - \psi_0))$ and $\widehat{\beta}_* = (\widehat{\beta}_{*,j})_{1\le j\le q}$. Note that $|\widehat{\beta}_\lambda - \beta|_q^2 = \sum_{j=1}^q (\widehat{\beta}_{\lambda,j} - \beta_j)^2 = \|\widehat{S}_\lambda - S\|^2$ and $|\widehat{\beta}_* - \beta|_q^2 = \|\widehat{S}_* - S\|^2$. Therefore, Theorem 5.1 implies the following oracle inequality.

**Theorem 5.6** *Assume that the conditions* $\mathbf{A}_1$)–$\mathbf{A}_3$) *hold. Then, for any* $T \geq 1$, $0 < \rho \leq 1/8$ *and* $\vartheta \in \Theta$,

$$\mathbf{E}_\vartheta |\widehat{\beta}_* - \beta|_q^2 \leq \frac{(1+\rho)^2(1+4\rho)}{1-6\rho} \min_{\lambda \in \Lambda} \mathbf{E}_\vartheta |\widehat{\beta}_\lambda - \beta|_q^2 + \frac{\mathbf{U}_{\vartheta,T}}{\rho T}, \tag{5.9}$$

*where the term* $\mathbf{U}_{\vartheta,T}$ *satisfies the property (5.2).*

Theorem 5.5 imply the efficiency property for the estimator $\widehat{\beta}_*$ based on the model selection procedure (4.17) constructed through the trigonometric basis (4.2) with the weight coefficients (4.21).

**Theorem 5.7** *Assume that the conditions of Theorem 5.4 hold. Then the estimate* $\widehat{\beta}_*$ *is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} \frac{\inf_{\widehat{\beta}_T \in \Xi_T} \sup_{\vartheta \in \Theta_{k,r}} \upsilon(\vartheta) \mathbf{E}_\vartheta |\widehat{\beta}_T - \beta|_q^2}{\sup_{\vartheta \in \Theta_{k,r}} \upsilon(\vartheta) \mathbf{E}_\vartheta |\widehat{\beta}_* - \beta|_q^2} = 1, \tag{5.10}$$

*where* $\Xi_T$ *is the set of all possible estimators for the vector* $\beta = (\beta_j)_{1\leq j\leq q}$.

***Remark 5.3*** Note, that in the estimators $\widehat{\beta}_{*,j}$ it is not used the parameter dimension $q$. Moreover, it can be equal to $+\infty$. In this case it is impossible to use neither LASSO method nor Danzig selector. It should be emphasized also that the efficiency property (5.10) is shown without using any sparse conditions for the parameters $\beta = (\beta_j)_{1\leq j\leq q}$ usually assumed for such problems (see, for example, Fan et al. 2014).

## 6 Properties of the model (4.5)

To prove the oracle inequalitiy (5.1) we need to modify the analytical tool developed in Galtchouk and Pergamenshchikov (2011). To this end we need to study the following functions

$$\Xi(\lambda) = \sum_{j=1}^n \lambda(j)\, \xi_{j,n} \quad \text{and} \quad \mathbf{B}(\lambda) = \frac{\check{\mathbf{x}}}{\sqrt{n}} \sum_{j=1}^n \lambda(j)\, \widetilde{\xi}_{j,n}, \tag{6.1}$$

where $\lambda \in \mathbb{R}^n$, the variables $\xi_{j,n}$ are defined in (4.5) and $\widetilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}\xi_{j,n}^2$.

**Proposition 6.1** *For any* $n \geq 1$ *and any* $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$,

$$\mathbf{E}_\vartheta\ \Xi^2(\lambda) \leq \sigma_{1,*} |\lambda|^2. \tag{6.2}$$

***Proof*** From Proposition 3.2, (4.1), (4.5) and (3.9) we can obtain directly that

$$\mathbf{E}_\vartheta\ \Xi^2(\lambda) = \frac{\check{\mathbf{x}}^2}{n^2} \mathbf{E}_\vartheta \sum_{l=1}^n \sigma_l^2 \left( \sum_{j=1}^n \lambda(j) \phi_j(z_l) \right)^2 \leq \sigma_{1,*} \frac{\check{\mathbf{x}}}{n} \sum_{j=1}^n \lambda^2(j) \frac{\check{\mathbf{x}}}{n} \sum_{l=1}^n \phi_k^2(z_l).$$

Hence Proposition 6.1. □

**Proposition 6.2** *For any* $n \geq 1$ *and any* $\lambda = (\lambda_1, \ldots, \lambda_n) \in [0,1]^n$,

$$\mathbf{E}_\vartheta\ \left(B^2(\lambda) | \mathcal{G}_{N_0}\right) \leq 6\sigma_{1,*} \check{\mathbf{x}}\, P_n(\lambda). \tag{6.3}$$

***Proof*** First we represent the variables $\widetilde{\xi}_{j,n}$ as

$$\widetilde{\xi}_{j,n} = \frac{\check{\mathbf{x}}}{n} \sum_{l=1}^{n} \left( \sigma_l^2 \phi_j^2(z_l) \eta_l + 2\mathbf{1}_{\{l \geq 2\}} \xi_l \, \mathbf{u}_{j,l} \right), \quad \eta_l = \xi_l^2 - 1,$$

where $\mathbf{u}_{j,l} = \sigma_l \phi_j(z_l) \sum_{r=1}^{l-1} \sigma_r \phi_j(z_r) \xi_r$. Using this in (6.1), we get

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

where $\gamma_{1,l} = \sigma_l^2 \sum_{j=1}^{n} \lambda(j) \phi_j^2(z_l)$ and $\gamma_{2,l} = \sum_{j=1}^{n} \lambda(j) \mathbf{u}_{j,l} \mathbf{1}_{\{l \geq 2\}}$. Now Proposition 3.2 implies, that

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

Due to the Buniakovski-Cauchy-Schwartz inequality

$$\gamma_{1,l}^2 = \sigma_l^4 \left( \sum_{j=1}^{n} \lambda(j) \phi_j^2(z_l) \right)^2 \leq \sigma_l^4 \left( \sum_{j=1}^{n} \lambda^2(j) \phi_j^2(z_l) \right) \left( \sum_{j=1}^{n} \phi_j^2(z_l) \right)$$
$$\leq \sigma_{1,*} \frac{n}{\check{\mathbf{x}}} \sigma_l^2 \sum_{j=1}^{n} \lambda^2(j) \phi_j^2(z_l).$$

Therefore,

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

Moreover, using the property (4.1), we get

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

Therefore, the term $M_{1,2}$ can be estimated as

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

These imply the upper bound (6.3). Hence Proposition 6.2. □

## 7 Sharp upper bound

To obtain the upper bound (5.8), first we assume that the parameters $k$, $r$ and the function $\mathbf{J}_\vartheta$ are known. In this case, we use the estimator (4.9) with the weight $\widetilde{\lambda}$ from the family (4.21) defined as

$$\widetilde{S} = \widetilde{S}_{\widetilde{\lambda}} \quad \text{and} \quad \widetilde{\lambda} = \lambda_{\widetilde{\alpha}}, \tag{7.1}$$

where $\widetilde{\alpha} = (k, \widetilde{l})$, $\widetilde{l} = \widetilde{l}_T = [\overline{r}(\vartheta)/\varepsilon]\,\varepsilon$, $\overline{r}(\vartheta) = r/\mathbf{J}_\vartheta$ and $\varepsilon = \varepsilon_T = 1/\ln(T+1)$. Now we need to study asymptotic properties of the vector $\widetilde{\lambda}$. To this end we set

$$\widetilde{\Upsilon}_T(\vartheta) = \sum_{j=1}^{n} (1-\widetilde{\lambda}(j))^2 \theta^2_{j,n} + \frac{\mathbf{J}_\vartheta}{\check{\mathbf{x}} n^2} \sum_{j=1}^{n} \widetilde{\lambda}^2(j). \tag{7.2}$$

**Proposition 7.1** *For any $k \geq 2$ and $r > 0$,*

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\widetilde{\Upsilon}_T(\vartheta) \leq l_*. \tag{7.3}$$

***Proof*** First of all, note that

$$0 < \inf_{\vartheta\in\Theta_{k,r}} \mathbf{J}_\vartheta \leq \sup_{\vartheta\in\Theta_{k,r}} \mathbf{J}_\vartheta < \infty. \tag{7.4}$$

This implies directly that

$$\lim_{T\to\infty} \sup_{\vartheta\in\Theta_{k,r}} \left| \frac{\widetilde{l}_T}{\overline{r}(\vartheta)} - 1 \right| = 0, \tag{7.5}$$

where $\overline{r}(\vartheta) = r/\mathbf{J}_\vartheta$. Moreover, note that

$$T^{\iota_k} \upsilon(\vartheta)\widetilde{\Upsilon}_T(\vartheta) \leq T^{\iota_k}\, \upsilon(\vartheta)\, \mathbf{S}_T + \frac{(\mathbf{J}_\vartheta)^{1-\iota_k}}{T^{1-\iota_k}\check{\mathbf{x}}} \sum_{j=1}^{n} \widetilde{\lambda}^2(j), \tag{7.6}$$

where $\mathbf{S}_T = \sum_{j=1}^{n} (1-\widetilde{\lambda}(j))^2\, \theta^2_{j,n}$. We decompose $\mathbf{S}_T$ as

$$\mathbf{S}_T = \sum_{j=j_0+1}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\, \theta^2_{j,n} + \sum_{j=[\widetilde{\omega}]+1}^{n} \theta^2_{j,n} := \mathbf{S}_{1,T} + \mathbf{S}_{2,T},$$

where $\widetilde{\omega} = \omega_{\widetilde{\alpha}} = \check{\omega}_k \left(T\widetilde{l}_T\right)^{1/(2k+1)}$. Lemmas A.5 and A.6 yield

$$\mathbf{S}_{1,T} \leq (1+\widetilde{\varepsilon}) \sum_{j=j_0}^{[\widetilde{\omega}]} (1-\widetilde{\lambda}(j))^2\, \theta^2_j + 2r(1+\widetilde{\varepsilon}^{-1})\, \frac{\widetilde{\omega}}{n^{2k}},$$

and

$$\mathbf{S}_{2,T} \leq (1+\widetilde{\varepsilon}) \sum_{j>\widetilde{\omega}} \theta^2_j + (1+\widetilde{\varepsilon}^{-1})\, \frac{r}{n^2\, \widetilde{\omega}^{2(k-1)}}.$$

Therefore,

$$\mathbf{S}_T \leq (1+\widetilde{\varepsilon})\mathbf{S}^*_T + 2r(1+\widetilde{\varepsilon}^{-1})\, \gamma_T, \tag{7.7}$$

where

$$\mathbf{S}^*_T = \sum_{j\ge 1}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 = \sum_{j\le\widetilde{\omega}}(1-\widetilde{\lambda}(j))^2\,\theta_j^2 + \sum_{j>\widetilde{\omega}}\theta_j^2 := \mathbf{S}^*_{1,T} + \mathbf{S}^*_{2,T}.$$

and $\gamma_T = \widetilde{\omega}n^{-2k} + n^{-2}\widetilde{\omega}^{-2(k-1)}$. Note, that

$$T^{\iota_k}\upsilon(\vartheta)\mathbf{S}^*_{1,T} = \frac{\upsilon(\vartheta)}{\check{\omega}_k^{2k}(\widetilde{l}_T)^{\iota_k}}\sum_{j=j_0}^{[\widetilde{\omega}]} j^{2k}\,\theta_j^2 \le \frac{\upsilon(\vartheta)}{\check{\omega}_k^{2k}(\widetilde{l}_T)^{\iota_k}}\,\varpi_{j_0}\sum_{j=j_0}^{[\widetilde{\omega}]} a_j\,\theta_j^2,$$

where $\varpi_n = \sup_{j\ge n} j^{2k}/a_j$. It is clear that $\lim_{n\to\infty}\varpi_n = \check{\mathbf{x}}^{2k}/\pi^{2k}$. Therefore, from (7.5) we obtain that

$$\limsup_{n\to\infty}\sup_{\vartheta\in\Theta_{k,r}}\frac{T^{\iota_k}\upsilon(\vartheta)\,\mathbf{S}^*_{1,T}}{\sum_{j=j_0}^{[\widetilde{\omega}]} a_j\,\theta_j^2} \le \frac{\check{\mathbf{x}}^{2k}}{\pi^{2k}\,\check{\omega}_k^{2k}\,r^{\iota_k}} = \left(\frac{k}{(2k+1)(k+1)r\pi}\right)^{\iota_k}.$$

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

$$\mathbf{S}^*_{2,T} = \sum_{j>\widetilde{\omega}}\theta_j^2 \le (1+\widetilde{\varepsilon})\frac{\check{\mathbf{x}}^{2k}}{\pi^{2k}\,\widetilde{\omega}^{2k}}\sum_{j>\widetilde{\omega}} a_j\,\theta_j^2 = \frac{(1+\widetilde{\varepsilon})\check{\mathbf{x}}^{2k}}{\pi^{2k}\,(T\widetilde{l}_T)^{\iota_k}\check{\omega}_k^{2k}}\sum_{j>\widetilde{\omega}} a_j\,\theta_j^2.$$

Therefore, in view of (7.5) we get that

$$\limsup_{n\to\infty}\sup_{\vartheta\in\Theta_{k,r}}\frac{T^{\iota_k}\upsilon(\vartheta)\,\mathbf{S}^*_{2,T}}{\sum_{j>\widetilde{\omega}} a_j\,\theta_j^2} \le \frac{\check{\mathbf{x}}^{2k}}{\pi^{2k}\check{\omega}_k^{2k}r^{\iota_k}} = \left(\frac{k}{(2k+1)(k+1)r\pi}\right)^{\iota_k}.$$

Note now, that in (7.7) for $k\ge 2$ we have $\lim_{T\to\infty}\sup_{S\in W_{k,r}} T^{\iota_k}\gamma_T = 0$, i.e.

$$\limsup_{T\to\infty}\sup_{\vartheta\in\Theta_{k,r}} T^{\iota_k}\,\upsilon(\vartheta)\,\mathbf{S}_T \le r^{1-\iota_k}\left(\frac{k}{(2k+1)(k+1)\pi}\right)^{\iota_k}. \qquad (7.8)$$

Moreover, we can check directly that

$$\limsup_{T\to\infty}\sup_{\vartheta\in\Theta_{k,r}}\frac{\mathbf{J}_\vartheta^{1-\iota_k}}{T^{1-\iota_k}\check{\mathbf{x}}}\sum_{j=1}^{n}\widetilde{\lambda}^2(j) \le \mathbf{u}^*_k, \qquad (7.9)$$

where

$$\mathbf{u}^*_k = \frac{\check{\omega}_k r^{1-\iota_k}2k^2}{\check{\mathbf{x}}(2k+1)(k+1)} = 2kr^{1-\iota_k}\left(\frac{k}{(2k+1)(k+1)\pi}\right)^{\iota_k}.$$

Using now (7.8) and (7.9) in (7.6) we obtain the limit equality (7.3) and, hence Proposition 7.1. □

Now, we can study the estimator $\widetilde{S}$. For this we need to use the mean variance

$$\mathbf{s}_n = \check{\mathbf{x}}\sum_{l=1}^{n}\sigma_l^2, \qquad (7.10)$$

where the variances $(\sigma_l)_{1\le l\le n}$ are defined in (3.4).

**Theorem 7.2** *For any $k\ge 2$ and $r>0$, the estimator $\widetilde{S}$ from (7.1) satisfies the inequality*

$$\limsup_{T\to\infty} T^{\iota_k}\sup_{\vartheta\in\Theta_{k,r}}\upsilon(\vartheta)\,\mathbf{E}_\vartheta\|\widetilde{S}-S\|_n^2 \le l_*. \qquad (7.11)$$

***Proof*** First note that, in view of Proposition 3.1, to prove this Theorem it suffices to show that

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\, \mathbf{E}_\vartheta\, \chi_{\mathbf{G}_*} \|\widetilde{S}-S\|_n^2 \le l_*. \tag{7.12}$$

Indeed, note that due to (4.3)–(4.6) on the set $\mathbf{G}_*$ we obtain that

$$\begin{aligned}\|\widetilde{S}-S\|_n^2 &= \sum_{j=1}^n \left(\widetilde{\lambda}^2(j)(\theta_{j,n}+\zeta_{j,n})^2 - 2\widetilde{\lambda}(j)(\theta_{j,n}+\zeta_{j,n})\theta_{j,n}+\theta_{j,n}^2\right)\\ &= \sum_{j=1}^n \left((\widetilde{\lambda}(j)-1)^2\theta_{j,n}^2 + 2\widetilde{\lambda}(j)(\widetilde{\lambda}(j)-1)\zeta_{j,n}\theta_{j,n}+\widetilde{\lambda}^2(j)\zeta_{j,n}^2\right),\end{aligned} \tag{7.13}$$

where $\widetilde{\lambda}$ is defined in (7.1). Taking into account the definition $\zeta_{j,n}$ in (4.5) and using the inequality

$$2xy \le \varepsilon x^2 + \varepsilon^{-1} y^2, \tag{7.14}$$

we get that on the set $\mathbf{G}_*$

$$\begin{aligned}\|\widetilde{S}-S\|_n^2 \le\ & (1+\varepsilon)\left(\sum_{j=1}^n (1-\widetilde{\lambda}(j))^2\theta_{j,n}^2 + \frac{\check{\mathbf{x}}}{n}B_1\right) + \left(1+\frac{2}{\varepsilon}\right)B_2\\ &+ 2\sqrt{\frac{\check{\mathbf{x}}}{n}}\sum_{j=1}^n \widetilde{\lambda}(j)(\widetilde{\lambda}(j)-1)\theta_{j,n}\xi_{j,n},\end{aligned} \tag{7.15}$$

where $B_1=\sum_{j=1}^n \widetilde{\lambda}^2(j)\xi_{j,n}^2$ and $B_2=\sum_{j=1}^n \widetilde{\lambda}^2(j)\mathbf{g}_{j,n}^2$. Since $\mathbf{E}\,\xi_{j,n}=0$, we get

$$\begin{aligned}\mathbf{E}_\vartheta \|\widetilde{S}-S\|_n^2\, \chi_{\mathbf{G}_*} \le\ & (1+\varepsilon)\left(\sum_{j=1}^n (1-\widetilde{\lambda}(j))^2\theta_{j,n}^2 + \mathbf{E}_\vartheta\, B_1\right)\\ &+ \left(1+\frac{2}{\varepsilon}\right)\mathbf{E}_\vartheta B_2\, \chi_{\mathbf{G}_*} - 2\sqrt{\frac{\check{\mathbf{x}}}{n}}\,\mathbf{E}_\vartheta\, B_3\chi_{\mathbf{G}_*^c},\end{aligned} \tag{7.16}$$

where $B_3=\sum_{j=1}^n \widetilde{\lambda}(j)(\widetilde{\lambda}(j)-1)\xi_{j,n}\theta_{j,n}$. Note that

$$\check{\mathbf{x}}\mathbf{E}_\vartheta\left(\xi_{j,n}^2|\mathcal{G}_{N_0}\right) = \frac{\mathbf{s}_n}{n} + \frac{\check{\mathbf{x}}}{n}\sum_{l=1}^n \sigma_l^2\overline{\phi}_j(z_l) \tag{7.17}$$

where $\mathbf{s}_n$ is defined in (7.10) and $\overline{\phi}_j(z_l)=\check{\mathbf{x}}\phi_j^2(z_l)-1$. Moreover, setting now $\bar{\mathbf{s}}_n=\mathbf{s}_n-\mathbf{J}_\vartheta/\check{\mathbf{x}}$, we can represent the term $\mathbf{E}_\vartheta\, B_1$ as

$$\begin{aligned}\mathbf{E}_\vartheta\, B_1 &= \frac{1}{n^2}\mathbf{E}_\vartheta \sum_{j=1}^n \widetilde{\lambda}^2(j)\left(\mathbf{s}_n + \check{\mathbf{x}}\sum_{l=1}^n \sigma_l^2\overline{\phi}_j(x_l)\right)\\ &= \frac{\mathbf{J}_\vartheta}{\check{\mathbf{x}}n^2}\sum_{j=1}^n \widetilde{\lambda}^2(j) + \frac{B_{11}}{n^2} + \frac{\check{\mathbf{x}}B_{12}}{n^2},\end{aligned} \tag{7.18}$$

where $B_{11} = \sum_{j=1}^{n} \widetilde{\lambda}^2(j)\,\mathbf{E}_\vartheta \bar{\mathbf{s}}_n$ and $B_{12} = \mathbf{E}_\vartheta \sum_{j=1}^{n} \widetilde{\lambda}^2(j) \sum_{l=1}^{n} \sigma_l^2 \overline{\phi}_j(x_l)$. Now from (7.16) and (7.18) it follows that

$$\mathbf{E}_\vartheta \,\|\widetilde{S} - S\|_n^2 \,\chi_{\mathbf{G}_*} \le (1+\varepsilon)\left(\widetilde{\Upsilon}_T + \frac{B_{11}}{n^2} + \frac{\check{\mathbf{x}} B_{12}}{n^2}\right) + \left(1 + \frac{2}{\varepsilon}\right) \mathbf{E}_\vartheta \, B_2 \,\chi_{\mathbf{G}_*} - 2\mathbf{E}_\vartheta \, B_3 \chi_{\mathbf{G}_*^c}. \tag{7.19}$$

The bound (7.9) and Proposition A.3 yield

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \frac{1}{n^2} |\, B_{11}| = 0.$$

As to the term $B_{12}$, due to Lemma A.7, one has

$$\frac{\check{\mathbf{x}}}{n^2} |B_{12}| \le \frac{\check{\mathbf{x}}}{n^2} \mathbf{E}_\vartheta \sum_{l=1}^{n} \sigma_l^2 \left| \sum_{j=1}^{n} \widetilde{\lambda}^2(j) \overline{\phi}_j(x_l) \right| \le 2^{k+1} \mathbf{E}_\vartheta \frac{\mathbf{s}_n}{n^2}. \tag{7.20}$$

Therefore, in view of Proposition A.3, we obtain that

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \frac{|B_{12}|}{n^2} = 0.$$

Moreover, taking into account that $B_2 \le \sum_{j=1}^{n} \mathbf{g}_{j,n}^2 = \|\mathbf{g}\|_n^2$, we obtain that

$$\mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*} B_2 \le \mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*} \|\mathbf{g}\|_n^2 = \frac{\check{\mathbf{x}}}{n} \sum_{k=1}^{n} \mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*} \mathbf{g}_k^2 \le \check{\mathbf{x}} \frac{\mathbf{g}_T^*}{T},$$

where $\mathbf{g}_T^*$ is given by (3.6). Therefore, in view of Proposition 3.3

$$\lim_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*} \, B_2 = 0. \tag{7.21}$$

Now through Proposition 6.1 we estimate the last term in (7.16), i.e.

$$\mathbf{E}_\vartheta \, B_3^2 \le \sigma_{1,*} \frac{\check{\mathbf{x}}}{n} \sum_{j=1}^{n} \theta_{j,n}^2 = \frac{\sigma_{1,*}\check{\mathbf{x}}}{n} \|S\|_n^2.$$

Therefore,

$$\mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*^c} |B_3| \le \sqrt{\frac{\sigma_{1,*}\check{\mathbf{x}}}{n}} \, \|S\|_n \, \mathbf{P}_\vartheta \left(\mathbf{G}_*^c\right)$$

and by Proposition 3.1, we get $\lim_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \mathbf{E}_\vartheta \, \mathbf{1}_{\mathbf{G}_*^c} |B_3| = 0$. Therefore, using Proposition 7.1 in (7.19) we obtain Theorem 7.2. □

Note that Lemma A.4 implies the following upper bound.

**Theorem 7.3** *The quadratic risk for the estimating procedure $\widetilde{S}$ from (7.1) has the following asymptotic upper bound*

$$\limsup_{T\to\infty} T^{\iota_k} \sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\mathcal{R}(\widetilde{S}, S) \le l_*. \tag{7.22}$$

***Remark 7.1*** It should be noted that the inequality (7.22) through Theorem 5.3 means that the estimator $\widetilde{S}$ is efficient. Unfortunately, we can't calculate this estimator, since it depends on unknown parameters $k$, $r$ and $\mathbf{J}_\vartheta$. But, this estimator belongs to the family $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ with the weight vectors defined in (4.21). Therefore, through the oracle inequality we can estimate the risk of the model selection procedure with the risk of the estimator (7.1) and, therefore, using the property (7.22) we can provide the efficiency property for the procedure (4.17).

## 8 Proofs

### 8.1 Proof of Theorem 5.1

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

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

with $\check{\theta}_{j,n} = \widetilde{\theta}_{j,n} - \theta_{j,n}\widehat{\theta}_{j,n}$. From (4.5) and (4.11) one obtains

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

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

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

$$\text{and} \quad \overline{P}_n(\lambda) = \widehat{P}_n(\lambda) - P_n(\lambda) = \frac{\check{\mathbf{x}}}{n}\sum_{j=1}^{n} \lambda^2(j)\widetilde{\sigma}_{j,n}, \tag{8.2}$$

we obtain from (8.1), that

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

where $M_1(\lambda) = n^{-1/2}\,\mathbf{B}(\lambda) + \Delta(\lambda)$, $B(\lambda)$ is defined in (6.1) and

$$\Delta(\lambda) = \sum_{j=1}^{n} \lambda(j)\,\mathbf{g}_{j,n}^2 + 2\sqrt{\frac{\check{\mathbf{x}}}{n}}\sum_{j=1}^{n} \lambda(j)\,\xi_{j,n}\mathbf{g}_{j,n} := \Delta_1(\lambda) + \Delta_2(\lambda). \tag{8.4}$$

In view of Proposition 6.2, for any $\lambda \in [0,1]^n$,

$$\mathbf{E}_\vartheta\,\left(\mathbf{B}^2(\lambda)|\,\mathcal{G}_{N_0}\right) \le 6\sigma_*\check{\mathbf{x}}\,P_n(\lambda). \tag{8.5}$$

To estimate the first term in (8.4) note that

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

Moreover, using the inequality (7.14), we get that, for any $0 < \varepsilon < 1$,

$$|\Delta_2(\lambda)| \le \varepsilon \frac{\check{\mathbf{x}}}{n} \sum_{j=1}^{n} \lambda^2(j)\, \xi_{j,n}^2 + \frac{\|\mathbf{g}\|_n^2}{\varepsilon} = \varepsilon P_n(\lambda) + \varepsilon \frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} + \frac{\|\mathbf{g}\|_n^2}{\varepsilon},$$

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

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

This implies

$$2|M_1(\lambda)| \le 2\frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2\frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} + 2\varepsilon P_n(\lambda) + 4\varepsilon^{-1}\|\mathbf{g}\|_n^2. \tag{8.7}$$

Since $P_n(\lambda^2) \le P_n(\lambda)$, we get that, for any $0 < \varepsilon < 1$ and $\lambda \in \Lambda$,

$$2\frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2\frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} \le \varepsilon P_n(\lambda) + \frac{2}{\varepsilon n}\left(\frac{\mathbf{B}^2(\lambda)}{P_n(\lambda)} + \frac{\mathbf{B}^2(\lambda^2)}{P_n(\lambda^2)}\right). \tag{8.8}$$

Note that the inequalities (3.9) imply that

$$P_{0,n}(\lambda) \le P_n(\lambda) \le P_{1,n}(\lambda), \tag{8.9}$$

where

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

From the inequalities (8.8) and (8.9) it follows

$$2\frac{|\mathbf{B}(\lambda)|}{\sqrt{n}} + 2\frac{|\mathbf{B}(\lambda^2)|}{\sqrt{n}} \le \varepsilon P_n(\lambda) + \frac{2}{\varepsilon\,\sigma_{0,*}\check{\mathbf{x}}}\mathbf{B}^*(\lambda), \tag{8.10}$$

where $\mathbf{B}^*(\lambda) = \left(\mathbf{B}^2(\lambda)/|\lambda|^2 + \mathbf{B}^2(\lambda^2)/|\lambda^2|^2\right)$. Choosing $\varepsilon = \rho/3$ in (8.7), we get

$$2|M_1(\lambda)| \le \rho P_n(\lambda) + \frac{6}{\rho}\Upsilon_n(\lambda), \qquad \Upsilon_n(\lambda) = \frac{\mathbf{B}^*(\lambda)}{\sigma_{0,*}\check{\mathbf{x}}} + 2\|\mathbf{g}\|_n^2. \tag{8.11}$$

Now we study $\mathbf{D}^* = \max_{\lambda\in\Lambda} |\mathbf{D}(\lambda - \lambda_0)|$. One can check directly that

$$\mathbf{E}_\vartheta\, \mathbf{D}^* \le 2\sum_{\lambda\in\Lambda} \mathbf{E}_\vartheta\, |\mathbf{D}(\lambda)| \le \frac{2\nu\,\Lambda_*}{n}\,\varpi_T^*, \tag{8.12}$$

where $\varpi_T^*$ is defined in (3.10). Similarly, denoting $P^* = \sup_{\lambda\in\Lambda}|\overline{P}_n(\lambda)|$, we get

$$\begin{aligned}\mathbf{E}_\vartheta\, P^* &\le \sum_{\lambda\in\Lambda} \mathbf{E}_\vartheta |\overline{P}_n(\lambda)| \le \sum_{\lambda\in\Lambda} \frac{\check{\mathbf{x}}}{n}\sum_{j=1}^{n}\lambda^2(j)\mathbf{E}_\vartheta|\widehat{\sigma}_{j,n} - \sigma_{j,n}| \\ &\le \sum_{\lambda\in\Lambda}\frac{\check{\mathbf{x}}}{n^2}|\lambda|^2\varpi_T^* \le \frac{\check{\mathbf{x}}}{n^2}\Lambda_*\nu\varpi_T^*.\end{aligned} \tag{8.13}$$

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

$$\begin{aligned}&\mathrm{Err}_n(\widehat{\lambda}) - \mathrm{Err}_n(\lambda_0) \\ &\quad = J_n(\widehat{\lambda}) - J_n(\lambda_0) + 2\,M(\widehat{\mu}) - \frac{2\check{\mathbf{x}}}{n}\mathbf{D}(\widehat{\mu})\end{aligned}$$

$$+2(M_1(\widehat{\lambda})-M_1(\lambda_0))-\rho(P_n(\widehat{\lambda})-P_n(\lambda_0))-\rho(\overline{P}_n(\widehat{\lambda})-\overline{P}_n(\lambda_0)),$$

where $\widehat{\mu}=\widehat{\lambda}-\lambda_0$. By the definition of $\widehat{\lambda}$ in (4.17) we obtain on the set $\mathbf{G}_*$

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) &\le \mathrm{Err}_n(\lambda_0)\,+\,2\,M(\widehat{\mu})+\frac{6}{\rho}\Upsilon_n(\widehat{\lambda})\\ &\quad+\frac{2\check{\mathbf{x}}}{n}\mathbf{D}^*+2\rho P_n(\lambda_0)-\rho(\overline{P}_n(\widehat{\lambda})-\overline{P}_n(\lambda_0)).\end{aligned} \tag{8.14}$$

To study the term $\Upsilon_n(\lambda)$ note, that the bounds (8.5) and (8.9) imply

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

Moreover, as to the norm $\|\mathbf{g}\|_n^2$ note that $g_l\,g_k=0$ for $l\neq k$, i.e.

$$\mathbf{E}_\vartheta\,\mathbf{1}_{\mathbf{G}_*}\|\mathbf{g}\|_n^2=\frac{\check{\mathbf{x}}^2}{n^2}\mathbf{E}_\vartheta\,\sum_{j=1}^n\left(\sum_{l=1}^n g_l\phi_j(z_l)\right)^2=\frac{\check{\mathbf{x}}}{n}\mathbf{E}_\vartheta\,\sum_{l=1}^n g_l^2\le\check{\mathbf{x}}\frac{g_T^*}{T},$$

where $\mathbf{g}_T^*$ is given by (3.6). Therefore,

$$\mathbf{E}_\vartheta\mathbf{1}_{\mathbf{G}_*}\Upsilon_n\le 2\check{\mathbf{x}}\left(\frac{6\sigma_{1,*}^2}{n\sigma_{0,*}}+\frac{\mathbf{g}_T^*}{T}\right). \tag{8.15}$$

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

$$M(\mu)=Z(\mu)+V(\mu)\quad\text{and}\quad\mu=\lambda-\lambda_0, \tag{8.16}$$

where

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

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

$$\overline{S}_\mu=\sum_{j=1}^n\mu(j)\,\theta_{j,n}\phi_j. \tag{8.17}$$

Due to the definition of $\xi_{j,n}$ in (4.5), we can estimate the term $Z(\mu)$ as

$$\mathbf{E}_\vartheta\mathbf{1}_{\mathbf{G}_*}\,Z^2(\mu)\;\le\;\frac{\sigma_{1,*}\check{\mathbf{x}}}{n}\,\|\overline{S}_\mu\|_n^2. \tag{8.18}$$

Moreover, using the inequality (7.14) with $\varepsilon=\rho$, we obtain

$$2V(\mu)=2\sum_{j=1}^n\,\mu(j)\,\theta_{j,n}\mathbf{g}_{j,n}\le\rho\,\|\overline{S}_\mu\|_n^2+\frac{\|\mathbf{g}\|_n^2}{\rho}. \tag{8.19}$$

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

$$2M(\mu)\le\;2\rho\|\overline{S}_\mu\|_n^2+\frac{Z^*}{n\rho}+\frac{\|\mathbf{g}\|_n^2}{\rho},\quad Z^*=\sup_{\mu\in\Lambda-\lambda_0}\frac{nZ^2(\mu)}{\|\overline{S}_\mu\|_n^2}. \tag{8.20}$$

It is clear, that the upper bound (8.18) yields

$$\mathbf{E}_\vartheta\, \mathbf{1}_{\mathbf{G}_*}\, Z^* \le \sum_{\mu\in\Lambda-\lambda_0} \frac{n\mathbf{E}_\vartheta\, \mathbf{1}_{\mathbf{G}_*}\, Z^2(\mu)}{\|\overline{S}_\mu\|_n^2} \le \nu\sigma_{1,*}\check{\mathbf{x}}. \tag{8.21}$$

To estimate the norm $\|\overline{S}_\mu\|_n^2$ note that in view of (4.5) on the set $\mathbf{G}_*$

$$\begin{aligned}\|\overline{S}_\mu\|_n^2 - \|\widehat{S}_\mu\|_n^2 &= \sum_{j=1}^n \mu^2(j)(\theta_{j,n}^2 - \widehat{\theta}_{j,n}^2) \le -2\sum_{j=1}^n \mu^2(j)\,\theta_{j,n}\,\zeta_{j,n}\\ &= -2Z_1(\mu) - 2V_1(\mu),\end{aligned} \tag{8.22}$$

where

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

Taking into account that $|\mu(j)| \le 1$, similarly to the inequality (8.18), we find

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

Moreover, similarly to (8.21) we estimate $Z_1^* = \sup_{\mu\in\Lambda-\lambda_0} nZ_1^2(\mu)/\|\overline{S}_\mu\|_n^2$ as

$$\mathbf{E}_\vartheta\, Z_1^*\,\mathbf{1}_{\mathbf{G}_*} \le \nu\sigma_{1,*}\check{\mathbf{x}}. \tag{8.23}$$

Furthermore, similarly to (8.19) we estimate the second term in (8.22) as

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

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

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

i.e.

$$\|\overline{S}_\mu\|_n^2 \le \frac{1}{1-2\rho}\,\|\widehat{S}_\mu\|_n^2 + \frac{1}{(1-2\rho)\rho}\left(\frac{Z_1^*}{n} + \|\mathbf{g}\|_n^2\right). \tag{8.24}$$

Using this inequality in (8.20) and putting $Z_2^* = Z^* + Z_1^*$ yield on the set $\mathbf{G}_*$

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

Using this bound in (8.14), we obtain that

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

Using here, that $1 - 6\rho > 1/4$ for $\rho < 1/8$ we obtain that for some $\check{\mathbf{l}} > 0$

$$\mathbf{E}_\vartheta \operatorname{Err}_n(\widehat{\lambda})\mathbf{1}_{\mathbf{G}_*} \le \frac{1+2\rho}{1-6\rho}\mathbf{E}_\vartheta \operatorname{Err}_n(\lambda_0)\mathbf{1}_{\mathbf{G}_*} + \frac{2\rho(1-2\rho)}{1-6\rho}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} P_n(\lambda_0) + \frac{\check{\mathbf{l}}}{\rho}\left(\frac{\sigma_{1,*}(\nu+\overline{\sigma}_*)}{n} + \frac{g^*_T}{T} + \frac{\nu\Lambda_*\varpi^*_T}{n^2}\right),$$

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

$$\mathbf{E}_\vartheta \operatorname{Err}_n(\widehat{\lambda})\mathbf{1}_{\mathbf{G}_*} \le \frac{1+4\rho}{1-6\rho}\mathbf{E}_\vartheta \operatorname{Err}_n(\lambda_0)\mathbf{1}_{\mathbf{G}_*} + \frac{\check{\mathbf{l}}}{\rho T}\mathbf{U}_{\vartheta,T},$$

where

$$\mathbf{U}_{\vartheta,T} = \frac{T\sigma_{1,*}(\nu+\overline{\sigma}_*)}{n} + g^*_T + \frac{T\nu\Lambda_*\varpi^*_T}{n^2} + \frac{T\|S\|_n\sqrt{\sigma_{1,*}}}{n}\sqrt{\mathbf{P}_\vartheta(\mathbf{G}^c_*)}.$$

Replacing here $\mathbf{E}_\vartheta \operatorname{Err}_n(\widehat{\lambda})\mathbf{1}_{\mathbf{G}_*}$ and $\mathbf{E}_\vartheta \operatorname{Err}_n(\lambda_0)\mathbf{1}_{\mathbf{G}_*}$ by $\mathbf{E}_\vartheta \|\widehat{S}_* - S\|^2_n - \|S\|^2_n\mathbf{P}_\vartheta(\mathbf{G}^c_*)$ and $\mathbf{E}_\vartheta \|\widehat{S}_{\lambda_0} - S\|^2_n - \|S\|^2_n\mathbf{P}_\vartheta(\mathbf{G}^c_*)$, respectively, and using Lemma A.4 with $\widetilde{\varepsilon} = \rho$, we obtain the inequality (5.1). Moreover, Propositions 3.1 and 3.3, the condition (2.9), Proposition 3.4 and the condition $\mathbf{A}_3$) imply the property (5.2). Hence Theorem 5.1. □

### 8.2 Proof of Theorem 5.3

First, we introduce the auxiliary class $\Theta^0_{k,r} = \left\{\vartheta = (S, b_0) : S \in W_{k,r}, \quad b_0 \equiv 1\right\}$. It is clear that $\Theta^0_{k,r} \subset \Theta_{k,r}$ and $\sup_{\vartheta\in\Theta_{k,r}} \upsilon(\vartheta)\,\mathcal{R}(\widetilde{S}, S) \ge \sup_{\vartheta\in\Theta^0_{k,r}} \upsilon(\vartheta)\,\mathcal{R}(\widetilde{S}, S)$. Using here Theorem 5.2 from Galtchouk and Pergamenshchikov (2011) we obtain Theorem 5.3. □

### 8.3 Proof of Theorem 5.4

Taking into account that for sufficiently large $T$ the estimator (7.1) belongs to the family $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ indexed by the set (4.21), we obtain that

$$\limsup_{T\to\infty} T^{\iota_k}\mathbf{E}_\vartheta \|\widehat{S}_* - S\|^2 \le \limsup_{T\to\infty} T^{\iota_k}\mathbf{E}_\vartheta \|\widehat{S}_{\widetilde{\lambda}} - S\|^2.$$

So, Theorem 7.2 implies immediately Theorem 5.4. □

## Conclusion

In the conclusion we emphasize that in this paper we develop an adaptive sequential model selection method for the drift estimation problem of stochastic differential equations with unknown diffusion coefficients observed at the discrete fixed time moments. It should be noted that in this case we can't use model selection methods developed for such problems (see Remark 3.1 for details). To study the proposed estimation procedures, we find the constructive sufficient conditions $\mathbf{A}_1$)–$\mathbf{A}_3$) on the observations frequency and the model selection procedures under which we obtain the sharp non asymptotic oracle inequality (5.1). Then, through this inequality using the weighted least squares estimators providing the efficient estimation we show in Theorem 5.5, that the constructed model selection procedure is efficient in adaptive setting, i.e. when the drift regularity is unknown. To this end, for the first time, the sharp upper bound for quadratic risk (5.8) (i.e. the celebrate Pinsker constant) is calculated in explicit form for the the model (1.3) (see, Remark 5.2 for details). Moreover, on the basis of the developed model selection procedures we provide in Theorems 5.6 and 5.7 the non asymptotic optimal (in sharp oracle inequalities sense) and asymptotically efficient estimation methods for the high dimension diffusion models (1.1) without using the parameter dimension or any sparse conditions.

**Acknowledgements** 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).

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

# A Appendix

## A.1 Property of the penalty term

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

$$\mathbf{E}_\vartheta\, \mathbf{1}_{\mathbf{G}_*}\, P_n(\lambda_0) \le \frac{1}{1-2\varepsilon}\,\mathbf{E}_\vartheta Err_n(\lambda_0)\mathbf{1}_{\mathbf{G}_*} + \frac{\check{\mathbf{x}}\mathbf{g}^*_T}{\varepsilon(1-2\varepsilon)T} + \frac{2\check{\mathbf{x}}\sigma_{1,*}}{n\varepsilon} + \frac{2\|S\|_n\sqrt{\check{\mathbf{x}}\sigma_{1,*}}}{\sqrt{n}}\sqrt{\mathbf{P}_\vartheta\left(\mathbf{G}^c_*\right)},$$

*where the term* $\mathbf{g}^*_T$ *is given in* (3.6).

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

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

Taking into account here that

$$\zeta^2_{j,n} = g^2_{j,n} + \frac{\check{\mathbf{x}}}{n}\xi^2_{j,n} + 2\sqrt{\frac{\check{\mathbf{x}}}{n}}g_{j,n}\xi_{j,n},$$

we obtain

$$\mathrm{Err}_n(\lambda) \ge \frac{\check{\mathbf{x}}}{n}\sum_{j=1}^n \lambda^2(j)\xi^2_{j,n} + 2\sqrt{\frac{\check{\mathbf{x}}}{n}}I_1 - 2\sqrt{\frac{\check{\mathbf{x}}}{n}}I_2,$$

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 $0 < \varepsilon < 1$,

$$2\sqrt{\frac{\check{\mathbf{x}}}{n}}I_1 \le \frac{1}{\varepsilon}\|g\|^2_n + \frac{\varepsilon\check{\mathbf{x}}}{n}\sum_{j=1}^n \lambda^2(j)\xi^2_{j,n}.$$

Therefore

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

and

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\mathrm{Err}_n(\lambda_0) \geq \frac{(1-\varepsilon)\check{\mathbf{x}}}{n}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\sum_{j=1}^{n}\lambda^2(j)\xi_{j,n}^2 - \frac{2\sqrt{\check{\mathbf{x}}}}{\sqrt{n}}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}I_2 - \frac{1}{\varepsilon}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\|g\|_n^2.$$

Taking into account here the definition of $\mathbf{B}(\cdot)$ in (6.1) and that $\mathbf{E}_\vartheta I_2 = 0$ we can rewrite the last inequality as

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\mathrm{Err}_n(\lambda_0) \geq (1-\varepsilon)\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}P_n(\lambda) + \frac{(1-\varepsilon)}{\sqrt{n}}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\mathbf{B}(\lambda^2) + \frac{2\sqrt{\check{\mathbf{x}}}}{\sqrt{n}}\mathbf{E}_\vartheta \mathbf{1}_{(\mathbf{G}_*)^c}I_2 - \frac{1}{\varepsilon}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\|g\|_n^2.$$

Now Propositions 6.1– 6.2 imply that

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\mathrm{Err}_n(\lambda_0) \geq (1-2\varepsilon)\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}P_n(\lambda) - \frac{2\check{\mathbf{x}}\sigma_{1,*}}{n\varepsilon} - \frac{2\|S\|_n\sqrt{\check{\mathbf{x}}\sigma_{1,*}}}{\sqrt{n}}\sqrt{\mathbf{P}_\vartheta(\mathbf{G}^c)} - \frac{1}{\varepsilon}\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*}\|g\|_n^2.$$

Hence Proposition A.1. □

## A.2 Asymptotic analysis tools

**Proposition A.2** *Assume that the conditions* $\mathbf{A}_1$)–$\mathbf{A}_2$) *hold. Then, for any* $\mathbf{x}_0 < \mathbf{x}_1$ *and* $a > 0$,

$$\lim_{T\to\infty} T^a \max_{\mathbf{x}_0\leq x\leq \mathbf{x}_1}\sup_{\vartheta\in\Theta}\mathbf{P}_\vartheta(|\widetilde{q}_T(x) - \mathbf{q}_\vartheta(x)| > \upsilon_T) = 0.$$

The proof is the same as for Lemma A.3 in Galtchouk and Pergamenshchikov (2015), so it is omitted.

***Proof of Proposition 3.3*** First, note that, to show the limit (3.7) it suffices to check that, for any $a > 0$,

$$\lim_{T\to\infty} T^{1-a}\sup_{\mathbf{x}_0\leq z\leq \mathbf{x}_1}\sup_{\vartheta\in\Theta}\left(\mathbf{E}_\vartheta\,\mathbf{g}_1^2(z)\mathbf{1}_{\Gamma(z)} + \mathbf{E}_\vartheta\,\mathbf{g}_2^2(z)\mathbf{1}_{\Gamma(z)}\right) = 0. \tag{A.1}$$

Indeed, using the definition of $\mathbf{g}_1(z)$ in (2.16) we represent it on the set $\Gamma(z)$ as $\mathbf{g}_1(z) = \mathbf{g}_{1,1}(z) + \mathbf{g}_{1,2}(z)$, where

$$\mathbf{g}_{1,1}(z) = \frac{1}{\delta H(z)}(1-\sqrt{\varkappa(z)})\sqrt{\varkappa(z)}\chi_{\tau(z)}(z,h)\int_{t_{\tau(z)-1}}^{t_{\tau(z)}} S(y_u)\,\mathrm{d}u$$

and

$$\mathbf{g}_{1,2}(z) = \frac{1}{\delta H(z)}\sum_{j=N_0+1}^{\tau(z)}\widetilde{\varkappa}_j(z)\,\chi_j(z,h)\int_{t_{j-1}}^{t_j} S(y_u)\,\mathrm{d}u - S(z).$$

To estimate the term $\mathbf{g}_{1,2}(z)$ note that

$$\mathbf{g}_{1,2}^2(z) \le \frac{\Psi_{\tau(z)}(z)}{\delta H^2(z)}, \quad \Psi_{\tau(z)}(z) = \chi_{\tau(z)}(z,h) \int_{t_{\tau(z)-1}}^{t_{\tau(z)}} S^2(y_u)\,\mathrm{d}u.$$

Moreover, note also here, that for some constant $C > 0$

$$\max_{N_0<j\le N} \max_{t_{j-1}\le u\le t_j} \sup_{\vartheta\in\Theta} \mathbf{E}_\vartheta \left(S^2(y_u)|\mathcal{F}_{t_{j-1}}\right) \le C(1+y_{t_{j-1}}^2).$$

From the definition (2.13) it follows that $\{\tau(z)=j\} \in \mathcal{F}_{t_{j-1}}$, i.e.

$$\mathbf{E}_\vartheta \left(\Psi_j(z)|\mathcal{F}_{t_{j-1}}\right) \le \delta C(1+y_{t_{j-1}}^2)\chi_j(z,h) \le \delta C.$$

Therefore, for some $C > 0$

$$\mathbf{E}_\vartheta \left(\Psi_{\tau(z)}(z)|\mathcal{F}_{t_{N_0}}\right) = \sum_{j=N_0+1}^{N} \mathbf{E}_\vartheta \left(\mathbf{1}_{\{\tau(z)=j\}}\mathbf{E}_\vartheta \left(\Psi_j(z)|\mathcal{F}_{t_{j-1}}\right)|\mathcal{F}_{t_{N_0}}\right) \le \delta C$$

and

$$\mathbf{E}_\vartheta \left(\mathbf{g}_{1,2}^2(z)|\mathcal{F}_{t_{N_0}}\right) \le \frac{C}{H^2(z)}.$$

Using the definition (2.12), the conditions $\mathbf{A}_1$)–$\mathbf{A}_2$), and Propositions 4.1–4.2 from Galtchouk and Pergamenshchikov (2015) we obtain the property (3.7). Hence Proposition 3.3. □

***Proof of Proposition 3.4*** Note that

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

Taking into account the definition of $N_0$ in (2.9) we obtain through Proposition 3.1 from Galtchouk and Pergamenshchikov (2019) the limit equality (3.11). Hence Proposition 3.4. □

Now, we study the heteroscedastic property in the model (3.4). To this end we study asymptotic properties of the average variance $\mathbf{s}_n$ defined in (7.10).

**Proposition A.3** *Assume that the condition* $\mathbf{A}_1$) *holds. Then*

$$\lim_{T\to\infty} \sup_{\vartheta\in\Theta_{k,r}} \mathbf{E}_\vartheta \left|\mathbf{s}_n - \frac{\mathbf{J}_\vartheta}{\check{\mathbf{x}}}\right| = 0. \tag{A.2}$$

***Proof*** Using the definition of $\sigma_l$ in (3.4) and taking into account the form of $h$ given in (2.13), we can represent the term $\mathbf{s}_n$ as

$$\mathbf{s}_n = \frac{1}{\check{\mathbf{x}}} \sum_{l=1}^{n} \widetilde{b}_\vartheta(z_l)(z_l - z_{l-1}) + R_1(\vartheta) + R_2(\vartheta), \tag{A.3}$$

where $\widetilde{b}_\vartheta(x) = b^2(x)/\mathbf{q}_\vartheta(x)$,

$$R_1(\vartheta) = \frac{1}{\check{\mathbf{x}}} \left(\frac{n^2}{\delta\,(N-N_0)} - 1\right) \sum_{l=1}^{n} \widetilde{b}_\vartheta(z_l)\,(z_l - z_{l-1})$$

and

$$R_2(\vartheta) = \sum_{l=1}^{n} \frac{nb^2(x_l)}{\delta\, h\,(N-N_0)} \left( \frac{1}{2\widetilde{q}_T(z_l) - \upsilon_T} - \frac{1}{2\mathbf{q}_\vartheta(z_l)} \right)(z_l - z_{l-1}).$$

First of all note, that the function $\widetilde{b}_\vartheta(\cdot)$ and its derivative are uniformly bounded, i.e. $\sup_{\vartheta\in\Theta_{k,r}} \max_{\mathbf{x}_0\le z\le \mathbf{x}_1} \left(\widetilde{b}_\vartheta(z) + |\widetilde{b}'_\vartheta(z)|\right) < \infty$. Therefore,

$$\lim_{T\to\infty} \sup_{\vartheta\in\Theta_{k,r}} \left| \sum_{l=1}^{n} \widetilde{b}_\vartheta(z_l)(z_l - z_{l-1}) - \mathbf{J}_\vartheta \right| = 0.$$

As the second term (A.3), note that, in view of the condition (3.2),

$$\lim_{T\to\infty} \frac{n^2}{\delta\,(N-N_0)} = 1.$$

Therefore, $\lim_{T\to\infty} \sup_{\vartheta\in\Theta_{k,r}} |R_1(\vartheta)| = 0$. Moreover, taking into account that $2\widetilde{q}_T(z_l) - \upsilon_T > \upsilon_T^{1/2}$, we obtain, for sufficiently large $T$, that for some $C > 0$

$$|R_2(\vartheta)| \le C \left( \sum_{l=1}^{n} \upsilon_T^{-1/2} \left( |\widetilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)| \right)(z_l - z_{l-1}) + \sqrt{\upsilon_T} \right).$$

Note here, that for any $1 \le l \le n$ and for sufficiently large $T$,

$$\upsilon_T^{-1/2} \mathbf{E}_\vartheta\, |\widetilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)| \le 2\upsilon_T^{-1} \mathbf{P}_\vartheta(|\widetilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)| > \upsilon_T) + \sqrt{\upsilon_T}$$

and, therefore, Proposition A.2 implies $\lim_{T\to\infty} \sup_{\vartheta\in\Theta_{k,r}} |R_2(\vartheta)| = 0$. □

**Lemma A.4** *Let $f$ be an absolutely continuous $[\mathbf{x}_0, \mathbf{x}_1] \to \mathbb{R}$ function with $\|\dot{f}\| < \infty$ and $g$ be $[\mathbf{x}_0, \mathbf{x}_1] \to \mathbb{R}$ a step-wise function $g(z) = \sum_{j=1}^{n} c_j\, \chi_{(z_{j-1}, z_j]}(z)$, where $c_j$ are some constants and the sequence $(z_j)_{0\le j\le n}$ is given in (3.1). Then, for any $\widetilde{\varepsilon} > 0$, the function $\Delta = f - g$ satisfies the following inequalities*

$$\frac{1}{\widetilde{\varepsilon}} \frac{\|\dot{f}\|^2}{n^2} \check{\mathbf{x}}^2 - \frac{\|\Delta\|^2}{1+\widetilde{\varepsilon}} \le \|\Delta\|_n^2 \le (1+\widetilde{\varepsilon})\|\Delta\|^2 + \left(1 + \frac{1}{\widetilde{\varepsilon}}\right) \frac{\|\dot{f}\|^2}{n^2} \check{\mathbf{x}}^2.$$

The proof is given in Lemma A.2 from Konev and Pergamenshchikov (2015).

### A.3 Properties of the trigonometric basis

**Lemma A.5** *For any $1 \le j \le n$ and any $\widetilde{\varepsilon} > 0$, the discrete trigonometric Fourier coefficients $(\theta_{j,n})_{1\le j\le n}$ introduced in (4.3) for $S \in W_{k,r}$ are bounded as*

$$\theta_{j,n}^2 \le (1+\widetilde{\varepsilon})\,\theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,\frac{\check{r}_k}{n^{2k}}, \quad \check{r}_k = \frac{2r(\pi^2+1)\check{\mathbf{x}}^{2k}}{\pi^{2k}}, \tag{A.4}$$

*where the coefficients $\theta_j$ are defined in (5.5).*

***Proof*** First we represent the function $S$ in $\mathcal{L}[\mathbf{x}_0, \mathbf{x}_1]$ as

$$S(x) = \sum_{l=1}^{n} \theta_l\, \phi_l(x) + \Delta_n(x) \quad \text{and} \quad \Delta_n(x) = \sum_{l>n} \theta_l\, \phi_l(x). \tag{A.5}$$

Since $\theta_{j,n} = (S, \phi_j)_n = \theta_j + (\Delta_n, \phi_j)_n$, we get, that for any $0 < \widetilde{\varepsilon} < 1$,

$$\theta^2_{j,n} \le (1+\widetilde{\varepsilon})\theta^2_j + (1+\widetilde{\varepsilon}^{-1})\|\Delta_n\|^2_n.$$

Moreover, through Lemma A.4 and the definition (5.5) we deduce

$$\|\Delta_n\|^2_n \le 2\sum_{l>n}\theta^2_l + 2\frac{\|\dot{\Delta}_n\|^2\check{\mathbf{x}}^2}{n^2} \le \frac{2r}{a_{n+1}} + \frac{2\|\dot{\Delta}_n\|^2\check{\mathbf{x}}^2}{n^2}.$$

Taking into account here that $2[l/2] \ge l-1$ for $l \ge 2$, we get

$$\|\Delta_n\|^2_n \le \frac{2r\check{\mathbf{x}}^{2k}}{\pi^{2k}n^{2k}} + 2\frac{\|\dot{\Delta}_n\|^2\check{\mathbf{x}}^2}{n^2}.$$

Similarly, for any $n \ge 1$,

$$\|\dot{\Delta}_n\|^2 = \frac{(2\pi)^2}{\check{\mathbf{x}}^2}\sum_{l>n}\theta^2_l\,[l/2]^2 = \frac{\check{\mathbf{x}}^{2(k-1)}}{\pi^{2(k-1)}}\sum_{l>n}\frac{a_l\theta^2_l}{(2[l/2])^{2(k-1)}}$$
$$\le \frac{\check{\mathbf{x}}^{2(k-1)}}{\pi^{2(k-1)}}\sum_{l>n}\frac{a_l\theta^2_l}{(l-1)^{2(k-1)}} \le \frac{r\check{\mathbf{x}}^{2(k-1)}}{\pi^{2(k-1)}n^{2(k-1)}}. \tag{A.6}$$

Hence Lemma A.5. □

**Lemma A.6** *For any $n \ge 2$, $1 \le m < n$ and $r > 0$, the coefficients $(\theta_{j,n})_{1\le j\le n}$ of functions $S$ from the class $W_{k,r}$ satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality*

$$\sum_{j=m+1}^{n}\theta^2_{j,n} \le (1+\widetilde{\varepsilon})\sum_{j\ge m+1}\theta^2_j + (1+\widetilde{\varepsilon}^{-1})\frac{\check{r}_1}{n^2m^{2(k-1)}}, \tag{A.7}$$

*where $\check{r}_1 = r\check{\mathbf{x}}^{2k}/\pi^{2(k-1)}$.*

***Proof*** First we note that

$$\sum_{j=m+1}^{n}\theta^2_{j,n} = \min_{x_1,\ldots,x_m}\|S - \sum_{j=1}^{m}x_j\phi_j\|^2_n \le \|\Delta_m\|^2_n,$$

where the function $\Delta_m(\cdot)$ is defined in (A.5). By applying Lemma A.4 with $f = \Delta_m$, $g = 0$, and taking into account the inequality (A.6), we obtain the bound (A.7). Hence Lemma A.6 □

**Lemma A.7** *For any $k \ge 1$,*

$$\sup_{n\ge 2} n^{-k}\sup_{x\in[\mathbf{x}_0,\mathbf{x}_1]}\left|\sum_{l=2}^{n} l^k\overline{\phi}_l(x)\right| \le 2^k, \tag{A.8}$$

*where $\overline{\phi}_l(x) = \check{\mathbf{x}}\phi^2_l(x) - 1$.*

Proof of this result is given in Lemma A.2 from Galtchouk and Pergamenshchikov (2009a).

## References


Bayisa FL, Zhou Z, Cronie O, Yu J (2019) Adaptive algorithm for sparse signal recovery. Digit Signal Proc 87:10–18

Comte F, Genon-Catalot V, Rozenholc Y (2009) Non-parametric estimation for a discretely observed integrated diffusion model. Stoch Process Appl 119:811–834

Dalalyan AS (2005) Sharp adaptive estimation of the drift function for ergodic diffusion. Ann Stat 33(6):2507–2528

Dalalyan AS, Kutoyants YuA (2002) Asymptotically efficient trend coefficient estimation for ergodic diffusion. Math Methods Stat 11(4):402–427

Fan J, Fan Y, Barut E (2014) Adaptive robust variable selection. Ann Stat 42(1):324–351

Florens-Zmirou D (1993) On estimating the diffusion coefficient from discrete observations. J Appl Probab 30(4):790–804

Fujimori K (2019) The Danzing selector for a linear model of diffusion processes. Stat Infer Stoch Process 22:475–498

De Gregorio A, Iacus SM (2012) Adaptive LASSO-type estimation for multivariate diffusion processes. Econ Theory 28(4):838–860

Galtchouk LI (1978) Existence and uniqueness of a solution for stochastic equations with respect to semimartingales. Theory Probab Appl 23:751–763

Galtchouk L, Pergamenshchikov S (2001) Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. Math Methods Stat 10(3):316–330

Galtchouk L, Pergamenshchikov S (2004) Nonparametric sequential estimation of the drift in diffusion via model selection. Math Methods Stat 13:25–49

Galtchouk L, Pergamenshchikov S (2005) Nonparametric sequential minimax estimation of the drift coefficient in diffusion processes. Sequ Anal 24(3):303–330

Galtchouk L, Pergamenshchikov S (2006) Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. Stat Infer Stoch Process 9:1–16

Galtchouk L, Pergamenshchikov S (2007) Uniform concentration inequality for ergodic diffusion processes. Stoch Process Appl 117:830–839

Galtchouk L, Pergamenshchikov S (2009a) Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. J Nonparametr Stat 21:1–16

Galtchouk L, Pergamenshchikov S (2009b) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. J Korean Stat Soc 38(4):305–322

Galtchouk L, Pergamenshchikov S (2011) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. J Nonparametr Stat 23(2):255–285

Galtchouk L, Pergamenshchikov S (2013) Uniform concentration inequality for ergodic diffusion processes observed at discrete times. Stoch Process Appl 123(1):91–109

Galtchouk L, Pergamenshchikov S (2014) Geometric ergodicity for classes of homogeneous Markov chains. Stoch Process Appl 124(10):3362–3391

Galtchouk L, Pergamenshchikov S (2015) Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions. Bernoulli 21(4):2569–2594

Galtchouk L, Pergamenshchikov S (2019) Non asymptotic sharp oracle inequalities for high dimensional ergodic diffusion models. Preprint https://hal.archives-ouvertes.fr/hal-02387034 (2019)

Gihman II, Skorohod AV (1968) Stochastic differential equations. Naukova Dumka, Kiev

Gobet E, Hoffmann M, Reiss M (2004) Nonparametric estimation of scalar diffusions based on low frequency data. Ann Stat 32(5):2223–2253

Hoffmann M (1999) Adaptive estimation in diffusion processes. Stoch Process Appl 79:135–163

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Data mining, inference and prediction. Springer series in statistics. Springer, Berlin

Jacod J (2000) Non-parametric kernel estimation of the coefficient of a diffusion. Scand J Stat 27(1):83–96

Kabanov YM, Pergamenshchikov SM (2003) Two scale stochastic systems: asymptotic analysis and control. Applications of mathematics, stochastic modelling and applied probability, vol 49. Springer Berlin

Karatzas I, Shreve SE (1998a) Brownian motion and stochastic calculus. Graduate texts in mathematics, 2nd edn. Springer Science+Business Media, New York

Karatzas I, Shreve SE (1998b) Methods of mathematical finance. Springer, New York

Kutoyants YuA (1977) Estimation of the signal parameter in a Gaussian noise. Probl Inf Transm 13(4):29–38

Kutoyants YuA (1984a) Parameter estimation for stochastic processes. Heldeman-Verlag, Berlin

Kutoyants YuA (1984b) On nonparametric estimation of trend coefficients in a diffusion process. Statistics and control of stochastic processes. Steklov seminar proceedings, Moscow, pp 230–250

Kutoyants YuA (2003) Statistical inferences for ergodic diffusion processes. Springer, Berlin

Konev VV, Pergamenshchikov SM (2015) Robust model selection for a semimartingale continuous time regression from discrete data. Stoch Process Appl 125:294–326

Lamberton D, Lapeyre B (1996) Introduction to stochastic calculus applied to finance. Chapman & Hall, London

Pinsker MS (1981) Optimal filtration of square integrable signals in gaussian noise. Probl Inf Transm 17:120–133