Check for updates

Efficient Improved Estimation Method for Non-Gaussian Regression from Discrete Data

Evgeny Pchelintsev[1(✉)] and Serguei Pergamenshchikov[2]

[1] Tomsk State University, Tomsk, Russian Federation
evgen-pch@yandex.ru
[2] Université de Rouen Normandie, Rouen, France
https://persona.tsu.ru/Home/UserProfile/393

Abstract. We study a robust adaptive nonparametric estimation problem for periodic functions observed in discrete fixed time moments with non-Gaussian Ornstein–Uhlenbeck noises. For this problem we develop a model selection method, based on the shrinkage (improved) weighted least squares estimates. We found constructive sufficient conditions for the observations frequency under which sharp oracle inequalities for the robust risks are obtained. Moreover, on the basis of the obtained oracle inequalities we establish for the proposed model selection procedures the robust efficiency property in adaptive setting. Then, we apply the constructed model selection procedures to estimation problems in Big Data models in continuous time. Finally, we provide Monte - Carlo simulations confirming the obtained theoretical results.

Keywords: Nonparametric regression · Non-Gaussian Ornstein–Uhlenbeck process · Discrete observations · Improved model selection method · Sharp oracle inequality · Asymptotic efficiency

1 Introduction

In this paper we consider the following nonparametric regression model in continuous time

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t, \quad 0 \le t \le T, \tag{1}$$

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function from $\mathbf{L}_2[0,1]$, the duration of observations T is integer and $(\xi_t)_{t\ge0}$ is defined by a Ornstein – Uhlenbeck – Lévy defined as

$$\mathrm{d}\xi_t = a\xi_t\mathrm{d}t + \mathrm{d}u_t, \quad u_t = \varrho_1\, w_t + \varrho_2\, z_t, \quad \xi_0 = 0. \tag{2}$$

Here $(w_t)_{t\ge0}$ is a standard Brownian motion, z_t is a pure jump Lévy process defined through the stochastic integral with respect to the compensated jump measure $\mu(\mathrm{d}s, \mathrm{d}x)$ with deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, i.e.

$$z_t = x * (\mu - \widetilde{\mu})_t = \int_0^t \int_{\mathbb{R}_*} v\,(\mu - \widetilde{\mu})(\mathrm{d}s\,\mathrm{d}v) \quad \text{and} \quad \mathbb{R}_* = \mathbb{R} \setminus \{0\},$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
A. N. Shiryaev et al. (Eds.): ICSM-5 2020, PROMS 371, pp. 164–177, 2021.
https://doi.org/10.1007/978-3-030-83266-7_12

$\Pi(\cdot)$ is the Lévy measure on $\mathbb{R}_*$, (see, for example in [2]), such that

$$\int_{\mathbb{R}_*} z^2\,\Pi(\mathrm{d}z) = 1 \quad \text{and} \quad \int_{\mathbb{R}_*} z^8\,\Pi(\mathrm{d}z) < \infty.$$

We assume that the unknown parameters $a \le 0$, ϱ_1 and ϱ_2 are such that

$$-a_{max} \le a \le 0, \quad 0 < \underline{\varrho} \le \varrho_1^2 \quad \text{and} \quad \sigma_Q = \varrho_1^2 + \varrho_2^2 \le \varsigma^*. \tag{3}$$

Moreover, we assume that the bounds a_{max}, $\underline{\varrho}$ and ς^* are functions of T, i.e. $a_{max} = a_{max}(T)$, $\underline{\varrho} = \underline{\varrho}_T$ and $\varsigma^* = \varsigma^*_T$, for which for any $\epsilon > 0$

$$\lim_{T\to\infty} \frac{a_{max}(T) + \varsigma^*_T}{T^\epsilon} = 0 \quad \text{and} \quad \liminf_{T\to\infty} T^\epsilon\, \underline{\varrho}_T > 0. \tag{4}$$

We denote by $\mathcal{Q}_T$ the family of all distributions of process (1)–(2) on the Skorokhod space $\mathbf{D}[0,n]$ satisfying the conditions (3)–(4). It should be noted that the process (2) is conditionally-Gaussian square integrated semimartingale with respect to σ-algebra $\mathcal{G} = \sigma\{z_t,\, t \ge 0\}$ which is generated by the jump process $(z_t)_{t\ge 0}$.

The problem is to estimate the unknown function S in the model (1) on the basis of observations

$$(y_{t_j})_{0\le j\le n}, \quad t_j = j\Delta \quad \text{and} \quad \Delta = \frac{1}{p}, \tag{5}$$

where $n = Tp$ and the observations frequency p is some fixed integer number. For this problem we use the quadratic risk, which for any estimate $\widehat{S}$, is defined as

$$\mathcal{R}_Q(\widehat{S}, S) := \mathbf{E}_{Q,S}\,\|\widehat{S} - S\|^2 \quad \text{and} \quad \|f\|^2 := \int_0^1 f^2(t)\mathrm{d}t, \tag{6}$$

where $\mathbf{E}_{Q,S}$ stands for the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1) with a fixed distribution Q of the noise $(\xi_t)_{0\le t\le n}$ and a given function S. Moreover, in the case when the distribution Q is unknown we use also the robust risk

$$\mathcal{R}^*_T(\widehat{S}, S) = \sup_{Q\in\mathcal{Q}_T} \mathcal{R}_Q(\widehat{S}, S). \tag{7}$$

Note that if $(\xi_t)_{t\ge 0}$ is a Brownian motion, then we obtain the well known white noise model (see, for example, [7] and [13]). Later, to take into account the dependence structure in the papers [6] and [10] it was proposed to use the Ornstein – Uhlenbeck noise processes, so called color Gaussian noises. Then, to study the estimation problem for non-Gaussian observations (1) in the papers [9,11] and [12] it was introduced impulse noises defined through the compound Poisson processes with unknown impulse distributions. However, compound Poisson processes can describe the impulse influence of only one fixed frequency and, therefore, such models are too restricted for practical applications. In this paper we consider more general pulse noises described by the Ornstein – Uhlenbeck – Lévy processes.

Our main goal in this paper is to develop improved estimation methods for the incomplete observations, i.e. when the process (1) is available for observations only in the fixed time moments (5). To this end we propose adaptive model selection method based on the improved weighted least squares estimates. For nonparametric estimation problem such approach was proposed in [15] for Lévy regression model.

2 Improved Estimation Method

First, we chose the trigonometric basis $(\phi_j)_{j\ge 1}$ in $\mathbf{L}_2[0,1]$, i.e. $\phi_1 \equiv 1$ and for $j \ge 2$

$$\phi_j(x) = \sqrt{2}\begin{cases} \cos(2\pi[j/2]x) & \text{for even} \quad j; \\ \sin(2\pi[j/2]x) & \text{for odd} \quad j, \end{cases} \tag{8}$$

where $[a]$ denotes the integer part of a. Note that if p is odd, then for any $1 \le i,j \le p$

$$(\phi_i,\phi_j)_p = \frac{1}{p}\sum_{l=1}^{p} \phi_i(t_l)\phi_j(t_l) = \mathbf{1}_{\{i=j\}}. \tag{9}$$

We use this basis to represent the function S on the lattice $\mathcal{T}_p = \{t_1,...,t_p\}$ in the Fourier expansion form

$$S(t) = \sum_{j=1}^{p}\theta_j\phi_j(t) \quad \text{and} \quad \theta_j = (S,\phi_j)_p := \frac{1}{p}\sum_{k=1}^{p} S(t_k)\phi_j(t_k).$$

The coefficients θ_j can be estimated from the discrete data (5) as

$$\widehat{\theta}_j = \frac{1}{T}\int_0^T \psi_j(t)\,\mathrm{d}y_t \quad \text{and} \quad \psi_j(t) = \sum_{k=1}^{n}\phi_j(t_k)\mathbf{1}_{(t_{k-1},t_k]}(t).$$

We note that the system of the functions $\{\psi_j\}_{1\le j\le p}$ is orthonormal in $\mathbf{L}_2[0,1]$. Now we set weighted least squares estimates for $S(t)$ as

$$\widehat{S}_\gamma(t) = \sum_{j=1}^{p}\gamma(j)\widehat{\theta}_j\psi_j(t) \tag{10}$$

with weights $\gamma = (\gamma(j))_{1\le j\le p}$ from a finite set $\Gamma \subset [0,\,1]^p$. Now for the weight coefficients we introduce the following size characteristics

$$\nu = \#(\Gamma) \quad \text{and} \quad \nu_* = \max_{\gamma\in\Gamma}\sum_{j=1}^{p}\gamma(j),$$

where $\#(\Gamma)$ is the number of the vectors γ in Γ.

Definition 1. *Function* $\mathbf{g}(T)$ *is called slowly increasing as* $T \to \infty$*, if for any* $\epsilon > 0$

$$\lim_{T\to\infty} T^{-\epsilon}\,\mathbf{g}_T = 0.$$

$\mathbf{H}_1$) *For any vector* $\gamma \in \Gamma$ *there exists some fixed integer* $7 \le d = d(\gamma) \le p$ *such that their first* d *components are equal to one, i.e.* $\gamma(j) = 1$ for $1 \le j \le d$*. Moreover, we assume that the parameters* ν and ν_* *are functions of* T*, i.e.* $\nu = \nu(T)$ and $\nu_* = \nu_*(T)$*, and the functions* $\nu(T)$ *and* $T^{-1/3}\nu_*(T)$ *are slowly increasing as* $T \to \infty$*.*

Using this condition, we define the shrinkage weighted least squares estimates for S

$$S^*_\gamma(t) = \sum_{j=1}^{p} \gamma(j)\theta^*_j \psi_j(t), \quad \theta^*_j = \left(1 - \frac{\mathbf{c}_T}{\sqrt{\sum_{j=1}^{d} \widehat{\theta}^2_j}} \mathbf{1}_{\{1\le j\le d\}}\right)\widehat{\theta}_j, \tag{11}$$

where

$$\mathbf{c}_T = \frac{\underline{\varrho}_T(d-6)}{2\left(\mathbf{r} + \sqrt{2d\varsigma^*/T}\right)T}$$

and the radius $\mathbf{r} > 0$ may be dependent of T, i.e. $\mathbf{r} = \mathbf{r}_T$ as a slowly increasing function for $T \to \infty$. To compare the estimates (10) and (11) we set

$$d_0 = \inf\{d \ge 7 \,:\, 5 + \ln d \le \check{a} d\} \quad \text{and} \quad \check{a} = \frac{1 - e^{-a_{max}}}{4a_{max}}.$$

Now we can compare the estimators (10) and (11) in mean square accuracy sense.

Theorem 1. *Assume that the condition* $\mathbf{H}_1$) *holds with* $d \ge d_0$*. Then for any* $p \ge d$ *and* $T \ge 3$

$$\sup_{Q\in\mathcal{Q}_T}\ \sup_{\|S\|\le \mathbf{r}} \left(\mathcal{R}_Q(S^*_\gamma, S) - \mathcal{R}_Q(\widehat{S}_\gamma, S)\right) < -\mathbf{c}^2_T. \tag{12}$$

Remark 1. The inequality (12) means that non-asymptotically, uniformly in $p \ge d$ the estimate (11) outperforms in square accuracy the estimate (10). Such estimators are called improved. Note that firstly for parametric regression models in continuous time similar estimators were proposed in [14] and [12]. Later, for Lévy models in nonparametric setting these methods were developed in [15].

3 Adaptive Model Selection Procedure

To obtain a good estimate from the class (11), we have to choose a weight vector $\gamma \in \Gamma$. The best way is to minimize the empirical squared error

$$\mathrm{Err}_p(\gamma) = \|S^*_\gamma - S\|^2$$

with respect to γ. Since this quantity depends on the unknown function S and, hence, depends on the unknown Fourier coefficients $(\theta_j)_{j\geq 1}$, the weight coefficients $(\gamma_j)_{j\geq 1}$ cannot be found by minimizing one. Then, one needs to replace the corresponding terms by their estimators. For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_p(\gamma) = \sum_{j=1}^{p} \gamma^2(j)(\theta^*_j)^2 - 2\sum_{j=1}^{p} \gamma(j) \left(\theta^*_j \widehat{\theta}_j - \frac{\widehat{\sigma}_T}{T}\right) + \rho \widehat{P}_T(\gamma). \tag{13}$$

Here ρ is some positive penalty coefficient, $\widehat{P}_T(\gamma)$ is the penalty term is defined as

$$\widehat{P}_T(\gamma) = \frac{\widehat{\sigma}_T}{T} \sum_{j=1}^{p} \gamma^2(j),$$

where $\widehat{\sigma}_T$ is the estimate for the variance σ_Q which is chosen for $\sqrt{T} \leq p \leq T$ in the following form

$$\widehat{\sigma}_T = \frac{T}{p} \sum_{j=[\sqrt{T}]+1}^{p} \widehat{\theta}^2_j. \tag{14}$$

The substituting the weight coefficients, minimizing the cost function (13), in (11) leads to the improved model selection procedure, i.e.

$$S^* = S^*_{\gamma^*} \quad \text{and} \quad \gamma^* = \text{argmin}_{\gamma\in\Gamma} J_p(\gamma). \tag{15}$$

It will be noted that γ^* exists because Γ is a finite set. If the minimizing sequence γ^* is not unique, one can take any minimizer. Unlike Pinsker's approach [16], here we do not use the regularity property of the unknown function to find the weights sequence γ^*, i.e. the procedure (15) is adaptive.

Now we study non-asymptotic property of the estimate (15). To this end we assume that

$\mathbf{H}_2$) *The observation frequency p is a function of T, i.e. $p = p(T)$ such that $\sqrt{T} \leq p \leq T$ and for any $\epsilon > 0$*

$$\lim_{T\to\infty} T^{\epsilon-5/6} p = \infty.$$

First, we study the estimate (14).

Proposition 1. *Assume that the conditions $\mathbf{H}_1$) and $\mathbf{H}_2$) hold and the unknown function S has the square integrated derivative $\dot{S}$. Then for $T \geq 3$ and $\sqrt{T} < p \leq T$*

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_T - \sigma_Q| \leq \mathbf{K}_T T^{-1/3} \left(1 + \|\dot{S}\|^2\right), \tag{16}$$

where the term $\mathbf{K}_T > 0$ is slowly increasing as $T \to \infty$.

Using this Proposition, we come to the following sharp oracle inequality for the robust risk of proposed improved model selection procedure.

Theorem 2. *Assume that the conditions* $\mathbf{H}_1) - \mathbf{H}_2)$ *hold and the function* S *has the square integrable derivative* $\dot{S}$*. Then for any* $T \geq 3$ *and* $0 < \rho < 1/2$ *the robust risk* (7) *of estimate* (15) *satisfies the following sharp oracle inequality*

$$\mathcal{R}^*_T(S^*, S) \leq \frac{1+5\rho}{1-\rho} \min_{\gamma \in \Gamma} \mathcal{R}^*_T(S^*_\gamma, S) + \frac{1}{\rho T} \mathbf{U}_T (1 + \|\dot{S}\|^2),$$

where the rest term $\mathbf{U}_T$ *is slowly increasing as* $T \to \infty$*.*

We use the condition $\mathbf{H}_1)$ to construct the special set Γ of weight vectors $(\gamma(j))_{j \geq 1}$ as it is proposed in [4] and [5] for which we will study the asymptotic properties of the model selection procedure (15). For this we consider the following grid

$$\mathcal{A}_T = \{1, \ldots, \mathbf{k}\} \times \{r_1, \ldots, r_m\},$$

where $r_i = i\delta$, $i = \overline{1,\, m}$ with $m = [1/\delta^2]$. We assume that the parameters $\mathbf{k} \geq 1$ and $0 < \delta \leq 1$ are functions of T, i.e. $\mathbf{k} = \mathbf{k}_T$ and $\delta = \delta(T)$, such that

$$\lim_{T \to \infty} \left(\frac{1}{\mathbf{k}_T} + \frac{\mathbf{k}_T}{\ln T} \right) = 0 \quad \text{and} \quad \lim_{T \to \infty} \left(\delta(T) + \frac{1}{T^\epsilon \delta(T)} \right) = 0$$

for any $\epsilon > 0$. One can take, for example,

$$\delta(T) = \frac{1}{\ln(T+1)} \quad \text{and} \quad \mathbf{k}(T) = k_0 + \sqrt{\ln(T+1)},$$

where $k_0 \geq 0$ is a fixed constant. For $\alpha = (\beta, r) \in \mathcal{A}_T$ we define the weights $\gamma_\alpha = (\gamma_\alpha(j))_{j \geq 1}$ as

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

where $j_*(\alpha) = \omega_\alpha / \ln(T+1)$,

$$\omega_\alpha = \left(\frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \, r\, v_T \right)^{1/(2\beta+1)} \quad \text{and} \quad v_T = T/\varsigma^*.$$

Finally, we set

$$\Gamma = \{\gamma_\alpha,\, \alpha \in \mathcal{A}_T\}. \tag{17}$$

Remark 2. It should be noted, that in this case the condition $\mathbf{H}_1)$ holds true with $d = [j_*(\alpha)]$ (see, for example, [15]). Therefore, the model selection procedure (15) with the coefficients (17) satisfies the oracle inequality obtained in Theorem 2.

4 Asymptotic Efficiency

To study the efficiency properties we use the approach proposed by Pinsker in [16], i.e. we assume that the unknown function S belongs to the functional Sobolev ball $W_{k,\mathbf{r}}$ defined as

$$W_{k,\mathbf{r}} = \left\{ f \in \mathcal{C}^{(k)}_{per}[0,1] \, : \, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le \mathbf{r} \right\}, \tag{18}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some unknown parameters, $\mathcal{C}^k_{per}[0,1]$ is the space of k times differentiable 1-periodic $\mathbb{R} \to \mathbb{R}$ functions such that for any $0 \le i \le k-1$ the periodic boundary conditions are satisfied, i.e. $f^{(i)}(0) = f^{(i)}(1)$. It should be noted that the ball $W_{k,\mathbf{r}}$ can be represented as an ellipse in $\mathbb{R}^\infty$ through the Fourier representation in $\mathbf{L}_2[0,1]$ for S, i.e.

$$S = \sum_{j=1}^{\infty} \tau_j \phi_j \quad \text{and} \quad \tau_j = \int_0^1 S(t)\phi_j(t)\mathrm{d}t.$$

In this case we can represent the ball (18)

$$W_{k,\mathbf{r}} = \left\{ f \in \mathbf{L}_2[0,1] \, : \, \sum_{j\ge 1} \mathbf{a}_j \tau_j^2 \le \mathbf{r} \right\}, \tag{19}$$

where $\mathbf{a}_j = \sum_{i=0}^{k} \|\phi^{(i)}\|^2 = \sum_{i=0}^{k} (2\pi[j/2])^{2i}$.

To compare the model selection procedure (15) with all possible estimation methods we denote by Σ_T the set of all estimators $\widehat{S}_T$ based on the observations $(y_{t_j})_{0\le j\le n}$. According to the Pinsker method, firstly one needs to find a lower bound for risks. To this end, we set

$$l_k(\mathbf{r}) \;=\; ((2k+1)\mathbf{r})^{1/(2k+1)} \left(\frac{k}{(\pi(k+1))} \right)^{2k/(2k+1)}. \tag{20}$$

Using this coefficient we obtain the following lower bound.

Theorem 3. *The robust risks (7) are bounded from below as*

$$\liminf_{T\to\infty} v_T^{2k/(2k+1)} \inf_{\widehat{S}_T \in \Sigma_T} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}^*_T(\widehat{S}_T, S) \ge l_k(\mathbf{r}), \tag{21}$$

where $v_T = T/\varsigma^*$.

Remark 3. The lower bound (21) is obtained on the basis of the Van - Trees inequality obtained in [15] for non-Gaussian Lévy processes.

To obtain the upper bound we need the following condition.

$\mathbf{H}_3$) *The parameter* ρ *in the cost function* (13) *is a function of* T, *i.e.* $\rho = \rho(T)$, *such that* $\lim_{T\to\infty} \rho(T) = 0$ *and*

$$\lim_{T\to\infty} T^{\epsilon} \rho(T) = +\infty$$

for any $\epsilon > 0$

Theorem 4. *Assume that the conditions* $\mathbf{H}_2$) – $\mathbf{H}_3$) *hold. Then the model selection procedure* (15) *constructed through the weights* (17) *has the following upper bound*

$$\limsup_{T\to\infty} v_T^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_T^*(S^*, S) \le l_k(\mathbf{r}).$$

It is clear that these theorems imply the following efficient property.

Theorem 5. *Assume that the conditions of Theorems 3 and 4 hold. Then the procedure* (15) *is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} v_T^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_T^*(S^*, S) = l_k(\mathbf{r})$$

and

$$\lim_{T\to\infty} \frac{\inf_{\widehat{S}_T\in\Sigma_T} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_T^*(\widehat{S}_T, S)}{\sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_T^*(S^*, S)} = 1. \tag{22}$$

Remark 4. Note that the parameter (20) defining the lower bound (21) is the well-known Pinsker constant, obtained in [16] for the model (1) with the Gaussian white noise process $(\xi_t)_{t\ge 0}$. For general semimartingale models the lower bound is the same as for the white noise model, but generally the normalization coefficient is not the same. In this case the convergence rate is given by $\left(T/\varsigma_T^*\right)^{-2k/(2k+1)}$ while in the white noise model the convergence rate is $(T)^{-2k/(2k+1)}$. So, if the upper variance threshold ς_T^* tends to zero, the convergence rate is better than the classical one, if it tends to infinity, it is worse and, if it is a constant, the rate is the same.

Remark 5. It should be noted that the efficiency property (22) is shown for the procedure (15) without using the Sobolev regularity parameters $\mathbf{r}$ and k, i.e. this procedure is efficient in adaptive setting.

5 Statistical Analysis for the Big Data Model

Now we apply our results for the high dimensional model (1), i.e. we consider this model with the parametric function

$$S(t) = \sum_{j=1}^{q} \beta_j \mathbf{u}_j(t), \tag{23}$$

where the parameter dimension q more than number of observations given in (5), i.e. $q > n$, the functions $(\mathbf{u}_j)_{1\le j\le q}$ are known and orthonormal in $\mathbf{L}_2[0,1]$. In this case we use the estimator (11) to estimate the vector of parameters $\beta = (\beta_j)_{1\le j\le q}$ as

$$\beta_\gamma^* = (\beta_{\gamma,j}^*)_{1\le j\le q} \quad \text{and} \quad \beta_{\gamma,j}^* = (\mathbf{u}_j, S_\gamma^*).$$

Moreover, we use the model selection procedure (15) as

$$\beta^* = (\beta^*_j)_{1\le j\le q} \quad \text{and} \quad \beta^*_j = (\mathbf{u}_j, S^*). \tag{24}$$

It is clear that

$$|\beta^*_\gamma - \beta|^2_q = \sum_{j=1}^{q} (\beta^*_{\gamma,j} - \beta_j)^2 = \|S^*_\gamma - S\|^2$$

and

$$|\beta^* - \beta|^2_q = \|S^* - S\|^2.$$

Therefore, Theorem 2 implies

Theorem 6. *Assume that conditions* $\mathbf{H}_1$) - $\mathbf{H}_2$) *hold and the function* (23) *has the square integrable derivative* $\dot{S}$. *Then for any* $T \ge 3$ *and* $0 < \rho < 1/2$

$$\sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,\beta} |\beta^* - \beta|^2_q \le \frac{1+5\rho}{1-\rho} \min_{\gamma\in\Gamma} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,\beta} |\beta^* - \beta|^2_q + \frac{1}{\rho T} \mathbf{U}_T (1 + \|\dot{S}\|^2),$$

where the term $\mathbf{U}_T$ *is slowly increasing as* $T \to \infty$.

Theorems 3 and 4 imply the efficiency property for the estimate (24) based on the model selection procedure (15) constructed on the weight coefficients (17) and the penalty threshold satisfying the condition $\mathbf{H}_3$).

Theorem 7. *Assume that the conditions* $\mathbf{H}_2$) – $\mathbf{H}_3$) *hold. Then the estimate* (24) *is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} \upsilon_T^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,\beta} |\beta^* - \beta|^2_q = l_k(\mathbf{r}) \tag{25}$$

and

$$\lim_{T\to\infty} \frac{\inf_{\widehat{\beta}_T\in\Xi_T} \sup_{S\in W_{k,\mathbf{r}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,\beta} |\widehat{\beta}_T - \beta|^2_q}{\sup_{S\in W_{k,\mathbf{r}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,\beta} |\beta^* - \beta|^2_q} = 1,$$

where Ξ_T *is the set of all possible estimators for the vector* β.

Remark 6. In the estimator (15) doesn't use the dimension q in (23). Moreover, it can be equal to $+\infty$. In this case it is impossible to use neither LASSO method nor Dantzig selector which are usually applied to similar models (see, for example, [17] and [1]). It should be emphasized also that the efficiency property (25) is shown without using any sparse conditions for the parameters $\beta = (\beta_j)_{1\le j\le q}$ usually assumed for such problems (see, for example, [3]).

6 Monte-Carlo Simulations

In this section we give the results of numerical simulations to assess the performance and improvement of the proposed model selection procedure (15). We simulate the model (1) with 1-periodic functions S of the forms

$$S_1(t) = t\,\sin(2\pi t) + t^2(1-t)\cos(4\pi t) \tag{26}$$

and

$$S_2(t) = \sum_{j=1}^{+\infty} \frac{1}{1+j^3} \sin(2\pi jt) \tag{27}$$

on [0, 1] and the Ornstein – Uhlenbeck – Lévy noise process ξ_t is defined as

$$\mathrm{d}\xi_t = -\xi_t \mathrm{d}t + 0.5\,\mathrm{d}w_t + 0.5\,\mathrm{d}z_t, \quad z_t = \sum_{j=1}^{N_t} Y_j,$$

where N_t is a homogeneous Poisson process of intensity $\lambda = 1$ and $(Y_j)_{j\geq 1}$ is i.i.d. $\mathcal{N}(0,\,1)$ sequence (see, for example, [11]). We use the model selection procedure (15) constructed through the weights (17) in which $\mathbf{k} = 100 + \sqrt{\ln(T+1)}$,

$$r_i = \frac{i}{\ln(T+1)}, \quad m = [\ln^2(T+1)], \quad \rho = \frac{1}{(3+\ln T)^2}$$

and $\varsigma^* = 0.5$ We define the empirical risk as

$$\mathcal{R}(S^*,\,S) = \frac{1}{p}\sum_{j=1}^{p} \widehat{\mathbf{E}}\Delta_T^2(t_j) \quad \text{and} \quad \widehat{\mathbf{E}}\Delta_T^2(t) = \frac{1}{T}\sum_{l=1}^{N} \Delta_{T,l}^2(t),$$

where $\Delta_T(t) = S_T^*(t) - S(t)$ and $\Delta_{T,l}(t) = S_{T,l}^*(t) - S(t)$ is the deviation for the l-th replication. In this example we take $p = T/2$ and $N = 1000$.

Table 1. The sample quadratic risks for different optimal weights

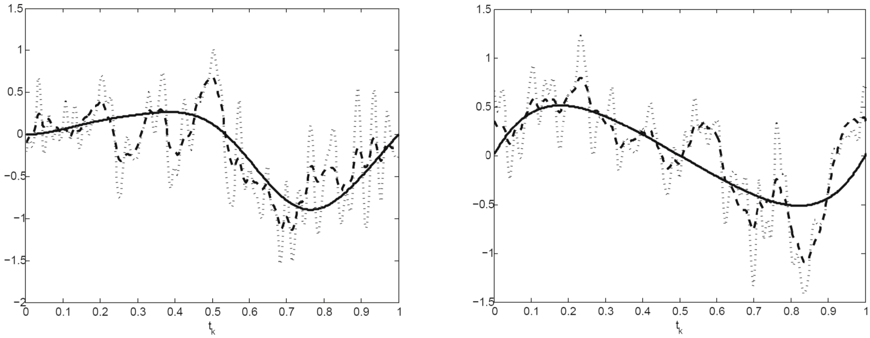
T	200	500	1000	10000
$\mathcal{R}(S^*_{\gamma^*},\,S_1)$	2.8235	0.8454	0.0626	0.0024
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}},\,S_1)$	6.0499	1.8992	0.4296	0.0419
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}},\,S_1)/\mathcal{R}(S^*_{\gamma^*},\,S_1)$	2.1	2.2	6.9	17.7
$\mathcal{R}(S^*_{\gamma^*},\,S_2)$	2.3174	1.0199	0.0817	0.0015
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}},\,S_2)$	7.1047	3.6592	0.8297	0.0299
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}},\,S_2)/\mathcal{R}(S^*_{\gamma^*},\,S_2)$	3.1	3.6	10.2	19.9

Tables 1 and 2 give the sample risks for the improved estimate (15) and the model selection procedure based on the weighted least squares estimates (10) from [11] for different observation period T. One can observe that the improvement increases as T increases for the both models (26) and (27).

Remark 7. The figures show the behavior of the procedures (10) and (11) in the depending on the observation time T. The continuous lines are the functions (26) and (27), the dotted lines are the model selection procedures based on the least squares estimates $\widehat{S}$ and the dashed lines are the improved model selection

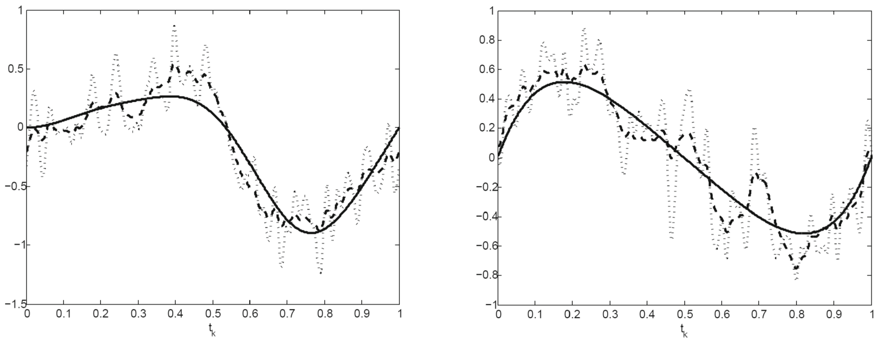
Table 2. The sample quadratic risks for the same optimal weights

T	200	500	1000	10000
$\mathcal{R}(S^*_{\widehat{\gamma}}, S_1)$	3.2017	0.9009	0.1284	0.0076
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}}, S_1)$	6.0499	1.8992	0.4296	0.0419
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}}, S_1)/\mathcal{R}(S^*_{\widehat{\gamma}}, S_1)$	1.9	2.1	3.3	5.5
$\mathcal{R}(S^*_{\widehat{\gamma}}, S_2)$	4.1586	1.9822	0.1032	0.0036
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}}, S_2)$	7.1047	3.6592	0.8297	0.0299
$\mathcal{R}(\widehat{S}_{\widehat{\gamma}}, S_2)/\mathcal{R}(S^*_{\widehat{\gamma}}, S_2)$	1.7	1.8	8.0	8.3



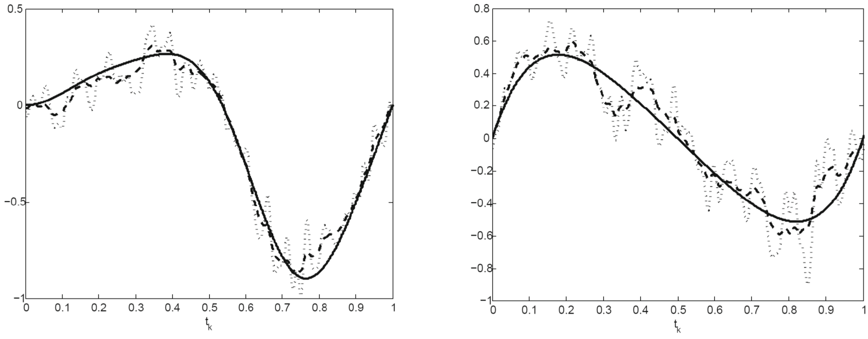
a) b)

Fig. 1. Behavior of the regression functions and their estimates for $T = 200$ (a) – for the function S_1 and b) – for the function S_2).



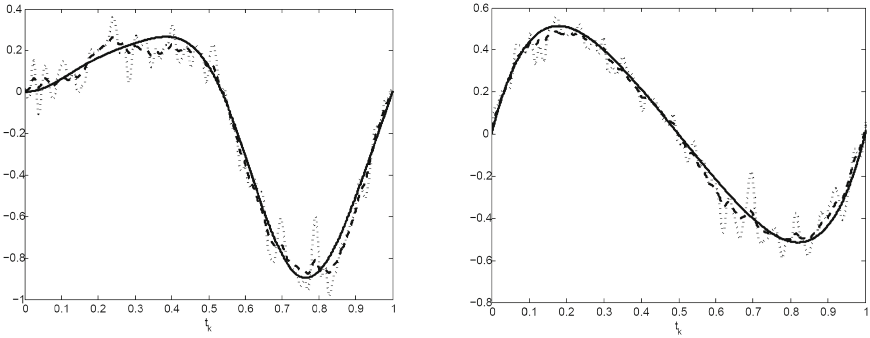
a) b)

Fig. 2. Behavior of the regressions function and their estimates for $T = 500$ (a) – for the function S_1 and b) – for the function S_2).



a) b)

Fig. 3. Behavior of the regression functions and their estimates for $T = 1000$ (a) – for the function S_1 and b) – for the function S_2).



a) b)

Fig. 4. Behavior of the regression functions and their estimates for $T = 10000$ (a) – for the function S_1 and b) – for the function S_2).

procedures S^*. From the Table 2 for the same γ with various observations numbers T we can conclude that theoretical result on the improvement effect (12) is confirmed by the numerical simulations. Moreover, for the proposed shrinkage procedure, from the Table 1 and Figs. 1, 2, 3 and 4, one can be noted that the gain is significant for finite periods T.

7 Conclusion

In the conclusion we would like emphasized that in this paper we studied the following issues:

– we considered the nonparametric estimation problem for continuous time regression model (1) with the noise defined by non-Gaussian Ornstein–Uhlenbeck process with unknown distribution under the condition that this process can be observed only in the fixed discrete time moments (5);

– we proposed adaptive robust improved estimation method via model selection approach and we developed new analytical tools to provide the improvement effect for the non-asymptotic estimation accuracy. It turns out that in this case the accuracy improvement is much more significant than for parametric models, since according to the well-known James–Stein result [8] the accuracy improvement increases when dimension of the parameters increases. It should be noted, that for the parametric models this dimension is always fixed, while for the nonparametric models it tends to infinity, that is, it becomes arbitrarily large with an increase in the number of observations. Therefore, the gain from the application of improved methods is essentially increasing with respect to the parametric case;
– we found constructive conditions for the observation frequency under which we shown sharp non-asymptotic oracle inequalities for the robust risks (7). Then, through the obtained oracle inequalities we provide the efficiency property for the developed model selection methods in adaptive setting, i.e. when the regularity of regression function is unknown;
– we apply the developed model selection procedure to the estimation problem for the Big Data model in continuous time without using the parameter dimension and without assuming any sparse condition for the model parameters ;
– finally, we give Monte - Carlo simulations which confirm the obtained theoretical results.

Acknowledgments. This research was supported by RSF, project no. 20-61-47043.

References

1. Candés, E., Tao, T.: The Dantzig selector: statistical estimation when p is much larger than n. Ann. Stat. **35**, 2313–2351 (2007). https://doi.org/10.1214/009053606000001523
2. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. Chapman & Hall, London (2004)
3. Fan, J., Fan, Y., Barut, E.: Adaptive robust variable selection. Ann. Stat. **42**, 321–351 (2014). https://doi.org/10.1214/13-AOS1191
4. Galtchouk, L., Pergamenshchikov, S.: Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. J. Nonparametr. Stat. **21**, 1–18 (2009). https://doi.org/10.1080/10485250802504096
5. Galtchouk, L., Pergamenshchikov, S.: Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. J. Korean Stat. Soc. **38**, 305–322 (2009). https://doi.org/10.1016/J.JKSS.2008.12.001
6. Höpfner, R., Kutoyants, Y.A.: On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion. Stat. Decis. **27**, 309–326 (2009)
7. Ibragimov, I.A., Khasminskii, R.Z.: Statistical Estimation: Asymptotic Theory. Springer, New York (1981). https://doi.org/10.1007/978-1-4899-0027-2
8. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability, vol. 1, pp. 361–380. University of California Press (1961)

9. Kassam, S.A.: Signal Detection in Non-Gaussian Noise. Springer, New York (1988). https://doi.org/10.1007/978-1-4612-3834-8
10. Konev, V.V., Pergamenshchikov, S.M.: General model selection estimation of a periodic regression with a Gaussian noise. Ann. Inst. Stat. Math. **62**, 1083–1111 (2010). https://doi.org/10.1007/s10463-008-0193-1
11. Konev, V.V., Pergamenshchikov, S.M.: Robust model selection for a semimartingale continuous time regression from discrete data. Stoch. Proc. Appl. **125**, 294–326 (2015). https://doi.org/10.1016/j.spa.2014.08.003
12. Konev, V., Pergamenshchikov, S., Pchelintsev, E.: Estimation of a regression with the pulse type noise from discrete data. Theory Probab. Appl. **58**, 442–457 (2014). https://doi.org/10.1137/s0040585x9798662x
13. Kutoyants, Y.A.: Identification of Dynamical Systems with Small Noise. Kluwer Academic Publishers Group, Berlin (1994)
14. Pchelintsev, E.: Improved estimation in a non-Gaussian parametric regression. Stat. Inference Stoch. Process. **16**, 15–28 (2013). https://doi.org/10.1007/s11203-013-9075-0
15. Pchelintsev, E.A., Pchelintsev, V.A., Pergamenshchikov, S.M.: Improved robust model selection methods for a Lévy nonparametric regression in continuous time. J. Nonparametr. Stat. **31**, 612–628 (2019). https://doi.org/10.1080/10485252.2019.1609672
16. Pinsker, M.S.: Optimal filtration of square integrable signals in Gaussian white noise. Probl. Transm. Inf. **17**, 120–133 (1981)
17. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B. **58**, 267–288 (1996). https://doi.org/10.1111/J.2517-6161.1996.TB02080.X