Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions

Leonid Galtchouk, Serguei Pergamenchtchikov

Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions. *

L.I. Galtchouk † S.M. Pergamenshchikov‡

Abstract

A truncated sequential procedure is constructed for estimating the drift coefficient at a given state point based on discrete data of ergodic diffusion process. A nonasymptotic upper bound is obtained for a pointwise absolute error risk. The optimal convergence rate and a sharp constant in the bounds are found for the asymptotic pointwise minimax risk. As a consequence, the efficiency is obtained of the proposed sequential procedure.

Key words: drift coefficient estimation, discrete data, efficient procedure, ergodic diffusion process, minimax, nonparametric sequential estimation.

AMS (1991) Subject Classification : primary 62G07; secondary 62G20

*The second author is partially supported by the RFBR-Grant 09-01-00172-a.

†IRMA, Strasbourg University 7, rue Rene Descartes, 67084, Strasbourg, France, e-mail: leonid.galtchouk@math.unistra.fr

‡Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France and Department of Mathematics and Mechanics,Tomsk State University, Lenin str. 36, 634041 Tomsk, Russia, e-mail: Serge.Pergamenchtchikov@univ-rouen.fr

1 Introduction

In this paper we consider the following diffusion model:

$$\mathrm{d}y_t = S(y_t)\,\mathrm{d}t + \sigma(y_t)\mathrm{d}W_t\,, \quad 0 \le t \le T\,, \tag{1.1}$$

where $(W_t)_{t\ge 0}$ is a scalar standard Wiener process, $S(\cdot)$ and $\sigma(\cdot)$ are unknown functions. This model appears in a number of applied problems of stochastic control, filtering, spectral analysis, identification of dynamic system, financial mathematics and others (see [1], [3], [23], [27], [28] and others for details).

The problem is to estimate the function $S(x)$ at a point x_0 based on the discrete time observations

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

where $N = [T/\delta]$ and the frequency $\delta = \delta_T \in (0,1)$ is a function of T that will be specified later.

The estimation problem of the function S was studied in a number of papers in the case of complete observations, that is when a continuous trajectory $(y_t)_{0\le t\le T}$ was observed. In the parametric case this problem was considered apparently for the first time in the paper [2] for diffusion model of the axis of the equator precession. In that paper a non-asymptotic distribution of the maximum likelihood estimator was found for a special Ornstein-Uhlenbeck process.

It should be noted that investigating non asymptotic properties of parametric estimators in the models like to (1.1) comes to the analysis of non linear functionals of observations. At the most cases this analysis is unproductive in non asymptotic setting. In order to overcome the technique difficulties the sequential analysis methods were used in [28] and [29] for estimating a scalar parameter. In [25] these methods were extended to estimating a multi-dimensional parameter as well. Moreover, in [26] truncated sequential procedures were developed that economizes the observation time.

In [7] and [8] a sequential approach was proposed for the pointwise nonparametric estimation in the ergodic models (1.1). Later in [10] the efficiency was studied of the proposed sequential procedures .

A sufficiently complete survey one can find in [27] on the nonparametric estimation in the ergodic model (1.1) when non sequential approaches are used.

In the cited papers estimation problems were studied based on complete observations $(y_t)_{0\le t\le T}$. In practice, usually one has at disposal discrete time observations even for continuous time models.

A natural question arises about proprieties and the behavior of estimates based on discrete time observations for such models. These problems were studied for several models. We cite some of them.

The LAN property was studied in the papers [17], [24] and [30] for parametric ergodic diffusion models observed at discrete times. Parametric estimation was investigated in [21] for a non ergodic diffusion. Nonparametric estimation setting for models of kind (1.1) was considered firstly for estimating the unknown diffusion coefficient $\sigma(\cdot)$ based on discrete time observations on the interval $[0,1]$ (see, for example, [6], [15], [19], [20] and the references therein). Later, in [18] kernel estimates of drift and diffusion coefficients were studied for non ergodic process (1.1) taking the values into the interval $[0,1]$.

So far as concerning the estimation in ergodic case, it should be noted that a sequential procedure was proposed in [19] for nonparametric estimating the drift coefficient of the process (1.1) in the integral metric. Some upper and lower asymptotic bounds were found for the $\mathbf{L}_p-$risks. Later, in the paper [5] a nonasymptotic oracle inequality was proved for the drift coefficient estimation problem in a special empiric quadratic risk based on discrete time observations.

This paper deals with the drift coefficient efficient nonparametric estimating at a given state point based on discrete time observations (1.2) in the absolute error risk. We find the optimal minimax convergence rate and we study the lower bound normalized by this convergence rate.

Our approach is based on the sequential analysis and it was developed in the papers [7], [8], and [10] for the nonparametric estimation.

Let us remind that in the case of complete observations (that is, when a whole trajectory is observed) the sequential estimate efficiency was proved by making use of uniform concentration inequalities (see [11]), besides the weak Hölder space functions S were used.

As it turns out later in [13], the efficient kernel estimate in the above given sense provides to construct a selection model adaptive procedure that appears efficient in the quadratic $\mathbf{L}_2-$metric.

Therefore, in order to realize this program in the case of discrete time observations, one needs to obtain the relative concentration inequalities, that is done in [14]. It should be noted that to obtain nonasymptotic concentration inequalities we make use of nonasymptotic bounds uniform over functions S and σ for the convergence rate in the ergodic theorem for the process (1.1). The latest result is proved in [12] and it based on a new approach using Lyapounov's functions and the coupling method.

Further in this paper, by making use of the concentration inequalities we find the explicit constant in the upper bound for weak Hölder's risk normalized by the optimal convergence rate and we prove that this upper bound is

best over all possible estimates. It means the procedure is efficient.

The paper is organized as follows. In Section 2 we describe the functional classes. In Section 3 the sequential procedure is constructed. In Section 4 we obtain a nonasymptotic upper bound for the absolute error pointwise risk of the sequential procedure. In Section 5 we show that the proposed procedure is asymptotically efficient for the pointwise risk. All proofs are given in Section 6. In the Appendix we give all necessary technical results.

2 Functional class

We consider the pointwise estimation problem for the function $S(\cdot)$ at a fixed point $x_0 \in \mathbb{R}$ for the model (1.1) with unknown diffusion coefficient σ. It is clear that to obtain a good estimate for the function $S(\cdot)$ at the point x_0 it is necessary to impose some conditions on the function $\vartheta = (S, \sigma)$ which provide that the observed process $(y_t)_{0\le t\le T}$ returns to any vicinity of the point x_0 infinitely many times.

In this Section we describe the weak Hölder functional class which guarantees the ergodicity property for this model. Firstly, for some $\mathbf{x}_* \ge |x_0| + 1$, $M > 0$ and $L > 1$ we denote by $\Sigma_{L,M}$ the class of functions S from $\mathbf{C}^1(\mathbb{R})$ such that

$$\sup_{|x|\le \mathbf{x}_*} \Big(|S(x)| + |\dot S(x)| \Big) \le M$$

and

$$-L \le \inf_{|x|\ge \mathbf{x}_*} \dot S(x) \le \sup_{|x|\ge \mathbf{x}_*} \dot S(x) \le -L^{-1} .$$

Moreover, for some fixed parameters $0 < \sigma_{min} \le \sigma_{max}$ we denote by $\mathcal{V}$ the class of the functions σ from $\mathbf{C}^2(\mathbb{R})$ such that

$$\begin{aligned} 0 < \sigma_{min} &\le \inf_{x\in\mathbb{R}} \min \left(|\sigma(x)| , |\dot\sigma(x)| , |\ddot\sigma(x)| \right) \\ &\le \sup_{x\in\mathbb{R}} \max \left(|\sigma(x)| , |\dot\sigma(x)| , |\ddot\sigma(x)| \right) \le \sigma_{max} < \infty . \end{aligned} \tag{2.1}$$

In this paper we make use of the week Hölder functions introduced in [9].

Definition 2.1. *We say that a function S satisfies the weak Hölder condition at the point $x_0 \in \mathbb{R}$ with the parameters $h,\epsilon > 0$ and exponent $\beta = 1 + \alpha$, $\alpha \in (0, 1)$, if $S \in \mathbf{C}^1(\mathbb{R})$ and its derivative satisfies the following inequality*

$$\left| \int_{-1}^{1} z \int_0^1 (\dot S(x_0 + uzh) - \dot S(x_0)) \, \mathrm{d}u \, \mathrm{d}z \right| \le \epsilon \, h^\alpha . \tag{2.2}$$

We will denote the set of all such functions by $\mathcal{H}^w_{x_0}(\epsilon, \beta, h)$.

Note that the inequality (2.2) implies that

$$\sup_{S\in\mathcal{H}^w_{x_0}(\epsilon,\beta,h)}\left|\Omega_{x_0,h}(S)\right| \le \epsilon\, h^{\beta}\,, \tag{2.3}$$

where

$$\Omega_{x_0,h}(S)=\int_{-1}^{1}\left(S(x_0+hz)-S(x_0)\right)\mathrm{d}z\,.$$

Let us denote by $\mathcal{H}^w_{x_0,M}(\epsilon,\beta,h)$ the set of all functions D from $\mathcal{H}^w_{x_0}(\epsilon,\beta,h)$ such that $\sup_{x\in\mathbb{R}}(|D(x)|+|\dot{D}(x)|)\le M/2$ and $D(x)=0$ for $|x|\ge x_*$.
Let S_0 be a function from $\Sigma_{L,M/2}$ such that

$$\lim_{h\to 0} h^{-\beta}\,\Omega_{x_0,h}(S_0)=0\,. \tag{2.4}$$

We denote

$$\mathcal{U}_M(x_0,\beta)=S_0+\mathcal{H}^w_{x_0,M}(\epsilon,\beta,h)\,, \tag{2.5}$$

where $h=T^{-1/(2\beta+1)}$ and

$$\varepsilon=\varepsilon_T=\frac{1}{(\ln T)^{1+\gamma}} \tag{2.6}$$

for some $0<\gamma<1$. Obviously that $\mathcal{U}_M(x_0,\beta)\subset\Sigma_{L,M}$. Now we set

$$\Theta_\beta=\mathcal{U}_M(x_0,\beta)\times\mathcal{V}\,. \tag{2.7}$$

It should be noted that, for any $\vartheta\in\Theta_\beta$, there exists the invariant density which is defined as

$$q_\vartheta(x)=\left(\int_{\mathbb{R}}\sigma^{-2}(z)\,e^{\widetilde{S}(z)}\mathrm{d}z\right)^{-1}\sigma^{-2}(x)\,e^{\widetilde{S}(x)}\,, \tag{2.8}$$

where $\widetilde{S}(x)=2\int_0^x\sigma^{-2}(v)S(v)\mathrm{d}v$ (see,e.g., [16], Ch.4, 18, Th2). It is easy to see that this density is uniformly bounded in the class (2.7) , i.e.

$$q^*=\sup_{x\in\mathbb{R}}\sup_{\vartheta\in\Theta_\beta}q_\vartheta(x)<+\infty \tag{2.9}$$

and bounded away from zero on the interval $[x_0-1,x_0+1]$, i.e.

$$q_*=\inf_{x_0-1\le x\le x_0+1}\inf_{\vartheta\in\Theta_\beta}q_\vartheta(x)>0\,. \tag{2.10}$$

For any $\mathbb{R} \to \mathbb{R}$ function f from $\mathbf{L}_1(\mathbb{R})$ we set

$$\mathbf{m}_\vartheta(f) = \int_{\mathbb{R}} f(x)\, q_\vartheta(x)\, \mathrm{d}x\,. \tag{2.11}$$

Assume that the frequency δ in the observations (1.2) is of the following asymptotic form (as $T \to \infty$)

$$\delta = \delta_T = \mathrm{O}\left(\frac{\varepsilon_T}{T}\right), \tag{2.12}$$

where the function ε_T is introduced in (2.6).

Now, for any estimate (i.e. any $(y_t)_{0\le t\le T}$ measurable function) $\tilde{S}_T(x_0)$ of $S(x_0)$, we define the pointwise risk as follows

$$\mathcal{R}_\vartheta(\widetilde{S}_T) = \mathbf{E}_\vartheta\, |\widetilde{S}_T(x_0) - S(x_0)|\,. \tag{2.13}$$

3 Sequential procedure

In order to construct an efficient pointwise estimate of S we begin with estimating the ergodic density $q = q_\vartheta$ at the point x_0 from first N_0 observations. We choose

$$N_0 = N^{\gamma_0} \quad \text{and} \quad 2/3 < \gamma_0 < 1\,. \tag{3.1}$$

We will make use of the following kernel estimate

$$\widehat{q}_T(x_0) = \frac{1}{2(N_0-1)\varsigma} \sum_{j=0}^{N_0-1} Q\left(\frac{y_{t_j} - x_0}{\varsigma}\right), \tag{3.2}$$

where $Q(y) = \mathbf{1}_{(|y|\le 1)}$ and $\varsigma = \varsigma_T$ is a function of T such that

$$\varsigma_T = \mathrm{o}(T^{-\gamma_0/2}) \quad \text{as} \quad T \to \infty\,.$$

For $T \ge 3$ we set

$$\widetilde{q}_T(x_0) = \begin{cases} (\upsilon_T)^{1/2}\,, & \text{if} \quad \widehat{q}_T(x_0) < (\upsilon_T)^{1/2}\,; \\ \widehat{q}_T(x_0)\,, & \text{if} \quad (\upsilon_T)^{1/2} \le \widehat{q}_T(x_0) \le (\upsilon_T)^{-1/2}\,; \\ (\upsilon_T)^{-1/2}\,, & \text{if} \quad \widehat{q}_T(x_0) > (\upsilon_T)^{-1/2}\,, \end{cases} \tag{3.3}$$

where

$$\upsilon_T = \frac{1}{(\ln T)^{a_0}} \quad \text{and} \quad a_0 = \frac{\sqrt{\gamma+1}-1}{10}\,.$$

The properties of the estimates $\widehat{q}_T(x_0)$ and $\widetilde{q}_T(x_0)$ are studied in the Appendix. Let us define the following stopping time

$$\varpi = \varpi_T = \inf\left\{ j \geq N_0 \,:\, \sum_{i=N_0}^{j} \phi_i \,\geq H_T \right\}, \tag{3.4}$$

where $\phi_i = \chi_{h,x_0}(y_{t_{i-1}})\mathbf{1}_{\{i\leq N\}} + \mathbf{1}_{\{i>N\}}$, $\chi_{h,x_0}(y) = Q\left((y-x_0)/h\right)$ and h is a positive bandwidth. We put $\varpi = \infty$ if the set $\{\cdot\}$ is empty. Obviously, that in our case $\varpi < \infty$ a.s. since $\sum_{i\geq N_0} \phi_i = +\infty$ a.s.

Now we have to choose the threshold H_T. Note that in order to construct an efficient estimate one should use all, i.e. N, observations. Therefore, the threshold H_T should provide the asymptotic relations $\varpi_T \approx N$ and

$$\sum_{i=N_0}^{N} \phi_i = \sum_{i=N_0}^{N} \chi_{h,x_0}(y_{t_{i-1}}) \quad \text{as} \quad T\to\infty\,.$$

In order to obtain these relations, note that due to the ergodic theorem

$$\sum_{i=N_0}^{N} \chi_{h,x_0}(y_{t_{i-1}}) \approx 2h(N-N_0)q_T(x_0)\,.$$

Hence, replacing in the right-hand side term the ergodic density with its corrected estimate yields the following definition of the threshold

$$H = H_T = h(N-N_0)(2\widetilde{q}_T(x_0) - \upsilon_T)\,. \tag{3.5}$$

Note that in [7] it has been shown that the such form of the threshold H_T provides the optimal convergence rate. It is clear that

$$\varpi \leq N + H_T < N + h(N-N_0)/\sqrt{\upsilon_T}\,,$$

i.e. the stopping time ϖ is bounded.

Now on the set $\Gamma_T = \{\varpi \leq N\}$ we define the correction coefficient $\alpha = \alpha_T$ as

$$\alpha_T = \frac{H_T - \sum_{j=N_0}^{\varpi-1} \chi_{h,x_0}(y_{t_{j-1}})}{\chi_{h,x_0}(y_{t_{\varpi-1}})}\,,$$

i.e. on the set Γ_T

$$\sum_{j=N_0}^{\varpi-1} \chi_{h,x_0}(y_{t_{j-1}}) + \alpha\chi_{h,x_0}(y_{t_{\varpi-1}}) = H_T\,.$$

Moreover, on the Γ_T^c we set $\alpha_T = 1$. Using this definition we introduce the weight sequence

$$\widetilde{\alpha}_j = \mathbf{1}_{\{j<\varpi\}} + \alpha \mathbf{1}_{\{j=\varpi\}} \,. \tag{3.6}$$

One can check directly that, for any $j \geq 1$, the coefficients $\widetilde{\alpha}_j$ are $\mathcal{F}_{t_{j-1}}$ measurable, where $\mathcal{F}_t = \sigma\,(y_u\,,\, 0 \leq u \leq t)$. Now we define the sequential estimate for $S(x_0)$ as

$$S^*_{h,T}(x_0) = \frac{1}{\delta H_T} \left(\sum_{j=N_0}^{N} \widetilde{\alpha}_j \chi_{h,x_0}(y_{t_{j-1}})\, \Delta y_{t_j} \right) \mathbf{1}_{\Gamma_T} \,. \tag{3.7}$$

In the next section we study the non asymptotic properties of this procedure.

Remark 3.1. *Note that the correction coefficient of type* (3.6) *was used firstly in the paper [4] in order to construct an unbiased estimate of a scalar parameter in autoregressive processes AR(1). Here we make use of the same idea but for a nonparametric procedure.*

4 Non asymptotic estimation

As we will see later in studying the estimate (3.7), the approximation term plays the crucial role. In our case, this term is of the following form

$$\Upsilon_{1,T} = \frac{1}{\delta H_T} \sum_{j=N_0}^{N} \widetilde{\alpha}_j \chi_{h,x_0}(y_{t_{j-1}})\, \varrho_j \,, \tag{4.1}$$

where $\varrho_j = \int_{t_{j-1}}^{t_j} (S(y_u) - S(y_{t_{j-1}}))\mathrm{d}u$. One can show the following result.

Proposition 4.1. *For any $T \geq 3$,*

$$\sup_{\vartheta \in \Sigma_{L,M} \times [0,\sigma_{max}]} \mathbf{E}_\vartheta\, \Upsilon^2_{1,T} \leq \widetilde{L}^2 L_1\, \delta \,, \tag{4.2}$$

where $\widetilde{L} = \max(L, M)$ and $L_1 = 2\left(\sigma^2_{max} + 2\delta(M^2 + L^3 D_* + L^2\, x^2_*)\right)$.

Moreover, we set

$$\Upsilon_{2,T} = \frac{1}{\delta H_T} \sum_{j=N_0}^{N} \widetilde{\alpha}_j \chi_{h,x_0}(y_{t_{j-1}})\, \varrho^*_j \,, \tag{4.3}$$

where $\varrho^*_j = \int_{t_{j-1}}^{t_j} (\sigma(y_u) - \sigma(y_{t_{j-1}}))\mathrm{d}W_u$.

Proposition 4.2. *For any $T \geq 3$ for which $0 < \delta \leq 1$, one has*

$$\sup_{\vartheta \in \Sigma_{L,M} \times [0, \sigma_{max}]} \mathbf{E}_\vartheta \left(\Upsilon_{2,T} \right)^2 \leq \frac{\sigma^2_{max} L_1}{h(N - N_0)\sqrt{\upsilon_T}} \,. \tag{4.4}$$

Proofs of Propositions 4.1–4.2 are given in the Appendix.
Now we introduce the approximation term, i.e.

$$B_T = \frac{1}{H_T} \sum_{j=N_0}^{N} \widetilde{\alpha}_j f_h(y_{t_{j-1}}) \tag{4.5}$$

with $f_h(y) = \chi_{h,x_0}(y)(S(y) - S(x_0))$. Taking into account this formula, we can represent the error of estimate (3.7) on the set Γ_T as

$$S^*_{h,T}(x_0) - S(x_0) = \Upsilon_{1,T} + B_T + \mathbf{M}_T \,, \tag{4.6}$$

where

$$\mathbf{M}_T = \frac{1}{\delta H_T} \left(\sum_{j=N_0}^{N} \widetilde{\alpha}_j \chi_{h,x_0}(y_{t_{j-1}}) \eta_j \right)$$

with $\eta_j = \int_{t_{j-1}}^{t_j} \sigma(y_u) \, \mathrm{d}W_u$. Obviously, for any function S from $\Sigma_{L,M}$, the term B_T can be bounded as

$$|B_T| \leq h \max_{|x - x_0| \leq h} |\dot{S}(x)| \leq M \, h \,.$$

Proposition 4.3. *For any $T \geq 3$, one has*

$$\sup_{\vartheta \in \Sigma_{L,M} \times [0, \sigma_{max}]} \mathbf{E}_\vartheta \, \mathbf{M}^2_T \leq \frac{\sigma^2_{\max}}{\delta h (N - N_0)\sqrt{\upsilon_T}} \,. \tag{4.7}$$

Hence, we obtain the following upper bound.

Theorem 4.4. *For any $h > 0$ and $T \geq 3$, one has*

$$\sup_{\vartheta \in \Sigma_{L,M} \times [\sigma_{min}, \sigma_{max}]} \mathbf{E}_\vartheta \, |S^*_{h,T}(x_0) - S(x_0)| \leq U^*(\delta, h, T) + M \, \Pi^*_T \,, \tag{4.8}$$

where

$$U^*(\delta, h, T) = \widetilde{L} \sqrt{\delta L_1} + M \, h + \frac{\sigma_{max}}{\sqrt{\delta \, h \, (N - N_0)} \upsilon_T^{1/4}}$$

and

$$\Pi^*_T = \sup_{\vartheta \in \Sigma_{L,M} \times [\sigma_{min}, \sigma_{max}]} \mathbf{P}_\vartheta \left(\Gamma^{\mathrm{c}}_T \right) \,.$$

Let us study now the last term in (4.8).

Proposition 4.5. *Assume that the parameter δ is of the following form* (2.12) *and $h \geq T^{-1/2}$. Then, for any $a > 0$,*

$$\lim_{T\to\infty} T^a \, \Pi^*_T = 0 \,. \tag{4.9}$$

Proof of this proposition is given in the Appendix.

5 Asymptotic efficiency

First of all we study a lower bound for the risk (2.13). To this end we set

$$\varsigma^*_\vartheta(x_0) = \frac{2q_\vartheta(x_0)}{\sigma(x_0)} \,. \tag{5.1}$$

This parameter provides a sharp asymptotic lower bound for the pointwise risk normalized by the minimax rate $\varphi_T = T^{\beta/(2\beta+1)}$.

Theorem 5.1. *The risk defined in* (2.13) *admits the following lower bound*

$$\underline{\lim}_{T\to\infty} \varphi_T \inf_{\widetilde{S}_T} \sup_{\vartheta\in\Theta_\beta} \sqrt{\varsigma^*_\vartheta(x_0)} \, \mathcal{R}_\vartheta(\widetilde{S}_T) \geq \mathbf{E}|\xi| \,, \tag{5.2}$$

where infimum is taken over all possible estimate $\widetilde{S}_T$, ξ is a $(0,1)$ gaussian random variable.

Theorem 5.2. *The kernel estimate $S^*_{h,T}$ defined in* (3.7) *with $h = T^{-1/(2\beta+1)}$ satisfies the following asymptotic inequality*

$$\overline{\lim}_{T\to\infty} \varphi_T \sup_{\vartheta\in\Theta_\beta} \sqrt{\varsigma^*_\vartheta(x_0)} \, \mathcal{R}_\vartheta(S^*_{h,T}) \leq \mathbf{E}|\xi| \,.$$

where ξ is a $(0,1)$ gaussian random variable.

Notice that the Theorem 5.1 and Theorem 5.2 imply the following efficiency property

Theorem 5.3. *The sequential procedure* (3.7) *with $h = T^{-1/(2\beta+1)}$ is asymptotically efficient in the following sense*

$$\overline{\lim}_{T\to\infty} \varphi_T \sup_{\vartheta\in\Theta_\beta} \sqrt{\varsigma^*_\vartheta(x_0)} \, \mathcal{R}_\vartheta(S^*_{h,T}) = \underline{\lim}_{T\to\infty} \varphi_T \inf_{\widetilde{S}_T} \sup_{\vartheta\in\Theta_\beta} \sqrt{\varsigma^*_\vartheta(x_0)} \, \mathcal{R}_\vartheta(\widetilde{S}_T) \,,$$

where infimum is taken over all possible estimate $\widetilde{S}_T$.

Remark 5.1. *Note that the constant* (5.1) *provides the sharp lower bound for the minimax pointwise risk. The calculation of this constant is possible by making use of the weak Hölder class. This functional class was introduced in [9] for regression models. For the first time, the constant* (5.1) *was obtained in the paper [10] at the pointwise estimation problem of the drift based on continuous time observations of the process* (1.1) *with the unit diffusion. Later this constant was used in the paper [13] to obtain the Pinsker constant for a quadratic risk at the adaptive estimation problem of the drift in the model* (1.1) *based on continuous time observations.*

Remark 5.2. *Note also that in this paper the efficient procedure is constructed when the regularity is known of the function to be estimated. In the case of unknown regularity we shall use an approach based on the model selection similarly to that in the paper [13] which deals with continuous time observations. The announced result will be published in the next paper which is in the work.*

6 Proofs

6.1 Lower bound

In this section the Theorem 5.1 will be proved. Let us introduce the model (1.1) with $\sigma = 1$, i.e.

$$\mathrm{d}y_t = S(y_t)\,\mathrm{d}t + \mathrm{d}W_t\,. \tag{6.1}$$

Now we define the risk corresponding to this model as follows

$$\mathcal{R}^*_S(\widetilde{S}_T) = \mathbf{E}_S\,|\widetilde{S}_T(x_0) - S(x_0)|\,, \tag{6.2}$$

where $\mathbf{E}_S$ denotes the expectation with respect to the distribution $\mathbf{P}_S$ of the process (6.1) in the space of continuous functions $\mathbf{C}[0,T]$. It is clear that

$$\sup_{\vartheta\in\Theta_\beta}\sqrt{\varsigma^*_\vartheta(x_0)}\,\mathcal{R}_\vartheta(\widetilde{S}_T) \geq \sup_{S\in\mathcal{U}_M(x_0,\beta)}\sqrt{2q_S(x_0)}\,\mathcal{R}^*_S(\widetilde{S}_T)\,, \tag{6.3}$$

where q_S is the invariant density for the process (6.1) which equals to q_ϑ with $\sigma = 1$. Let now g be a continuously differentiable probability density on the interval $[-1,1]$. Then, for any $u\in\mathbb{R}$ and $0<\nu<1/4$, we set

$$S_{u,\nu}(x) = S_0(x) + \frac{u}{\varphi_T}\,V_\nu\left(\frac{x-x_0}{h}\right)\,,$$

where $h = T^{-1/(2\beta+1)}$ and

$$V_\nu(x) = \frac{1}{\nu}\int_{-\infty}^{\infty}\left(\mathbf{1}_{(|u|\le 1-2\nu)} + 2\mathbf{1}_{(1-2\nu\le|u|\le 1-\nu)}\right) g\left(\frac{u-x}{\nu}\right)\mathrm{d}u\,.$$

It is easy to see directly that, for any $0 < \nu < 1/4$,

$$V_\nu(0) = 1 \quad \text{and} \quad \int_{-1}^{1} V_\nu(x)\,\mathrm{d}x = 2\,.$$

Therefore, denoting $D(x) = S_{u,\nu}(x) - S_0(x)$, we obtain

$$\int_{-1}^{1}(D(x_0+hz) - D(x_0))\mathrm{d}z = 0\,.$$

Moreover, note that

$$|\dot{D}(x)| = |u|\,\varphi_T^{-1}h^{-1}|\dot{V}_\nu\left(\frac{x-x_0}{h}\right)| \le |u|\,T^{-\alpha/(2\beta+1)}\nu^{-2}\dot{g}^*\,,$$

where $\dot{g}^* = \sup_x|\dot{g}(x)|$. Therefore, for any fixed $u \in \mathbb{R}$ and $0 < \nu < 1/4$, the functions $S_{u,\nu}$ belong to the class $\mathcal{U}_M(x_0,\beta)$ for sufficiently large T. It implies that, for any $b > 0$ and for sufficiently large T, we can estimate from below the right-hand term in the inequality (6.3) as

$$\sup_{S\in\mathcal{U}_M(x_0,\beta)}\sqrt{2q_S(x_0)}\,\mathcal{R}^*_S(\widetilde{S}_T) \ge \sup_{|u|\le b}\sqrt{2q_{S_u}(x_0)}\,\mathcal{R}^*_{S_u}(\widetilde{S}_T)\,.$$

Taking into account here that

$$\lim_{T\to\infty}\sup_{|u|\le b}\left|q_{S_u}(x_0) - q_{S_0}(x_0)\right| = 0\,,$$

we obtain the inequality (5.2) by making use of the Theorem 4.1 from [10]. Thus we obtain the Theorem 5.1. □

6.2 Upper bound

We begin with stating the following result for the term (4.5).

Proposition 6.1. *The function B_T defined in* (4.5) *satisfies the following asymptotic property*

$$\limsup_{T\to\infty}\varphi_T\sup_{\vartheta\in\Theta_\beta}\mathbf{E}_\vartheta\,|B_T| = 0\,. \tag{6.4}$$

The result is proved in the Appendix.
Now we prove Theorem 5.2. To this end we set

$$\widetilde{\phi}(u) = \sum_{j=N_0}^{+\infty} \phi_i \, \mathbf{1}_{\left\{t_{i-1}<u\le t_i\right\}} ,$$

where the random variables $(\phi_i)_{i\ge 1}$ are defined in (3.4). Using this function we introduce the stopping time

$$\tau = \tau_T = \inf\left\{ t \ge T_0 \, : \, \int_{T_0}^{t} \widetilde{\phi}(u)\,\mathrm{d}u \ge \delta H_T \right\} ,$$

where $T_0 = t_{N_0} = \delta N_0$. As usually, we put $\tau = \infty$ if the set $\{\cdot\}$ is empty. Obviously that

$$\tau \le T + \delta H \le T + \delta h (N - N_0)/\sqrt{v_T} \, .$$

Due to the equality $\int_{T_0}^{\infty} \widetilde{\phi}(u)\,\mathrm{d}u = \infty$, we obtain immediately that the random variable

$$\xi_T = \frac{1}{\sqrt{\delta H_T}} \int_{T_0}^{\tau} \widetilde{\phi}(u)\,\mathrm{d}W_u \tag{6.5}$$

is gaussian $\mathcal{N}(0,1)$ (see, for example [28], Ch.17). Now, using this property, we can rewrite the deviation (4.6) on set Γ_T as

$$S^*_{h,T}(x_0) - S(x_0) = B^*_T + \mathbf{M}^{(1)}_T + \sigma(x_0)\mathbf{M}^{(2)}_T + \frac{\sigma(x_0)}{\sqrt{\delta H_T}}\,\xi_T \, , \tag{6.6}$$

where $B^*_T = \Upsilon_{1,T} + \Upsilon_{2,T} + B_T$,

$$\mathbf{M}^{(1)}_T = \frac{1}{\delta H_T} \sum_{j=N_0}^{N} \widetilde{\alpha}_j \, \chi_{h,x_0}(y_{t_{j-1}}) \, (\sigma(y_{t_{j-1}}) - \sigma(x_0)) \Delta W_{t_j}$$

and

$$\mathbf{M}^{(2)}_T = \frac{1}{\delta H_T} \left(\sum_{j=N_0}^{\varpi} \widetilde{\alpha}_j \, \phi_j \, \Delta W_{t_j} - \int_{T_0}^{\tau} \widetilde{\phi}(u)\,\mathrm{d}W_u \right) .$$

First we note that the definition of the sequence $(\widetilde{\alpha}_j)_{j\ge 1}$ in (3.6) implies

$$\sum_{j=N_0}^{N} \widetilde{\alpha}_j \chi_{h,x_0}(y_{t_{j-1}}) \le H_T \quad \text{a.s.} \tag{6.7}$$

Therefore, through the condition (2.1)

$$\mathbf{E}_\vartheta \left(\mathbf{M}_T^{(1)}\right)^2 = \mathbf{E}_\vartheta \left(\frac{1}{\delta H_T^2} \sum_{j=N_0}^{N} \widetilde{\alpha}_j^2 \chi_{h,x_0}(y_{t_{j-1}}) \left(\sigma(y_{t_{j-1}}) - \sigma(x_0)\right)^2 \right)$$

$$\leq \mathbf{E}_\vartheta \, \frac{h\sigma^2_{max}}{\delta H_T} \, .$$

Taking into account here that, for $T \geq 3$,

$$H_T \geq h(N - N_0)(2\sqrt{\upsilon_T} - \upsilon_T) \geq h(N - N_0)\sqrt{\upsilon_T} \, , \tag{6.8}$$

we obtain

$$\lim_{T\to\infty} \varphi_T \sup_{\vartheta\in\Theta_\beta} \mathbf{E}_\vartheta \, |\mathbf{M}_T^{(1)}| = 0 \, .$$

Now we study the term $\mathbf{M}_T^{(2)}$. To this end note that $t_{\varpi-1} < \tau \leq t_\varpi$. Therefore, we can represent this term as

$$\mathbf{M}_T^{(2)} = \frac{1}{\delta H_T} \left(\alpha \, \phi_\varpi \Delta W_{t_\varpi} - \phi_\varpi \left(W_\tau - W_{t_\varpi - 1} \right) \right) \, .$$

Moreover, taking into account that the stopping times ϖ and τ are bounded, one gets

$$\mathbf{E} \left(\Delta W_{t_\varpi} \right)^2 = \delta \quad \text{and} \quad \mathbf{E} \left(W_\tau - W_{t_{\varpi-1}} \right)^2 = \mathbf{E} \left(\tau - t_{\varpi-1} \right) \leq \delta \, .$$

Therefore, from here and (6.8), we get

$$\mathbf{E}_\vartheta \left(\mathbf{M}_T^{(2)}\right)^2 \leq 2\mathbf{E}_\vartheta \frac{1}{\delta H_T^2} \, \leq \frac{2}{\delta \, h^2 \upsilon_T (N - N_0)^2}$$

and

$$\lim_{T\to\infty} \varphi_T \sup_{\vartheta\in\Theta_\beta} \mathbf{E}_\vartheta \, |\mathbf{M}_T^{(2)}| = 0 \, .$$

To put an end to the proof of this theorem we have to show that

$$\lim_{T\to\infty} \sup_{\vartheta\in\Theta_\beta} \mathbf{E}_\vartheta \, K_T \, |\xi_T| \, = 0 \, , \tag{6.9}$$

where

$$K_T = \left| \frac{1}{\sqrt{2\widetilde{q}_T(x_0) - \upsilon_T}} - \frac{1}{\sqrt{2q_\vartheta(x_0)}} \right| \, .$$

It is easy to see that, for any $T > 0$, the random variable ξ_T is conditionally gaussian with respect to $\mathcal{F}_{T_0}$ with the parameters $(0, 1)$. Therefore,

$$\mathbf{E}_\vartheta \, K_T \, |\xi_T| = \sqrt{\frac{2}{\pi}} \, \mathbf{E}_\vartheta \, K_T \, .$$

Taking into account here Lemma 8.5 we come to the equality (6.9). Hence Theorem 5.2. □

7 Conclusions

In the paper we studied the estimation problem of the function S when its smoothness is known. In the case of unknown smoothness, in order to construct an adaptive estimate based on discrete time observations (1.2) in the model (1.1) we shall use the approach developed in [7] for continuous time observations. The approach make use of Lepskii's procedure and sequential estimating. Note that Lepskii's procedure works here just thanks to sequential estimating since, for the sequential estimate of the function S, the stochastic term in the deviation (6.6) is a gaussian random variable. This provides correct estimating the tail distribution of a kernel estimate and adapting for the pointwise risk. Moreover, for adaptive estimating in the case of quadratic risk, we shall apply the selection model developed in [13] to sequential kernel estimates (3.7). Note once more, that gaussianity of the stochastic term in (6.6) is a cornerstone result for obtaining a sharp oracle inequality. It permits to find Pinsker's constant like to [13] and then to study the proposed procedure efficiency.

These both programs will be realized in the next paper.

8 Appendix

8.1 Geometric ergodicity

First of all we recall that in [12] we have proved the following result.

Theorem 8.1. *For any $\epsilon > 0$, there exist constants $R = R(\epsilon) > 0$ and $\kappa = \kappa(\epsilon) > 0$ such that*

$$\sup_{u \geq 0} e^{\kappa u} \sup_{\|g\|_* \leq 1} \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Sigma_{L,M} \times \mathcal{V}} \frac{|\mathbf{E}_{\vartheta,x} \, g(y_u) - \mathbf{m}_\vartheta(g)|}{(1 + x^2)^\epsilon} \leq R \, ,$$

where $\mathbf{E}_{\vartheta,x} \, (\cdot) = \mathbf{E}_\vartheta \, (\cdot | y_0 = x)$, $\|g\|_* = \sup_x |g(x)|$.

8.2 Concentration inequalities

For any $\mathbb{R} \to \mathbb{R}$ function f belonging to $\mathbf{L}_1(\mathbb{R})$ we set

$$\mathbf{D}_n(f) = \sum_{k=1}^{n} \left(f(y_{t_k}) - \mathbf{m}_\vartheta(f) \right) . \tag{A.1}$$

Now we assume that the frequency δ in the observations (1.2) is of the following form

$$\delta = \delta_T = \frac{1}{(T+1) l_T} , \tag{A.2}$$

where the function l_T is such that,

$$\lim_{T\to\infty} \frac{l_T}{T^{1/2}} = 0 \quad \text{and} \quad \lim_{T\to\infty} \frac{l_T}{\ln T} = +\infty , \tag{A.3}$$

in particular, the function $l_T = (\ln T)^{1+\gamma}$ from (2.6) is of this kind. Moreover, let $\varkappa = \varkappa_T$ be a positive function satisfying the following properties

$$\lim_{T\to\infty} \varkappa_T = 0 \quad \text{and} \quad \lim_{T\to\infty} \frac{\varkappa_T^5 l_T}{\ln T} = +\infty . \tag{A.4}$$

Theorem 8.2. *([14]) Assume that the frequency δ satisfies* (A.2)- (A.3). *Then, for any $a > 0$,*

$$\lim_{T\to\infty} T^a \sup_{h \geq T^{-1/2}} \sup_{\vartheta \in \Theta_\beta} \mathbf{P}_\vartheta \left(|\mathbf{D}_N(\chi_{h,x_0})| \geq \varkappa_T T \right) = 0 . \tag{A.5}$$

8.3 Proof of Proposition 4.1

First, we note that by the Bunyakovskii - Cauchy - Schwarz inequality

$$\mathbf{E}_\vartheta \left(\varrho_j^2(S) | \mathcal{F}_{t_{j-1}} \right) \leq \delta \widetilde{L}^2 \int_{t_{j-1}}^{t_j} \mathbf{E}_\vartheta \left((y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}} \right) \mathrm{d}u ,$$

where $\widetilde{L} = \max(L,\, M)$. Note now that, for $t_{j-1} \leq u \leq t_j$,

$$\begin{aligned} \mathbf{E}_\vartheta ((y_u - y_{t_{j-1}})^2 | \mathcal{F}_{t_{j-1}}) &\leq 2\delta \left(\int_{t_{j-1}}^{u} \mathbf{E}_\vartheta (S^2(y_v) | \mathcal{F}_{t_{j-1}}) \, \mathrm{d}v + \sigma^2_{max} \right) \\ &\leq 2\delta \left(2 \int_{t_{j-1}}^{u} (M^2 + L^2 \mathbf{E}_\vartheta (y_v^2 | \mathcal{F}_{t_{j-1}})) \mathrm{d}v + \sigma^2_{max} \right) . \end{aligned}$$

Due to Proposition 8.6, we can estimate the last conditional expectation as

$$\sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\ \sup_{t_{j-1}\le u\le t_j}\mathbf{E}_\vartheta\left(y_u^2|\mathcal{F}_{t_{j-1}}\right)\le D_*L+y_{t_{j-1}}^2\,.$$

Therefore, taking into account that $\chi_{h,x_0}(y_{t_{j-1}})y_{t_{j-1}}^2\le x_*^2$, we obtain

$$\sup_{t_{j-1}\le u\le t_j}\ \sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\chi_{h,x_0}(y_{t_{j-1}})\,\mathbf{E}_\vartheta\left((y_u-y_{t_{j-1}})^2|\mathcal{F}_{t_{j-1}}\right)\le L_1\,\delta\,. \quad \text{(A.6)}$$

Therefore,

$$\sup_{j\ge1}\ \sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\chi_{h,x_0}(y_{t_{j-1}})\,\mathbf{E}_\vartheta\left(\varrho_j^2(S)|\mathcal{F}_{t_{j-1}}\right)\le\widetilde{L}^2\,L_1\,\delta^3\,.$$

Making use of the inequality (6.7) yields the following upper bound, through the Bunyakovskii - Cauchy - Schwarz inequality,

$$\begin{aligned}\mathbf{E}_\vartheta\Upsilon_T^2&\le\mathbf{E}_\vartheta\frac{1}{\delta^2H_T}\sum_{j=N_0}^{N}\widetilde{\alpha}_j\chi_{h,x_0}(y_{t_{j-1}})\varrho_j^2\\&=\mathbf{E}_\vartheta\frac{1}{\delta^2H_T}\sum_{j=N_0}^{N}\widetilde{\alpha}_j\chi_{h,x_0}(y_{t_{j-1}})\mathbf{E}_\vartheta\left(\varrho_j^2|\mathcal{F}_{t_{j-1}}\right)\\&\le\widetilde{L}^2\,L_1\,\delta\,.\end{aligned}$$

Hence Proposition 4.1. □

8.4 Proof of Proposition 4.2

Note that by the condition (2.13)

$$\begin{aligned}\mathbf{E}_\vartheta\left((\varrho_j^*)^2|\mathcal{F}_{t_{j-1}}\right)&=\int_{t_{j-1}}^{t_j}\mathbf{E}_\vartheta\left(\sigma(y_u)-\sigma(y_{t_{j-1}}))^2|\mathcal{F}_{t_{j-1}}\right)\mathrm{d}u\\&\le\sigma_{max}^2\int_{t_{j-1}}^{t_j}\mathbf{E}_\vartheta\left((y_u-y_{t_{j-1}})^2|\mathcal{F}_{t_{j-1}}\right)\mathrm{d}u\,.\end{aligned}$$

Therefore, using the inequality (A.6) we obtain

$$\sup_{j\ge1}\ \sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\chi_{h,x_0}(y_{t_{j-1}})\,\mathbf{E}_\vartheta\left((\varrho_j^*)^2|\mathcal{F}_{t_{j-1}}\right)\le\sigma_{max}^2\,L_1\,\delta^2\,.$$

From here and (6.7) and, taking into account that $0 < \widetilde{\alpha}_j \le 1$, we obtain

$$\mathbf{E}_\vartheta\, \Upsilon^2_{2,T} = \mathbf{E}_\vartheta\, \frac{1}{\delta^2 H_T^2} \sum_{j=N_0}^{N} \widetilde{\alpha}_j^2\, \chi_{h,x_0}(y_{t_{j-1}}) \mathbf{E}_\vartheta \left((\varrho_j^*)^2 | \mathcal{F}_{t_{j-1}} \right)$$

$$\le\, \sigma^2_{max}\, L_1\, \mathbf{E}_\vartheta\, \frac{1}{H_T}\,.$$

Now the inequality (6.8) yields (4.4). Hence Proposition 4.2. □

8.5 Proof of Proposition 4.3

Indeed, taking into account the inequalities (6.7) and (6.8), we obtain that, for any $T \ge 3$,

$$\mathbf{E}_\vartheta\, \mathbf{M}_T^2 \,= \mathbf{E}_\vartheta\, \frac{1}{\delta^2 H_T^2} \sum_{j=N_0}^{N} \chi_{h,x_0}(y_{t_{j-1}})\, \widetilde{\alpha}_j^2\, \mathbf{E}_\vartheta \left(\eta_j^2 | \mathcal{F}_{t_{j-1}} \right)$$

$$\le \mathbf{E}_\vartheta \frac{1}{\delta^2 H_T^2} \sum_{j=N_0}^{N} \widetilde{\alpha}_j\, \chi_{h,x_0}(y_{t_{j-1}}) \int_{t_{j-1}}^{t_j} \sigma^2(y_u)\, \mathrm{d}u$$

$$\le \frac{\sigma^2_{\max}}{\delta h (N - N_0) \sqrt{\upsilon_T}}\,.$$

Hence Proposition 4.3. □

8.6 Proof of Proposition 6.1

We start with setting

$$r_T = \frac{(2\widetilde{q}_T - \upsilon_T)h}{\mathbf{m}_\vartheta(\chi_{h,x_0})} \quad \text{and} \quad N_1 = N_0 + r_T(N - N_0)\,.$$

Note that $N_1 - N_0 \le \left(q_* \sqrt{\upsilon_T}\right)^{-1} N := N_1^*$, for sufficiently large T. Moreover, we set

$$G_T = \frac{1}{H_T} \sum_{j=N_0}^{N_1} f_h(y_{t_{j-1}}) \quad \text{and} \quad \widehat{G}_T = G_T - B_T\,.$$

Using (2.11) we can represent the term G_T as

$$G_T = \frac{N_1 - N_0}{H_T}\, \mathbf{m}_\vartheta(f_h) \,+\, \frac{1}{H_T} \sum_{j=N_0}^{N_1} \widetilde{f}_h(y_{t_{j-1}}) := G_1(T) + G_2(T)\,,$$

where $\widetilde{f}_h(y) = f_h(y) - \mathbf{m}_\vartheta(f_h)$. Taking into account that $\mathbf{m}_\vartheta(\chi_{h,x_0}) \geq 2hq_*$, we obtain

$$|G_1(T)| = \frac{r_T(N-N_0)h}{H_T}\,|\mathbf{m}^*_\vartheta(h)| \leq \frac{1}{2q_*}|\mathbf{m}^*_\vartheta(h)|\,, \qquad \mathbf{m}^*_\vartheta(h) = \frac{\mathbf{m}_\vartheta(f_h)}{h}\,.$$

Let us represent the last term as

$$\mathbf{m}^*_\vartheta(h) = q_\vartheta(x_0)\Omega_{x_0,h}(S) + \widetilde{\mathbf{m}}_\vartheta(h)\,,$$

where $\widetilde{\mathbf{m}}_\vartheta(h) = \int_{-1}^1 (S(x_0+hz)-S(x_0))\,(q_\vartheta(x_0+hz)-q_\vartheta(x_0))\,\mathrm{d}z$. Furtherer, by the definition (2.5), one has

$$\Omega_{x_0,h}(S) = \Omega_{x_0,h}(S_0) + \Omega_{x_0,h}(D)\,,$$

for some function D from $\mathcal{H}^w_{x_0}(\epsilon,\beta,h)$. Therefore, the properties (2.3)-(2.4) and (2.6) yield

$$\lim_{h\to 0}\,\varphi_T \sup_{S\in\mathcal{U}_M(x_0,\beta)}\,\left|\Omega_{x_0,h}(S)\right| = 0\,.$$

Obviously, that

$$\limsup_{h\to 0}\,h^{-2}\sup_{\vartheta\in\Theta_\beta}\,|\widetilde{\mathbf{m}}_\vartheta(h)| < \infty\,.$$

Hence,

$$\limsup_{T\to\infty}\,\varphi_T\sup_{\vartheta\in\Theta_\beta}\,\mathbf{E}_\vartheta\,|G_1(T)| = 0\,.$$

Now we note that,

$$\mathbf{E}_\vartheta\,G_2^2(T) \,=\, \mathbf{E}_\vartheta\,\frac{1}{H_T^2}\left(\sum_{j=N_0}^{N_1-1}\,\Psi_j\,+\,\widetilde{f}_h^2(y_{t_{N_1-1}})\right)\,,$$

where $\Psi_j = \widetilde{f}_h^2(y_{t_{j-1}}) + 2\widetilde{f}_h(y_{t_{j-1}})\sum_{l=j+1}^{N_1}\mathbf{E}_\vartheta\left(\widetilde{f}_h(y_{t_{l-1}})|\mathcal{F}_{t_{j-1}}\right)$ and $\mathcal{F}_t = \sigma\{y_s\,,\,0\leq s\leq t\}$. Taking into account that $(y_t)_{t\geq 0}$ is a homogeneous Markov process and that $|\widetilde{f}_h(y)| \leq 2Mh$, we estimate from above the last conditional expectation, through the Theorem 8.1 (for $\epsilon = 1/2$), as

$$\left|\mathbf{E}_\vartheta\left(\widetilde{f}_h(y_{t_{l-1}})|\mathcal{F}_{t_{j-1}}\right)\right| = \left|\mathbf{E}_{\vartheta,y_{t_{j-1}}}\,\widetilde{f}_h(y_{t_{l-j}})\right| \leq 2MhR\left(1+y^2_{t_{j-1}}\right)^{1/2}\,e^{-\kappa t_{l-j}}$$

$$\leq 2MhR\left(1+|y_{t_{j-1}}|\right)\,e^{-\kappa\delta(l-j)}\,.$$

Therefore,

$$|\Psi_j| \le 4M^2h^2\left(1+\frac{2R(1+|y_{t_{j-1}}|)}{e^{\kappa\delta}-1}\right).$$

From here, bounding $e^{\kappa\delta}-1$ by $\kappa\delta$, we get

$$\begin{aligned}\mathbf{E}_\vartheta\, G_2^2(T) &\le 8M^2h^2\mathbf{E}_\vartheta\frac{1}{H_T^2}\sum_{j=N_0}^{N_1}\left(1+\frac{R}{\kappa\delta}(1+|y_{t_{j-1}}|)\right)\\ &\le 8M^2h^2\left(1+\frac{R}{\kappa\delta}\right)\mathbf{E}_\vartheta\frac{(N_1-N_0)}{H_T^2}\\ &\quad+8M^2h^2\frac{R}{\kappa\delta}\mathbf{E}_\vartheta\frac{1}{H_T^2}\sum_{j=N_0}^{N_1}\left(\mathbf{E}_\vartheta(y^2_{t_{j-1}}|\mathcal{F}_{t_{N_0-1}})\right)^{1/2}.\end{aligned}$$

By making use of Proposition 8.6 one obtains

$$\mathbf{E}_\vartheta\, G_2^2(T) \le 8M^2h^2\left(1+\frac{R}{\kappa\delta}\right)\mathbf{E}_\vartheta\frac{N_1-N_0}{H_T^2}\left(1+\sqrt{D_*L}+|y_{t_{N_0-1}}|\right).$$

Now from (6.8) it follows that

$$\frac{N_1-N_0}{H_T^2}=\frac{1}{H_T\mathbf{m}_\vartheta(\chi_{h,x_0})}\le\frac{1}{2h^2\sqrt{\upsilon_T}(N-N_0)q_*}.$$

Thus,

$$\sup_{\vartheta\in\Theta_\beta}\mathbf{E}_\vartheta\, G_2^2(T)\le\frac{G^*}{\delta\sqrt{\upsilon_T}(N-N_0)},$$

where $G^*=4M^2(\kappa+R)\left(1+2\sqrt{D_*L}+|y_0|\right)/(\kappa q_*)$. From this equality we obtain immediately

$$\lim_{T\to\infty}\varphi_T\sup_{\vartheta\in\Theta_\beta}\mathbf{E}_\vartheta\,|G_2(T)|=0.$$

Let us estimate the term $\widehat{G}_T$. Taking into account the lower bound (6.8) we get

$$\begin{aligned}|\widehat{G}_T| &\le\frac{Mh}{H_T}\left|\sum_{j=N_0}^{N_1}\chi_{h,x_0}(y_{t_{j-1}})-H_T\right|+\frac{2Mh}{H_T}\\ &\le\frac{M}{\sqrt{\upsilon_T}(N-N_0)}\sum_{j=N_0}^{N_1}\left|\widetilde{\chi}_h(y_{t_{j-1}})\right|+\frac{2M}{\sqrt{\upsilon_T}(N-N_0)},\end{aligned}$$

where $\widetilde{\chi}_h(y) = \chi_{h,x_0}(y) - \mathbf{m}_\vartheta(\chi_{h,x_0})$. By making use of the Theorem 8.1 with $\epsilon = 1/2$ one gets

$$\sum_{j \geq N_0} \mathbf{E}_{\vartheta,y_0} \, |\widetilde{\chi}_h(y_{t_{j-1}})| \leq \frac{R e^{-\kappa\delta(N_0-1)}}{1 - e^{-\kappa\delta}} \left(1 + \sqrt{D_* L} + |y_0|\right) .$$

This inequality implies directly

$$\lim_{T\to\infty} \varphi_T \sup_{\vartheta\in\Theta_\beta} \mathbf{E}_\vartheta \, |\widehat{G}_T| = 0 \, .$$

Hence Proposition 6.1. □

8.7 Properties of the estimate (3.3)

Lemma 8.3. *Assume that the parameter δ is of the form (2.12). Then, for any $a > 0$,*

$$\lim_{T\to\infty} T^a \sup_{\vartheta\in\Theta_\beta} \mathbf{P}_\vartheta(|\widehat{q}_T(x_0) - q_\vartheta(x_0)| > \upsilon_T) = 0 \, .$$

Proof. Denoting $\psi_\varsigma(y) = (1/\varsigma)Q((y - x_0)/\varsigma)$ one has

$$\widehat{q}_T(x_0) - q_\vartheta(x_0) = \frac{1}{2} \int_{-1}^{1} (q_\vartheta(x_0 + \varsigma z) - q_\vartheta(x_0)) \, \mathrm{d}z + \frac{1}{2(N_0 - 1)} \mathbf{D}_{N_0-1}(\psi_\varsigma) \, .$$

Therefore

$$\mathbf{P}_\vartheta(|\widehat{q}_T(x_0) - q_\vartheta(x_0)| > \upsilon_T)$$

$$\leq \mathbf{P}_\vartheta(|\int_{-1}^{1} (q_\vartheta(x_0 + \varsigma z) - q_\vartheta(x_0)) \, \mathrm{d}z| > \upsilon_T) + \mathbf{P}_\vartheta(\frac{1}{(N_0 - 1)} \mathbf{D}_{N_0-1}(\psi_\varsigma) > \upsilon_T) \, .$$

The first term on the right-hand side equals to zero for sufficiently large T since

$$|\int_{-1}^{1} (q_\vartheta(x_0 + \varsigma z) - q_\vartheta(x_0)) \, \mathrm{d}z| \leq \varsigma^2 \ddot{q}^* < \upsilon_T \, ,$$

for sufficiently large T, where $\ddot{q}^* = \sup_x \sup_\vartheta |\ddot{q}_\vartheta(x)| < \infty$. Applying Theorem 8.2 to the second term on the right-hand side of the same inequality yields the Lemma 8.3. □

Lemma 8.4. *Assume that the parameter δ is of the form* (2.12). *Then, for any $a > 0$,*

$$\lim_{T\to\infty} T^a \sup_{\vartheta\in\Theta_\beta} \mathbf{P}_\vartheta(|\widetilde{q}_T(x_0) - q_\vartheta(x_0)| > \upsilon_T) = 0\,.$$

Proof. Note, that for sufficiently large T (for that $\ln T \geq \max(q^{*2}, 1/q_*^2)$),

$$|\widetilde{q}_T(x_0) - q_\vartheta(x_0)| \leq |\widehat{q}_T(x_0) - q_\vartheta(x_0)|\,.$$

The Lemma follows immediately from Lemma 8.3. □

Lemma 8.5. *Assume that the parameter δ is of the form* (2.12). *Then,*

$$\limsup_{T\to\infty} \frac{1}{\upsilon_T^{1/2}} \sup_{\vartheta\in\Theta_\beta} \mathbf{E}_\vartheta \left|\frac{1}{\widetilde{q}_T(x_0) - \upsilon_T} - \frac{1}{q_\vartheta(x_0)}\right| \leq \frac{4}{q^*} < \infty\,. \tag{A.7}$$

Proof. Indeed, for sufficiently large T for which

$$\upsilon_T \leq \min\left(1/(q^*)^2\,,\, 1/4\right)\,,$$

we obtain

$$\begin{aligned}\mathbf{E}_\vartheta \left|\frac{1}{\widetilde{q}_T(x_0) - \upsilon_T} - \frac{1}{q_\vartheta(x_0)}\right| &\leq \frac{2\upsilon_T^{1/2}}{q_*} + \frac{2}{q_*\upsilon_T^{1/2}}\,\mathbf{E}_\vartheta\,|\widetilde{q}_T(x_0) - q_\vartheta(x_0)| \\ &\leq \frac{4\upsilon_T^{1/2}}{q_*} + \frac{2}{q_*\upsilon_T}\,\mathbf{P}_\vartheta\left(|\widetilde{q}_T(x_0) - q_\vartheta(x_0)| > \upsilon_T\right)\,.\end{aligned}$$

Now Lemma 8.4 implies the equality (A.7). Hence Lemma 8.5. □

8.8 Moment inequality for the process (1.1)

We state the moment bound from [12].

Proposition 8.6. *Let $(y_t)_{t\geq 0}$ be a solution of the equation* (1.1). *Then, for any $z \in \mathbb{R}$ and $m \geq 1$,*

$$\sup_{u\geq 0} \sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]} \mathbf{E}_{\vartheta,z}\,(y_u)^{2m} \leq (2m-1)!!\,(D_*L + z^2)^m\,, \tag{A.8}$$

where $\mathbf{E}_{\vartheta,z}(\cdot) = \mathbf{E}_\vartheta(\cdot|y_0 = z)$ and $D_* = (M + Lx_* + 2x_*)^2(L + M) + \sigma_{max}^2$.

Proof. To obtain this inequality we make use of the method proposed in ([22], p.20) for linear stochastic equations. First of all, note that thanks to Theorem 4.7 from [28], for any $T > 0$, there exists some $\epsilon > 0$ such that, for each $\vartheta \in \Theta_\beta$ and $z \in \mathbb{R}$,

$$\sup_{0\le t\le T} \mathbf{E}_{\vartheta,z}\, e^{\epsilon y_t^2} \,<\infty\,. \tag{A.9}$$

Let us denote $D_\vartheta(y) = 2yS(y) + \sigma^2(y) + \varkappa y^2$ and $\varkappa = L^{-1}$. Taking into account that $0 < \varkappa < 1$ and $x_* \ge 1$ we obtain that, for $|y| \le x_*$,

$$|D_\vartheta(y)| \le x_*^2(2M+1) + \sigma^2_{max}.$$

Let now $|y| \ge x_*$. Denoting by y_* the projection of y onto the interval $[-x_*\,,\,x_*]$ we obtain that

$$\begin{aligned} 2yS(y) &= 2yS(y_*) + 2y_*\,(S(y) - S(y_*)) + 2(y-y_*)\,(S(y) - S(y_*)) \\ &\le 2|y|M + 2Lx_*\,|y-y_*| - 2\varkappa|y-y_*|^2 \\ &\le 2(M + Lx_* + 2x_*)|y| - 2\varkappa y^2\,. \end{aligned}$$

Therefore,

$$\sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\ \sup_{y\in\mathbb{R}}\ D_\vartheta(y)\ \le D_*\,.$$

By the Ito formula we obtain

$$\begin{aligned} \mathrm{d}y_u^{2m} &= -m\varkappa y_u^{2m}\mathrm{d}t + m y_u^{2(m-1)}\left(D_\vartheta(y_u) + 2(m-1)\sigma^2(y_u)\right)\mathrm{d}t \\ &\quad + 2m y_u^{2m-1}\sigma(y_u)\mathrm{d}W_t\,. \end{aligned}$$

Moreover, the property (A.9) yields that, for any $m \ge 1$,

$$\mathbf{E}_\vartheta \int_0^t e^{-m\varkappa(t-s)}\, y_s^{2m-1}\sigma(y_s)\mathrm{d}W_s = 0\,.$$

Therefore, $\mathbf{E}_\vartheta y_t^{2m} \le z^{2m} + m(2m-1)D_* \int_0^t e^{-m\varkappa(t-s)}\,\mathbf{E}_\vartheta y_s^{2(m-1)}\,\mathrm{d}s$. Now the induction implies directly the bound (A.8). Hence Proposition 8.6. □

Proposition 8.7. *Let $(y_t)_{t\ge0}$ be a solution of the equation (1.1). Then, for any $z \in \mathbb{R}$ and $m \ge 1$, and for any stopping time τ taking values in $[0,T]$, one has*

$$\sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]} \mathbf{E}_{\vartheta,z}\,(y_\tau)^{2m} \ \le\ B^*(m,z)\,T \tag{A.10}$$

and

$$\sup_{\vartheta\in\Sigma_{L,M}\times[0,\sigma_{max}]}\mathbf{E}_{\vartheta,z}\sup_{0\le u\le T}(y_u)^{2m}\ \le\ B_1^*(m,z)\,T\,, \tag{A.11}$$

where $B^*(m,z)=(2m-1)!!\,(D_*L+z^2)^m\,(D_*+2(m-1)\sigma^2_{max})$ *and* $B_1^*(m,z)=1+mB^*(m+1,z)$.

The proof of this proposition follows immediately from Proposition 1.1.5 in [22].

8.9 Proof of Proposition 4.5

It is clear, that to show (3.7) it suffices to check that, for any $a>0$,

$$\lim_{T\to\infty}T^a\sup_{\vartheta\in\Sigma_{L,M}\times[\sigma_{min},\sigma_{max}]}\mathbf{P}_\vartheta(\Gamma_T^c)=0\,. \tag{A.12}$$

Indeed, by the definition of ϖ

$$\begin{aligned}\mathbf{P}_\vartheta(\Gamma_T^c)&=\mathbf{P}_\vartheta\left(\sum_{j=N_0}^{N}\chi_{h,x_0}(y_{t_{j-1}})\ <H_T\right)\\&=\mathbf{P}_\vartheta\left(\mathbf{D}_{N_0,N-1}(\chi_{h,x_0})<(2\widetilde{q}_T-\upsilon_T-\mathbf{m}^*_\vartheta(\chi_{h,x_0}))\,(N-N_0)h\right)\,,\end{aligned}$$

where $\mathbf{D}_{k,n}(f)=\mathbf{D}_n(f)-\mathbf{D}_k(f)$ and

$$\mathbf{m}^*_\vartheta(\chi_{h,x_0})=\frac{\mathbf{m}_\vartheta(\chi_{h,x_0})}{h}=\int_{-1}^{1}q_\vartheta(x_0+hz)\,\mathrm{d}z\,.$$

Taking into account the definition of υ_T in (3.3) we obtain that, for sufficiently large T,

$$\sup_{\vartheta\in\Theta_\beta}\int_{-1}^{1}|q_\vartheta(x_0+hz)-q_\vartheta(x_0)|\,\mathrm{d}z\le\upsilon_T/4\,.$$

Therefore, for such T,

$$\begin{aligned}\mathbf{P}_\vartheta(\varpi>N)&\le\mathbf{P}_\vartheta\left(|\widetilde{q}_T(x_0)-q_\vartheta(x_0)|>\upsilon_T/8\right)\\&\quad+\mathbf{P}_\vartheta\left(|\mathbf{D}_{N_0,N-1}(\chi_{h,x_0})|>\ Nh\upsilon_T/2\right)\,.\end{aligned}$$

Now we estimate the last term as

$$\begin{aligned}\mathbf{P}_\vartheta\left(|\mathbf{D}_{N_0,N-1}(\chi_{h,x_0})|>\ Nh\upsilon_T/2\right)&\le\mathbf{P}_\vartheta\left(|\mathbf{D}_{N-1}(\chi_{h,x_0})|>\ Nh\upsilon_T/4\right)\\&\quad+\mathbf{P}_\vartheta\left(|\mathbf{D}_{N_0}(\chi_{h,x_0})|>\ Nh\upsilon_T/4\right)\,.\end{aligned}$$

By applying Lemma 8.3 and the inequality (A.5) we obtain (A.12). □

References

[1] M. Arato. *Linear stochastic systems with constant coefficients. A statistical approach.* Springer-Verlag, Berlin, Heidelberg, New York, 1982.

[2] M. Arato, A. N. Kolmogorov, and Ya. G. Sinai. On parameter estimation of a complex stationary gaussian process. *Reports of the Acad. Sciences USSR*, 146:747–750, 1962.

[3] A. Bensoussan. *Optimal control of partially observed systems.* Cambridge University Press, Cambridge, 1992.

[4] V. Z. Borisov and V. V. Konev. Sequential estimation of parameters of discrete processes. *Automat. and Remote Control*, 10:58–64, 1977.

[5] F. Comte, V. Genon-Catalot, and Y. Rozenholc. Nonparametric estimation for a discretely observed integrated diffusion model. *Stochastic Processes and their Applications*, 119:811–834, 2009.

[6] D. Florens-Zmirou. On estimating the diffusion coefficient from discret observations. *J. Appl. Probab.*, 30:790–804, 1993.

[7] L. I. Galtchouk and S. M. Pergamenshchikov. Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Mathematical Methods of Statistics*, 10:316–330, 2001.

[8] L. I. Galtchouk and S. M. Pergamenshchikov. Non parametric sequential minimax estimation of the drift coefficient in diffusion processes. *Sequential Analysis*, 24(3):303–330, 2005.

[9] L. I. Galtchouk and S. M. Pergamenshchikov. Asymptotically efficient estimates for non parametric regression models. *Statistics and Probability Letters*, 76(8):852–860, 2006.

[10] L. I. Galtchouk and S. M. Pergamenshchikov. Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, 9(1):1–16, 2006.

[11] L. I. Galtchouk and S. M. Pergamenshchikov. Uniform concentration inequality for ergodic diffusion processes. *Stochastic Processes and their applications*, 117:830–839, 2007.

[12] L. I. Galtchouk and S. M. Pergamenshchikov. Geometric ergodicity for families of homogenuous markov chains. *Preprint, http://hal.archives-ouvertes.fr/hal-00455976*, 2010.

[13] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *Journal of Nonparametric Statistics*, 23(2):255–285, 2011.

[14] L. I. Galtchouk and S. M. Pergamenshchikov. Uniform concentration inequality for discrete-time ergodic diffusion processes. *Preprint*, 2011.

[15] V. Genon-Catalot, C. Laredo, and D. Picard. Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.*, 19:317–335, 1992.

[16] I. I. Gihman and A. V. Skorohod. *Stochastic differential equations*. Naukova Dumka, Kiev, 1968.

[17] E. Gobet. Lan property for ergodic diffusions with discrete observations. *Ann. Inst. Henri Poincaré Probab. Statist.*, 38:711–737, 2002.

[18] E. Gobet, M. Hoffmann, and M. Reiss. Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, 32:2223–2253, 2004.

[19] M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79:135–163, 1999.

[20] J. Jacod. Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.*, 27(1):83–96, 2000.

[21] J. Jacod. Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli*, 12(3):383–401, 2006.

[22] Yu. M. Kabanov and S. M. Pergamenshchikov. *Two-Scale Stochastic Systems. Asymptotic Analysis and Control*. Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[23] I. Karatzas and S. E. Shreve. *Methods of Mathematical Finance*. Springer, New York, 1998.

[24] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24:211–229, 1997.

[25] V. V. Konev and S.M. Pergamenshchikov. Sequential estimation of the parameters of random processes with continuous time. In F. Tarasenko, editor, *Math. Statistics and Appl.*, pages 93–101. Publishing Hause of Tomsk State University, Tomsk, 1981.

[26] V. V. Konev and S.M. Pergamenshchikov. On truncated sequential estimation of the parameters of diffusion processes. In S. Aivazian, editor, *Methods of Economical Analysis*, pages 3–31. Central Economical and Mathematical Institute of Russian Academy of Science, Moscow, 1992.

[27] Yu. A. Kutoyants. *Statistical Inferences for Ergodic Diffusion Processes*. Springer, Berlin, 2003.

[28] R. Sh. Liptser and A. N. Shiryaev. *Statistics of a random process*. Springer-Verlag, Berlin, Heidelberg, New York,v.I,II, 1978.

[29] A. A Novikov. Sequential estimation of the parameters of the diffusion processes. *Theory Probability and Appl.*, 16:394–396, 1971.

[30] N. Yoshida. Estimation for diffusion processes from discrete observations. *J. Multivariate Analysis*, 41:220–242, 1992.