UDC 519.2 MSC 62G08; 62G05
DOI 10.17223/19988621/58/2

E.A. Pchelintsev, S.M. Pergamenshchikov

IMPROVED MODEL SELECTION METHOD FOR AN ADAPTIVE ESTIMATION IN SEMIMARTINGALE REGRESSION MODELS[1]

This paper considers the problem of robust adaptive efficient estimating of a periodic function in a continuous time regression model with the dependent noises given by a general square integrable semimartingale with a conditionally Gaussian distribution. An example of such noise is the non-Gaussian Ornstein–Uhlenbeck–Lévy processes. An adaptive model selection procedure, based on the improved weighted least square estimates, is proposed. Under some conditions on the noise distribution, sharp oracle inequality for the robust risk has been proved and the robust efficiency of the model selection procedure has been established. The numerical analysis results are given.

Key words: *improved non-asymptotic estimation*, *least squares estimates*, *robust quadratic risk*, *non-parametric regression*, *semimartingale noise*, *Ornstein–Uhlenbeck–Lévy process*, *model selection*, *sharp oracle inequality*, *asymptotic efficiency.*

1. Introduction

Consider a regression model in continuous time

$$dy_t = S(t)\,dt + d\xi_t, \quad 0 \le t \le n, \tag{1.1}$$

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function, $S \in \mathbf{L}_2[0,1]$, $(\xi_t)_{0\le t\le n}$ is an unobservable noise which is a square integrated semimartingale with the values in the Skorokhod space $\mathbf{D}[0,n]$ such that, for any function f from $\mathbf{L}_2[0,1]$, the stochastic integral

$$I_n(f) = \int_0^n f(s)\,d\xi_s \tag{1.2}$$

has the following properties

$$\mathbf{E}_Q I_n(f) = 0 \ \text{ and } \ \mathbf{E}_Q I_n^2(f) \le \kappa_Q \int_0^n f^2(s)\,ds\,. \tag{1.3}$$

Here $\mathbf{E}_Q$ denotes the expectation with respect to the distribution Q of the noise process $(\xi_t)_{0\le t\le n}$ on the space $\mathbf{D}[0,n]$, $\kappa_Q > 0$ is some positive constant depending on the distribution Q. The noise distribution Q is unknown and assumed to belong to some probability family Q_n specified below. Note that the semimartingale regression models in continuous time were introduced by Konev and Pergamenshchikov in [8, 9] for the signal estimation problems. It should be noted also that the class of the noise processes

[1] This work is supported by RSF, Grant no 17-11-01049.

$(\xi_t)_{t\ge 0}$ satisfying conditions (1.3) is rather wide and comprises, in particular, the Lévy processes which are used in different applied problems (see [2], for details). Moreover, as is shown in Section 2, non-Gaussian Ornstein–Uhlenbeck-based models enter this class.

The problem is to estimate the unknown function S in the model (1.1) on the basis of observations $(y_t)_{0\le t\le n}$. In this paper we use the quadratic risk, i.e. for any estimate $\hat{S}$ we set

$$\mathcal{R}_Q\left(\hat{S}_n, S\right) := \mathbf{E}_{Q,S}\left\|\hat{S}_n - S\right\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t)\,dt\,, \tag{1.4}$$

where $\mathbf{E}_{Q,S}$ stands the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process in (1.1) 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

$$\mathcal{R}^*\left(\hat{S}_n, S\right) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q\left(\hat{S}_n, S\right). \tag{1.5}$$

The goal of this paper is to develop the adaptive robust efficient model selection method for the regression (1.1) with dependent noises having conditionally Gaussian distribution using the improved estimation approach. This paper proposes the shrinkage least squares estimates which enable us to improve the non-asymptotic estimation accuracy. For the first time such idea was proposed by Fourdrinier and Pergamenshchikov in [4] for regression models in discrete time and by Konev and Pergamenshchikov in [10] for Gaussian regression models in continuous time. We develop these methods for the general semimartingale regression models in continuous time. It should be noted that for the conditionally Gaussian regression models we cannot use the well-known improved estimators proposed in [7] for Gaussian or spherically symmetric observations. To apply the improved estimation methods to the non-Gaussian regression models in continuous time one needs to use the modifications of the well-known James – Stein estimators proposed in [13, 14] for parametric estimation problems and developed in [16, 18]. We develop the new analytical tools which allow one to obtain the sharp non-asymptotic oracle inequalities for robust risks under general conditions on the distribution of the noise in the model (1.1). This method enables us to treat both the cases of dependent and independent observations from the same standpoint, it does not assume the knowledge of the noise distribution and leads to the efficient estimation procedure with respect to the risk (1.5). The validity of the conditions, imposed on the noise in the equation (1.1) is verified for a non-Gaussian Ornstein–Uhlenbeck process.

The rest of the paper is organized as follows. In the next Section 2, we describe the Ornstein–Uhlenbeck process as the example of a semimartingale noise in the model (1.1). In Section 3 we construct the shrinkage weighted least squares estimates and study the improvement effect. In Section 4 we construct the model selection procedure on the basis of improved weighted least squares estimates and state the main results in the form of oracle inequalities for the quadratic risk (1.4) and the robust risk (1.5). In Section 5 it is shown that the proposed model selection procedure for estimating S in (1.1) is asymptotically efficient with respect to the robust risk (1.5). In Section 6 we illustrate the performance of the proposed model selection procedure through numerical simulations. Section 7 gives the proofs of the main results.

2. Ornstein–Uhlenbeck–Lévy process

Now we consider the noise process $(\xi_t)_{t\geq 0}$ in (1.1) defined by a non-Gaussian Ornstein–Uhlenbeck process with the Lévy subordinator. Such processes are used in the financial Black–Scholes type markets with jumps (see, for example, [1], and the references therein). Let the noise process in (1.1) obey the equations

$$d\xi_t = a\xi_t dt + du_t,\ \ \xi_0 = 0, \tag{2.1}$$

$$u_t = \varrho_1 w_t + \varrho_2 z_t \ \text{ and } \ z_t = x*(\mu-\tilde{\mu})_t, \tag{2.2}$$

where $(w_t)_{t\geq 0}$ is a standard Brownian motion, $\mu(ds\,dx)$ is a jump measure with deterministic compensator $\tilde{\mu}(ds\,dx) = ds\,\Pi(dx)$, $\Pi(\cdot)$ is a Lévy measure, i.e. some positive measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$ (see, for example, in [3]), such that

$$\Pi\left(x^2\right) = 1 \ \text{ and } \ \Pi\left(x^8\right) < \infty. \tag{2.3}$$

We use the notation $\Pi\left(|x|^m\right) = \int_{\mathbb{R}_*} |y|^m\, \Pi(dy)$. Note that the Lévy measure $\Pi(\mathbb{R}_*)$ could be equal to $+\infty$. We use * for the stochastic integrals with respect to random measures, i.e.

$$x*(\mu-\tilde{\mu})_t = \int_0^t \int_{\mathbb{R}_*} y(\mu-\tilde{\mu})(ds,dy).$$

Moreover, we assume that the nuisance parameters $a\leq 0$, ϱ_1 and ϱ_2 satisfy the conditions

$$-a_{\max} \leq a \leq 0\,,\ 0 < \underline{\varrho} \leq \varrho_1^2 \ \text{ and } \ \sigma_Q = \varrho_1^2 + \varrho_2^2 \leq \varsigma^*, \tag{2.4}$$

where the bounds $a_{\max}$, $\underline{\varrho}$ and ς^* are functions of *n*, i.e. $a_{\max} = a_{\max}(n)$, $\underline{\varrho} = \underline{\varrho}_n$ and $\varsigma^* = \varsigma_n^*$ such that for any $\delta > 0$

$$\lim_{n\to\infty} n^{-\delta} a_{\max}(n) = 0\,,\ \liminf_{n\to\infty} n^{\delta}\underline{\varrho}_n > 0 \ \text{ and } \ \lim_{n\to\infty} n^{-\delta}\varsigma_n^* = 0\,. \tag{2.5}$$

We denote by Q_n the family of all distributions of process (1.1) – (2.1) on the Skorokhod space $\mathbf{D}[0,n]$ satisfying the conditions (2.4) and (2.5). It should be noted that, in view of Corollary 7.2 in [17], the condition (1.3) for the process (2.1) holds with $\kappa_Q = 2\sigma_Q$.

Note also that the process (2.1) is conditionally Gaussian square integrable semimartingale with respect to σ-algebra $\mathcal{G} = \sigma\{z_t,\ t\geq 0\}$ which is generated by the jump process $(z_t)_{t\geq 0}$ defined in (2.2.).

3. Improved estimation

For estimating the unknown function *S* in (1.1) we will consider it's Fourier expansion. Let $\left(\phi_j\right)_{j\geq 1}$ be an orthonormal basis in $\mathbf{L}_2[0,1]$. We extend these functions periodically on $\mathbb{R}$, i.e. $\phi_j(t) = \phi_j(t+1)$ for any $t\in\mathbb{R}$.

B$_1$) *Assume that the basis functions are uniformly bounded, i.e. for some constant* $\phi^* \geq 1$, *which may be depend on n*,

$$\sup_{1\le j\le n}\sup_{0\le t\le 1} |\phi_j(t)| \le \phi^* < \infty . \qquad (3.1)$$

B$_2$) *Assume that there exist some* $d_0 \geq 7$ *and* $\check{a} \geq 1$ *such that*

$$\sup_{d\ge d_0} \frac{1}{d}\int_0^1 \Phi^*_d(v)dv \le \check{a}\,,\quad \Phi^*_d(v) = \max_{t\ge v}\left|\sum_{j=1}^{d} \phi_j(t)\phi_j(t-v)\right| . \qquad (3.2)$$

For example, we can take the trigonometric basis defined as $\mathrm{Tr}_1 \equiv 1$, $\mathrm{Tr}_j(t) = \sqrt{2}\cos(\omega_j t)$ for even j and $\mathrm{Tr}_j(t) = \sqrt{2}\sin(\omega_j t)$ for odd $j \geq 2$, where the frequency $\omega_j = 2\pi[j/2]$ and $[x]$ denotes integer part of x. As is shown in Lemma A1 in [17], these functions satisfy the condition **B$_2$)** with $d_0 = \inf\{d \ge 7 : 5+\ln d \le d\}$ and $\check{a} = (1-\mathrm{e}^{-a_{\max}})/(4a_{\max})$.

We write the Fourier expansion of the unknown function S in the form

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t),$$

where the corresponding Fourier coefficients

$$\theta_j = (S,\phi_j) = \int_0^1 S(t)\phi_j(t)\,dt \qquad (3.3)$$

can be estimated as

$$\hat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)dy_t\,. \qquad (3.4)$$

We replace the differential $S(t)dt$ by the stochastic observed differential dy_t. In view of (1.1), one obtains

$$\hat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,,\quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j) \qquad (3.5)$$

and $I_n(\phi_j)$ is given in (1.2). As in [11], we define a class of weighted least squares estimates for $S(t)$ as

$$\hat{S}_\gamma = \sum_{j=1}^{n} \gamma(j)\hat{\theta}_{j,n}\phi_j \qquad (3.6)$$

with the weights $\gamma = (\gamma(j))_{1\le j\le n} \in \mathbb{R}^n$ which belong to some finite set Γ from $[0,1]^n$. We put

$$\nu = \mathrm{card}(\Gamma) \text{ and } |\Gamma|_* = \max_{\gamma\in\Gamma}\sum_{j=1}^{n}\gamma(j)\,, \qquad (3.7)$$

where $\mathrm{card}(\Gamma)$ is the number of the vectors γ in Γ. In the sequel we assume that all vectors from Γ satisfies the following condition.

D$_1$) *Assume that for any vector* $\gamma \in \Gamma$ *there exists some fixed integer* $d = d(\gamma)$ *such that the first d components of the vector are equal to one, i.e.* $\gamma(j) = 1$ for $1 \le j \le d$ *for any* $\gamma \in \Gamma$.

D$_2$) *There exists* $n_0 \ge 1$ *such that for any* $n \ge n_0$ *there exists a* σ*-field* $\mathcal{G}_n$ *for which the random vector* $\tilde{\xi}_{d,n} = \left(\xi_{j,n}\right)_{1\le j\le d}$ *is the* $\mathcal{G}_n$*-conditionally Gaussian in* $\mathbb{R}^d$ *with the covariance matrix*

$$\mathbf{G}_n = \left(\mathbf{E}(\xi_{i,n}, \xi_{j,n} \mid \mathcal{G}_n)\right)_{1\le i,j\le d} \tag{3.8}$$

and for some nonrandom constant $l_n^* > 0$

$$\inf_{Q\in\mathcal{Q}_n} \left(tr\mathbf{G}_n - \lambda_{\max}(\mathbf{G}_n)\right) \ge l_n^* \ a.s., \tag{3.9}$$

where $\lambda_{\max}(A)$ *is the maximal eigenvalue of the matrix A.*

As is shown in Proposition 7.11 in [17], the condition **D**$_2$) holds for the non-Gaussian Ornstein–Uhlenbeck-based model (1.1) – (2.1) with $l_n^* = \underline{\varrho}_n (d-6)/2$ and $d \ge d_0$.

Further we will use the improved estimation method proposed for parametric models in [14] for the first *d* Fourier coefficients in (3.5). To this end we set $\tilde{\theta}_n = (\hat{\theta}_{j,n})_{1\le j\le d}$. In the sequel we will use the norm $|x|_d^2 = \sum_{j=1}^{d} x_j^2$ for any vector $x = \left(x_j\right)_{1\le j\le d}$ from $\mathbb{R}^d$. Now we define the shrinkage estimators as

$$\theta_{j,n}^* = (1 - g(j))\hat{\theta}_{j,n}, \tag{3.10}$$

with $g(j) = \left(\mathbf{c}_n / |\tilde{\theta}_n|_d\right)\mathbf{1}_{\{1\le j\le d\}}$; $\mathbf{1}_A$ is the indicator of the set *A*,

$$\mathbf{c}_n = \frac{l_n^*}{\left(r_n^* + \sqrt{d\kappa_*/n}\right)n}, \quad \kappa_* = \sup_{Q\in\mathcal{Q}_n} \kappa_Q.$$

The positive parameter r_n^* is such that

$$\lim_{n\to\infty} r_n^* = \infty \text{ and } \lim_{n\to\infty} n^{-\delta} r_n^* = 0 \tag{3.11}$$

for any $\delta > 0$.

Now we introduce a new class of shrinkage weighted least squares estimates for *S* as

$$S_\gamma^* = \sum_{j=1}^{n} \gamma(j)\theta_{j,n}^* \phi_j. \tag{3.12}$$

Let $\Delta_Q(S) := \mathcal{R}_Q\left(S_\gamma^*, S\right) - \mathcal{R}_Q\left(\hat{S}_\gamma, S\right)$ denote the difference of quadratic risks of the estimates (3.12) and (3.6).

Theorem 3.1. *Assume that the conditions* **D**$_1$) – **D**$_2$) *hold. Then for any* $n \ge n_0$

$$\sup_{Q\in\mathcal{Q}_n} \sup_{\|S\|\le r_n^*} \Delta_Q(S) < -\mathbf{c}_n^2. \tag{3.13}$$

Remark 3.1. The inequality (3.13) shows that for any $n \ge n_0$ the estimate (3.12) outperforms non-asymptotically the estimate (3.6) in mean square accuracy.

4. Model selection

This Section gives the construction of a model selection procedure for estimating a function S in (1.1) on the basis of improved weighted least square estimates and states the sharp oracle inequality for the robust risk of proposed procedure.

The model selection procedure for the unknown function S in (1.1) will be constructed on the basis of a family of estimates $\left(S^*_\gamma\right)_{\gamma\in\Gamma}$.

The performance of any estimate S^*_γ will be measured by the empirical squared error

$$\mathrm{Err}_n(\lambda) = \left\|S^*_\lambda - S\right\|^2 .$$

In order to obtain a good estimate, we have to formulate the rule for choosing a weight vector $\gamma \in \Gamma$ in (3.12). It is obvious, that the best way is to minimize the empirical squared error with respect to γ. Making use of the estimate definition (3.12) and the Fourier transformation of S imply

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n} \gamma^2(j)\left(\theta^*_{j,n}\right)^2 - 2\sum_{j=1}^{n}\gamma(j)\theta^*_{j,n}\theta_j + \sum_{j=1}^{n}\theta_j^2 . \tag{4.1}$$

Since the Fourier coefficients $\left(\theta_j\right)_{j\ge1}$ are unknown, the weight coefficients $\left(\gamma_j\right)_{j\ge1}$ cannot be found by minimizing this quantity. To circumvent this difficulty one needs to replace the terms $\theta^*_{j,n}\theta_j$ by their estimators $\tilde{\vartheta}_{j,n}$. We set

$$\tilde{\vartheta}_{j,n} = \theta^*_{j,n}\hat{\theta}_{j,n} - \frac{\hat{\sigma}_n}{n} , \tag{4.2}$$

where $\hat{\sigma}_n$ is the estimate for the noise variance of $\sigma_Q = \mathbf{E}_Q\xi^2_{j,n}$ which we choose in the following form

$$\hat{\sigma}_n = \sum_{j=\left[\sqrt{n}\right]+1}^{n} \hat{t}^2_{j,n} \quad \text{and} \quad \hat{t}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\,dy_t . \tag{4.3}$$

For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\gamma) = \sum_{j=1}^{n} \gamma^2(j)\left(\theta^*_{j,n}\right)^2 - 2\sum_{j=1}^{n}\gamma(j)\tilde{\vartheta}_{j,n} + \rho\hat{P}_n(\gamma) , \tag{4.4}$$

where ρ is some positive constant, $\hat{P}_n(\gamma)$ is the penalty term defined as

$$\hat{P}_n(\gamma) = \frac{\hat{\sigma}_n |\gamma|^2_n}{n} . \tag{4.5}$$

Substituting the weight coefficients, minimizing the cost function

$$\gamma^* = \mathrm{agrmin}_{\gamma\in\Gamma} J_n(\gamma) \tag{4.6}$$

in (3.12) leads to the improved model selection procedure

$$S^* = S^*_{\lambda^*} . \tag{4.7}$$

It will be noted that γ^* exists because Γ is a finite set. If the minimizing sequence γ^* in (4.6) is not unique, one can take any minimizer.

To prove the sharp oracle inequality, the following conditions will be needed for the family Q_n of distributions of the noise $(\xi_t)_{t\ge 0}$ in (1.1). Namely, we need to impose some stability conditions for the noise Fourier transform sequence $\left(\xi_{j,n}\right)_{1\le j\le n}$ introduced in [15].

C₁) *There exists a proxy variance* $\sigma_Q > 0$ *such that for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{\mathbf{L}_{1,n}(Q)}{n^{\varepsilon}} = 0\,, \; \mathbf{L}_{1,n}(Q) = \sum_{j=1}^{n} \left| \mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q \right| . \tag{4.8}$$

C₂) *Assume that for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{\mathbf{L}_{2,n}(Q)}{n^{\varepsilon}} = 0\,, \; \mathbf{L}_{2,n}(Q) = \sup_{|x|\le 1} \mathbf{E}_Q \left(\sum_{j=1}^{n} x_j \left(\xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2 \right) \right)^2 . \tag{4.9}$$

Theorem 4.1. *If the conditions* **C₁**) *and* **C₂**) *hold for the distribution Q of the process* $(\xi_t)_{t\ge 0}$ *in* (1.1), *then, for any* $n \ge 1$ *and* $0 < \rho < 1/2$, *the risk* (1.4) *of estimate* (4.7) *for S satisfies the oracle inequality*

$$\mathcal{R}_Q\left(S^*, S\right) \le \frac{1+5\rho}{1-\rho} \min_{\gamma\in\Gamma} \mathcal{R}_Q\left(S_\gamma^*, S\right) + \frac{\mathbf{B}_n(Q)}{\rho n}\,, \tag{4.10}$$

where $\mathbf{B}_n(Q) = \mathbf{U}_n(Q)\left(1 + |\Gamma|_* \, \mathbf{E}_Q \, |\hat{\sigma}_n - \sigma_Q|\right)$ *and the coefficient* $\mathbf{U}_n(Q)$ *is such that for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{\mathbf{U}_n(Q)}{n^{\varepsilon}} = 0\,. \tag{4.11}$$

In the case, when the value of σ_Q in **C₁**) is known, one can take $\hat{\sigma}_n = \sigma_Q$. Then

$$P_n(\gamma) = \frac{\sigma_Q |\gamma|_n^2}{n}\,, \tag{4.12}$$

and we can rewrite the oracle inequality (4.10) with $\mathbf{B}_n(Q) = \mathbf{U}_n(Q)$. Now we study the estimate (4.3). To obtain the oracle inequality for the robust risk (1.5) we need some additional condition on the distribution family Q_n. We set

$$\varsigma^* = \varsigma_n^* = \sup_{Q\in Q_n} \sigma_Q\,. \tag{4.13}$$

$\mathbf{C}_1^*$) *Assume that the limit equations* (4.8) – (4.9) *hold uniformly in* $Q \in Q_n$ *and* $\varsigma_n^* / n^{\varepsilon} \to 0$ *as* $n \to \infty$ *for any* $\varepsilon > 0$.

Now we impose some conditions on the set of the weight coefficients Γ.

$\mathbf{C}_2^*$) *Assume that the set* Γ *is such that* $\nu / n^{\varepsilon} \to 0$ *and* $|\Gamma|_* / n^{1/2+\varepsilon} \to 0$ *as* $n \to \infty$ *for any* $\varepsilon > 0$.

As is shown in [17], both conditions $\mathbf{C}_1^*$) and $\mathbf{C}_2^*$) hold for the model (1.1) with Ornstein–Uhlenbeck noise process (2.1). Using Proposition 4.2 from [17] we can obtain the following result.

Theorem 4.2. *Assume that the conditions* $\mathbf{C}_1^*$) *and* $\mathbf{C}_2^*$) *hold and the function S(t) is continuously differentiable. Then the robust risk* (1.5) *of the estimate* (4.7) *satisfies the oracle inequality, for any* $n \ge 2$ *and* $0 < \rho < 1/2$,

$$\mathcal{R}_n^*\left(S^*, S\right) \le \frac{1+5\rho}{1-\rho} \min_{\gamma \in \Gamma} \mathcal{R}_n^*\left(S_\gamma^*, S\right) + \frac{1}{\rho n} \mathbf{B}_n^*\left(1 + \| \dot{S} \|^2\right),$$

where the term $\mathbf{B}_n^*$ *has the property* (4.11).

Now we specify the weight coefficients $(\gamma(j))_{j \ge 1}$ as proposed in [5, 6] for a heteroscedastic regression model in discrete time. Firstly, we define the normalizing coefficient $v_n = n / \varsigma^*$. Consider a numerical grid of the form

$$\mathcal{A}_n = \left\{1, \ldots, k^*\right\} \times \{r_1, \ldots, r_m\},$$

where $r_i = i\varepsilon$ and $m = \left[1/\varepsilon^2\right]$. We assume that the parameters $k^* \ge 1$ and $0 < \varepsilon \le 1$ are functions of *n*, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that $k^*(n) \to +\infty$, $\varepsilon(n) \to 0$, $k^*(n)/\ln n \to 0$ and $n^\delta \varepsilon(n) \to \infty$ as $n \to \infty$ for any $\delta > 0$. One can take, for example, $\varepsilon(n) = 1/\ln(n+1)$ and $k^*(n) = \sqrt{\ln(n+1)}$. For each $\alpha = (\beta, r) \in \mathcal{A}_n$, we introduce the weight sequence $\gamma_\alpha = (\gamma_\alpha(j))_{j \ge 1}$ as

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

where $\omega_\alpha = \left(\tau_\beta r v_n\right)^{1/(2\beta+1)}$, $\tau_\beta = (\beta+1)(2\beta+1)/\left(\pi^{2\beta}\beta\right)$ and $d(\alpha) = [\omega_\alpha / \ln(n+1)]$. We set

$$\Gamma = \{\gamma_\alpha, \alpha \in \mathcal{A}_n\}. \tag{4.14}$$

It will be noted that such weight coefficients satisfy the condition $\mathbf{D}_1$).

5. Asymptotic efficiency

In order to study the asymptotic efficiency we define the following functional Sobolev ball

$$W_{k,\mathbf{r}} = \left\{ f \in C_p^k[0,1] : \sum_{i=0}^{k} \left\| f^{(i)} \right\|^2 \le \mathbf{r} \right\}, \tag{5.1}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some unknown parameters, $C_p^k[0,1]$ is the space of *k* times differentiable 1-periodic functions such that $f^{(i)}(0) = f^{(i)}(1)$ for any $0 \le i \le k-1$. Let Σ_n denote all estimators $\hat{S}_n$, i.e. measurable functions with respect to $\sigma\{y_t, 0 \le t \le n\}$. In the sequel, we denote by Q^* the distribution of the process $(y_t)_{0 \le t \le n}$ with $\xi_t = \varsigma^* w_t$, i.e. white noise model with the intensity ς^*.

Theorem 5.1. *Assume that* $Q^* \in \mathcal{Q}_n$. *The robust risk* (1.5) *has the following lower bound*

$$\liminf_{n\to\infty} \inf_{\hat{S}_n \in \Sigma_n} v_n^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_n^*\left(\hat{S}_n, S\right) \geq l_k(\mathbf{r}) .$$

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

We show that this lower bound is sharp in the following sense.

Theorem 5.2. *The model selection procedure* (4.7), *with the weight coefficients* (4.14), *satisfies the following upper bound*

$$\limsup_{n\to\infty} v_n^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_n^*\left(S^*, S\right) \leq l_k(\mathbf{r}) .$$

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

Corollary 5.3. *Assume that* $Q^* \in \mathcal{Q}_n$. *Then the model selection procedure* (4.7), *with the weight coefficients* (4.14), *is asymptotically efficient*, *i.e.*

$$\lim_{n\to\infty} v_n^{2k/(2k+1)} \sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}_n^*\left(S^*, S\right) = l_k(\mathbf{r}) .$$

Theorem 5.1 and Theorem 5.2 are proved in the same way as Theorems 1 and 2 in [9].

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 (4.7). We simulate the model (1.1) with 1-periodic function S of the form

$$S(t) = t\sin(2\pi t) + t^2(1-t)\cos(2\pi t) , \quad 0 \leq t \leq 1 , \tag{6.1}$$

and the Ornstein–Uhlenbeck noise process $(\xi_t)_{t\geq 0}$ defined by the equation

$$d\xi_t = -\xi_t dt + 0.5 dw_t + 0.5 dz_t , \quad z_t = \sum\nolimits_{j=1}^{N_t} Y_j ;$$

here N_t is a homogeneous Poisson process with the intensity $\lambda = 1$ and $\left(Y_j\right)_{j\geq 1}$ is i.i.d. Gaussian (0,1) (see, for example, [12]).

We use the model selection procedure (4.7) with the weights (4.14) in which $k^* = 100 + \sqrt{\ln(n+1)}$, $r_i = i/\ln(n+1)$, $m = [\ln^2(n+1)]$, $\varsigma^* = 0.5$ and $\rho = (3+\ln n)^{-2}$. We define the empirical risk as

$$\overline{R}(\tilde{S}, S) = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\Delta_n^2(t_j) \text{ and } \hat{\mathbf{E}}\Delta_n^2(t) = \frac{1}{N}\sum_{l=1}^{N} \Delta_{n,l}^2(t) ,$$

where $\Delta_n(t) = \tilde{S}_n(t) - S(t)$ and $\Delta_{n,l}(t) = \tilde{S}_n^l(t) - S(t)$ is the deviation for the l-th replication. In this example we take the frequency of observations $p = 100001$ and numbers of replications $N = 1000$.

Table 1 gives the values for the sample risks of the improved estimate (4.7) and the model selection procedure based on the weighted LSE (3.15) from [11] for different numbers of observation period n. Table 2 gives the values for the sample risks of the model selection procedure based on the weighted LSE (3.15) from [11] and it's improved version for different numbers of observation period n.

Table 1

The sample quadratic risks for different optimal γ

n	100	200	500	1000
$\overline{R}(S^*_{\gamma^*}, S)$	0.0289	0.0089	0.0021	0.0011
$\overline{R}(\hat{S}_{\hat{\gamma}}, S)$	0.0457	0.0216	0.0133	0.0098
$\overline{R}(\hat{S}_{\hat{\gamma}}, S) / \overline{R}(S^*_{\gamma^*}, S)$	1.6	2.4	6.3	8.9

Table 2

The sample quadratic risks for same optimal $\hat{\gamma}$

n	100	200	500	1000
$\overline{R}(S^*_{\hat{\gamma}}, S)$	0.0391	0.0159	0.0098	0.0066
$\overline{R}(\hat{S}_{\hat{\gamma}}, S)$	0.0457	0.0216	0.0133	0.0098
$\overline{R}(\hat{S}_{\hat{\gamma}}, S) / \overline{R}(S^*_{\hat{\gamma}}, S)$	1.2	1.4	1.3	1.5

Remark 6.1. Figures 1 and 2 show the behaviour of the procedures (3.6) and (4.7) depending on the values of observation periods n. The bold line is the function (6.1), the continuous line is the model selection procedure based on the least squares estimators $\hat{S}$ and the dashed line is the improved model selection procedure S^*. From the Table 2 for the same $\hat{\gamma}$ with various observations numbers n one can conclude that theoretical result on the improvement effect (3.13) is confirmed by the numerical simulations. Moreover, for the proposed shrinkage procedure, Table 1 and Figures 1, 2, we can conclude that the gain is considerable for non-large n.

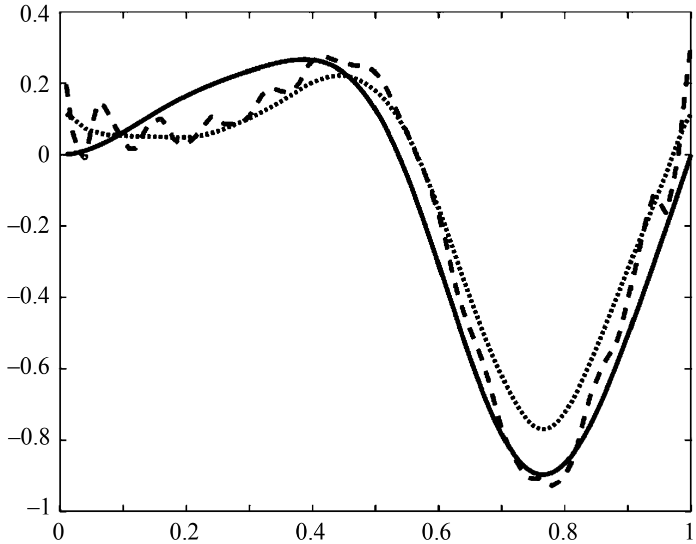

Fig. 1. Behavior of the regression function and its estimates for n = 500

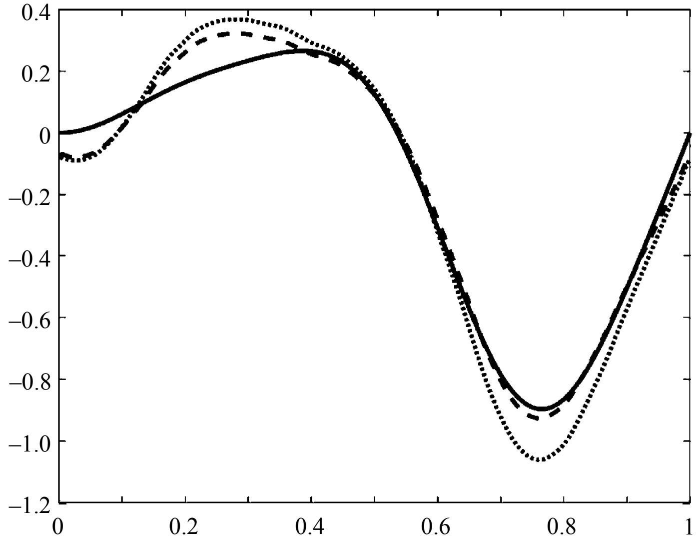

Fig. 2. Behavior of the regression function and its estimates for $n = 1000$

7. Proofs

7.1. Proof of Theorem 3.1. Consider the quadratic error of the estimates (3.12)

$$\|S^*_\gamma - S\|^2 = \sum_{j=1}^{n}\left(\gamma(j)\theta^*_{j,n} - \theta_j\right)^2 = \sum_{j=1}^{d}\left(\gamma(j)\theta^*_{j,n} - \theta_j\right)^2 + \sum_{j=d+1}^{n}\left(\gamma(j)\hat\theta_{j,n} - \theta_j\right)^2$$

$$= \sum_{j=1}^{n}\left(\gamma(j)\hat\theta_{j,n} - \theta_j\right)^2 + c_n^2 - 2c_n \sum_{j=d+1}^{d}\left(\hat\theta_{j,n} - \theta_j\right)\frac{\hat\theta_{j,n}}{\|\tilde\theta_n\|_d}$$

$$= \|\hat S_\gamma - S\|^2 + c_n^2 - 2c_n \sum_{j=d+1}^{d}\left(\hat\theta_{j,n} - \theta_j\right)\iota_j(\tilde\theta_n),$$

where $\iota_j(x) = x_j/\|x\|_d$ for $x = (x_j)_{1\le j\le d} \in \mathbb{R}^d$. Therefore the risk for the improved estimator S^*_γ can be represented as

$$\mathcal{R}_Q\left(S^*_\gamma, S\right) = \mathcal{R}_Q\left(\hat S_\gamma, S\right) + c_n^2 - 2c_n \mathbf{E}_{Q,S}\sum_{j=d+1}^{d}\left(\hat\theta_{j,n} - \theta_j\right)I_{j,n},$$

where $I_{j,n} = \mathbf{E}\left(\iota_j(\tilde\theta_n)\left(\hat\theta_{j,n} - \theta_j\right)|\mathcal{G}_n\right)$. Now, taking into account that the vector $\tilde\theta_n = (\hat\theta_{j,n})_{1\le j\le d}$ is the $\mathcal{G}_n$-conditionally Gaussian in $\mathbb{R}^d$ with mean $\tilde\theta = (\theta_j)_{1\le j\le d}$ and covariance matrix $n^{-1}\mathbf{G}_n$, we obtain

$$I_{j,n} = \int_{\mathbb{R}^d} \iota_j(x)\left(x - \theta_j\right)\mathbf{p}(x \mid \mathcal{G}_n)dx.$$

Here $\mathbf{p}(x \mid \mathcal{G}_n)$ is the conditional distribution density of the vector $\tilde\theta_n$, i.e.

$$\mathbf{p}(x \mid \mathcal{G}_n) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \mathbf{G}_n}} \exp\left(-\frac{(x-\theta)' \mathbf{G}_n^{-1}(x-\theta)}{2} \right).$$

Changing the variables by $u = \mathbf{G}_n^{-1/2}(x-\theta)$ yields

$$I_{j,n} = \frac{1}{(2\pi)^{d/2}} \sum_{l=1}^{d} g_{jl} \int_{\mathbb{R}^d} \tilde{\iota}_{j,n}(u) u_l \exp\left(-\frac{\|u\|_d^2}{2} \right) du, \qquad (7.1)$$

where $\tilde{\iota}_{j,n}(u) = \iota_j\left(\mathbf{G}_n^{1/2} u + \theta\right)$ and g_{ij} denotes (i,j)-th element of $\mathbf{G}_n^{1/2}$. Furthermore, integrating by parts, the integral $I_{j,n}$ can be rewritten as

$$I_{j,n} = \sum_{l=1}^{d}\sum_{k=1}^{d} \mathbf{E}\left(g_{jl} g_{kl} \frac{\partial \iota_j}{\partial u_k}(u)|_{u=\tilde{\theta}_n} \mid \mathcal{G}_n \right).$$

Now, taking into account that $z'Az \le \lambda_{\max}(A)\|z\|^2$ and the condition $\mathbf{D}_2$), we obtain that

$$\Delta_Q(S) = c_n^2 - 2c_n \mathbf{E}_{Q,S}\left(\frac{tr\mathbf{G}_n}{\|\tilde{\theta}_n\|_d} - \frac{\tilde{\theta}_n' \mathbf{G}_n \tilde{\theta}_n}{\|\tilde{\theta}_n\|^3} \right) \le c_n^2 - 2c_n l_n^* n^{-1} \mathbf{E}_{Q,S} \frac{1}{\|\tilde{\theta}_n\|_d}.$$

Recall, that the prime denotes the transposition. Moreover, in view of the Jensen inequality, we can estimate the last expectation from below as

$$\mathbf{E}_{Q,S}\left(\|\tilde{\theta}_n\|_d\right)^{-1} = \mathbf{E}_{Q,S}\left(\|\tilde{\theta} + n^{-1/2}\xi_n\|_d\right)^{-1} \ge \left(\|\theta_n\|_d + n^{-1/2}\mathbf{E}_{Q,S}\|\xi_n\|_d\right)^{-1}.$$

Now we note that the condition (1.3) implies that

$$\mathbf{E}_{Q,S}\|\xi_n\|_d^2 \le \kappa_* d.$$

So, for $\|S\|^2 \le r_n^*$

$$\mathbf{E}_{Q,S}\left(\|\tilde{\theta}_n\|\right)^{-1} \ge \left(r_n^* + \sqrt{d\kappa_*/n}\right)^{-1}$$

and therefore
$$\Delta_Q(S) = c_n^2 - 2c_n \frac{l_n^*}{\left(r_n^* + \sqrt{d\kappa_*/n}\right)n} \le -c_n^2.$$

Hence Theorem 3.1.

7.2. Proof of Theorem 4.1. Substituting (4.4) in (4.1) yields for any

$$\mathrm{Err}_n(\gamma) = J(\gamma) + 2\sum_{j=1}^{n}\gamma(j)\left(\theta_{j,n}^*\hat{\theta}_{j,n} - \frac{\hat{\sigma}_n}{n} - \theta_{j,n}^*\theta_j\right) + \|S\|^2 - \rho\hat{P}_n(\gamma). \qquad (7.2)$$

Now we set $L(\gamma) = \sum_{j=1}^{n}\gamma(j)$,

$$B_{1,n}(\gamma) = \sum_{j=1}^{n}\gamma(j)\left(\mathbf{E}_Q\xi_{j,n}^2 - \sigma_Q\right), \quad B_{2,n}(\gamma) = \sum_{j=1}^{n}\gamma(j)\tilde{\xi}_{j,n},$$

$$M(\gamma) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma(j)\theta_j\xi_{j,n} \quad \text{and} \quad B_{3,n}(\gamma) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma(j)g(j)\hat{\theta}_{j,n}\xi_{j,n}.$$

Taking into account the definition (4.5), we can rewrite (7.2) as

$$\mathrm{Err}_n(\gamma) = J(\gamma) + 2\frac{\sigma_Q - \hat{\sigma}_n}{n} L(\gamma) + 2M(\gamma) + \frac{2}{n} B_{1,n}(\gamma)$$
$$+ 2\sqrt{P_n(\gamma)}\frac{B_{2,n}(\overline{\gamma})}{\sqrt{\sigma_Q n}} - 2B_{3,n}(\gamma) + \|S\|^2 - \rho\hat{P}_n(\gamma) \tag{7.3}$$

with $\overline{\gamma} = \gamma/|\gamma|_n$. Let $\gamma_0 = (\gamma_0(j))_{1\le n}$ be a fixed sequence in Γ and γ^* be as in (4.6). Substituting γ_0 and γ^* in (7.3), we consider the difference

$$\mathrm{Err}_n(\gamma^*) - \mathrm{Err}_n(\gamma_0) \le 2\frac{\sigma_Q - \hat{\sigma}_n}{n} L(x) + 2M(x) + \frac{2}{n} B_{1,n}(x) + 2\sqrt{P_n(\gamma^*)}\frac{B_{2,n}(\overline{\gamma}^*)}{\sqrt{\sigma_Q n}}$$
$$-2\sqrt{P_n(\gamma_0)}\frac{B_{2,n}(\overline{\gamma}_0)}{\sqrt{\sigma_Q n}} - 2B_{3,n}(\gamma^*) + 2B_{3,n}(\gamma_0) - \rho\hat{P}_n(\gamma^*) + \rho\hat{P}_n(\gamma_0),$$

where $x = \gamma^* - \gamma_0$. Note that $L(x) \le 2|\Gamma|_*$ and $|B_{1,n}(x)| \le \mathbf{L}_{1,n}(Q)$. Applying the elementary inequality

$$2|ab| \le \varepsilon a^2 + \varepsilon^{-1} b^2 \tag{7.4}$$

with any $\varepsilon > 0$, we get

$$2\sqrt{P_n(\gamma)}\frac{B_{2,n}(\overline{\gamma})}{\sqrt{\sigma_Q n}} \le \varepsilon P_n(\gamma) + \frac{B_{2,n}^2(\overline{\gamma})}{\varepsilon\sigma_Q n} \le \varepsilon P_n(\gamma) + \frac{B_2^*}{\varepsilon\sigma n},$$

where
$$B_2^* = \max_{\gamma\in\Gamma}\left(B_{2,n}^2(\overline{\gamma}) + B_{2,n}^2(\overline{\gamma}^2)\right)$$

with $\gamma^2 = (\gamma_j^2)_{1\le j\le n}$. Note that from the definition of function $\mathbf{L}_{2,n}(Q)$ in the condition $\mathbf{C}_2)$ it follows that

$$\mathbf{E}_Q B_2^* \le \sum_{\gamma\in\Gamma}\left(\mathbf{E}_Q B_{2,n}^2(\overline{\gamma}) + \mathbf{E}_Q B_{2,n}^2(\overline{\gamma}^2)\right) \le 2\nu\mathbf{L}_{2,n}(Q). \tag{7.5}$$

Moreover, by the same argument one can estimate the term $B_{3,n}$. Note that

$$\sum_{j=1}^n g_\gamma^2(j)\hat{\theta}_j^2 = c_n^2 \le \frac{c_n^*}{n}, \tag{7.6}$$

where $c_n^* = n\max_{\gamma\in\Gamma} c_n^2$. Therefore by the Cauchy–Schwarz inequality, we can estimate the term $B_{3,n}(\gamma)$ as

$$|B_{3,n}(\gamma)| \le \frac{|\gamma|_n}{\sqrt{n}} c_n \left(\sum_{j=1}^n \overline{\gamma}^2(j)\xi_j^2\right)^{1/2} = \frac{|\gamma|_n}{\sqrt{n}} c_n \left(\sigma_Q + B_{2,n}(\overline{\gamma}^2)\right)^{1/2}.$$

So, applying the elementary inequality (7.4) with some arbitrary $\varepsilon > 0$, we have

$$2|B_{3,n}(\gamma)| \le \varepsilon P_n(\gamma) + \frac{c_n^*}{\varepsilon\sigma_Q n}\left(\sigma_Q + B_2^*\right).$$

Using the bounds, one has

$$\mathrm{Err}_n\left(\gamma^*\right) \le \mathrm{Err}_n\left(\gamma_0\right) + \frac{4|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n} L(x) + 2M(x) + \frac{2}{n}\mathbf{L}_{1,n}(Q)$$
$$+\frac{2}{\varepsilon}\frac{c^*}{\sigma_Q n}\left(\sigma_Q + B_2^*\right) + \frac{2}{\varepsilon}\frac{B_2^*}{\sigma_Q n} + 2\varepsilon P_n\left(\gamma^*\right) + 2\varepsilon P_n\left(\gamma_0\right) - \rho\hat{P}_n\left(\gamma^*\right) + \rho\hat{P}_n\left(\gamma_0\right).$$

Setting $\varepsilon = \rho/4$ and substituting $\rho = 1$ (where it is possible) we have

$$\mathrm{Err}_n\left(\gamma^*\right) \le \mathrm{Err}_n\left(\gamma_0\right) + \frac{5|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n} + 2M(x) + \frac{2}{n}\mathbf{L}_{1,n}(Q)$$
$$+\frac{16\left(c^*+1\right)\left(\sigma_Q + B_2^*\right)}{\rho\sigma_Q n} - \frac{\rho}{2}\hat{P}_n\left(\gamma^*\right) + \frac{\rho}{2}P_n\left(\gamma_0\right) + \frac{\rho}{2}\hat{P}_n\left(\gamma_0\right).$$

Moreover, taking into account here that

$$\left|\hat{P}_n\left(\gamma_0\right) - P_n\left(\gamma_0\right)\right| \le \frac{|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n}$$

and that $\rho < 1/2$, we obtain that

$$\mathrm{Err}_n\left(\gamma^*\right) \le \mathrm{Err}_n\left(\gamma_0\right) + \frac{6|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n} + 2M(x) + \frac{2}{n}\mathbf{L}_{1,n}(Q)$$
$$+\frac{16\left(c^*+1\right)\left(\sigma_Q + B_2^*\right)}{\rho\sigma_Q n} - \frac{\rho}{2}P_n\left(\gamma^*\right) + \frac{3\rho}{2}P_n\left(\gamma_0\right). \tag{7.7}$$

Now we estimate the third term in the right-hand side if this inequality. Firstly, we note that

$$2|M(x)| \le \varepsilon\left\|S_x\right\|^2 + \frac{Z^*}{n\varepsilon}, \tag{7.8}$$

where $S_x = \sum_{j=1}^n x_j \theta_j \phi_j$ and

$$Z^* = \sup_{x\in\Gamma_1} \frac{nM^2(x)}{\left\|S_x\right\|^2}.$$

with the set $\Gamma_1 = \Gamma - \gamma_0$. Using Proposition 7.1 from [17], we can obtain that for any fixed $x = (x_j)_{1\le j\le d} \in \mathbb{R}^d$.

$$\mathbf{E}M^2(x) = \frac{\mathbf{E}I_n^2\left(S_x\right)}{n^2} = \frac{\sigma_Q\left\|S_x\right\|^2}{n} = \frac{\sigma_Q}{n}\sum_{j=1}^n x_j^2\theta_j^2 \tag{7.9}$$

and therefore

$$\mathbf{E}_Q Z^* \le \sum_{x\in\Gamma_1} \frac{nM^2(x)}{\left\|S_x\right\|^2} \le \sigma_Q \nu. \tag{7.10}$$

The norm $\left\|S_{\gamma^*}^* - S_{\gamma_0}^*\right\|$ can be estimated from below as

$$\left\|S_{\gamma^*}^* - S_{\gamma_0}^*\right\| = \sum_{j=1}^n \left(x(j) + \beta(j)\right)\hat{\theta}_j^2 \ge \left\|\hat{S}_x\right\|^2 + 2\sum_{j=1}^n x(j)\beta(j)\hat{\theta}_j^2,$$

where $\beta(j)=\gamma_0(j)g_j(\gamma_0)-\gamma(j)g_j(\gamma)$. Therefore, in view of (3.5),

$$\|S_x\|^2-\left\|S^*_\gamma-S^*_{\gamma_0}\right\|^2\leq\|S_x\|^2-\left\|\hat{S}_x\right\|^2-2\sum_{j=1}^n x(j)\beta(j)\hat{\theta}_j^2$$
$$\leq -2M\left(x^2\right)-2\sum_{j=1}^n x(j)\beta(j)\hat{\theta}_j\theta_j-\frac{2}{\sqrt{n}}\Upsilon(x),$$

where $\Upsilon(\gamma)=\sum_{j=1}^n\gamma(j)\beta(j)\hat{\theta}_j\xi_j$. Note that the first term in this inequality can be estimated as

$$2M\left(x^2\right)\leq\varepsilon\|S_x\|^2+\frac{Z_1^*}{n\varepsilon}\quad\text{and}\quad Z_1^*=\sup_{x\in\Gamma_1}\frac{nM^2\left(x^2\right)}{\|S_x\|^2}.$$

Note that, similarly to (7.10) one can estimate the last term as

$$\mathbf{E}_Q Z_1^*\leq\sigma_Q\nu.$$

From here it follows that, for any $0<\varepsilon<1$,

$$\|S_x\|^2\leq\frac{1}{1-\varepsilon}\left(\left\|S^*_\gamma-S^*_{\gamma_0}\right\|^2+\frac{Z_1^*}{n\varepsilon}-2\sum_{j=1}^n x(j)\beta(j)\hat{\theta}_j\theta_j-\frac{2}{\sqrt{n}}\Upsilon(x)\right).\qquad(7.11)$$

Further we note that the property (7.6) yields

$$\sum_{j=1}^n\beta^2(j)\hat{\theta}_j^2\leq 2\sum_{j=1}^n g_\gamma^2(j)\hat{\theta}_j^2+2\sum_{j=1}^n g_{\gamma_0}^2(j)\hat{\theta}_j^2\leq\frac{4c^*}{n\varepsilon}.\qquad(7.12)$$

Given $|x(j)|\leq 1$ and using the inequality (7.4), we get that for any $\varepsilon>0$

$$2\left|\sum_{j=1}^n x(j)\beta(j)\hat{\theta}_j\theta_j\right|\leq\varepsilon\|S_x\|^2+\frac{4c^*}{n\varepsilon}.$$

To estimate the last in the right hand of (7.11) we use first the Cauchy-Schwartz inequality and then the bound (7.12), i.e.

$$\frac{2}{\sqrt{n}}|\Upsilon(\gamma)|\leq\frac{2|\Gamma|_*}{\sqrt{n}}\left(\sum_{j=1}^n\beta^2(j)\hat{\theta}_j^2\right)^{1/2}\left(\sum_{j=1}^n\bar{\gamma}^2(j)\xi_j^2\right)^{1/2}$$
$$\leq\varepsilon P_n(\gamma)+\frac{c^*}{\varepsilon\sigma n}\sum_{j=1}^n\bar{\gamma}^2(j)\xi_j^2\leq\varepsilon P_n(\gamma)+\frac{c^*\left(\sigma_Q+B_2^*\right)}{\varepsilon\sigma_Q n}.$$

Therefore,

$$\frac{2}{\sqrt{n}}|\Upsilon(\gamma)|\leq\frac{2}{\sqrt{n}}\left|\Upsilon\left(\gamma^*\right)\right|+\frac{2}{\sqrt{n}}|\Upsilon(\gamma_0)|\leq\varepsilon P_n\left(\gamma^*\right)+\varepsilon P_n(\gamma_0)+\frac{2c^*\left(\sigma_Q+B_2^*\right)}{\varepsilon\sigma_Q n}.$$

Combining all these bounds in (7.11), we obtain that

$$\|S_x\|^2\leq\frac{1}{1-\varepsilon}\left(\frac{Z_1^*}{n\varepsilon}+\left\|S^*_{\gamma^*}-S^*_{\gamma_0}\right\|^2+\frac{6c_n^*\left(\sigma+B_2^*\right)}{\varepsilon\sigma n}+\varepsilon P_n\left(\gamma^*\right)+\varepsilon P_n(\gamma_0)\right).$$

Using (7.8) and this bound

$$\left\|S^*_{\gamma^*} - S^*_{\gamma_0}\right\|^2 \le 2\left(Err_n\left(\gamma^*\right) + Err_n\left(\gamma_0\right)\right),$$

we have

$$2|M(x)| \le \frac{Z^* + Z_1^*}{n(1-\varepsilon)\varepsilon} + \frac{2\varepsilon\left(Err_n\left(\gamma^*\right) + Err_n(\gamma_0)\right)}{1-\varepsilon} + \frac{6c_n^*\left(\sigma_Q + B_2^*\right)}{n\sigma_Q(1-\varepsilon)} + \frac{\varepsilon^2}{1-\varepsilon}\left(P_n\left(\gamma^*\right) + P_n(\gamma_0)\right).$$

Choosing here $\varepsilon \le \rho/2 < 1/2$ we have that

$$2|M(x)| \le \frac{2\left(Z^* + Z_1^*\right)}{n\varepsilon} + \frac{2\varepsilon\left(Err_n\left(\gamma^*\right) + Err_n(\gamma_0)\right)}{1-\varepsilon} + \frac{12c_n^*\left(\sigma_Q + B_2^*\right)}{n\sigma_Q} + \varepsilon\left(P_n\left(\gamma^*\right) + P_n(\gamma_0)\right).$$

From here and (7.7), it follows that

$$\begin{aligned}\mathrm{Err}_n\left(\gamma^*\right) \le{}& \frac{1+\varepsilon}{1-3\varepsilon}\mathrm{Err}_n(\gamma_0) + \frac{6|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n(1-3\varepsilon)} + \frac{2\mathbf{L}_{1,n}(Q)}{n(1-3\varepsilon)} \\ &+ \frac{28\left(1+c_n^*\right)\left(\sigma_Q + B_2^*\right)}{\rho(1-3\varepsilon)n\sigma_Q} + \frac{2\left(Z^* + Z_1^*\right)}{n(1-3\varepsilon)} + \frac{2\rho P_n(\gamma_0)}{1-3\varepsilon}.\end{aligned}$$

Choosing here $\varepsilon = \rho/3$ and estimating $(1-\rho)^{-1}$ by 2 where this is possible, we get

$$\begin{aligned}\mathrm{Err}_n\left(\gamma^*\right) \le{}& \frac{1+\rho/3}{1-\rho}\mathrm{Err}_n(\gamma_0) + \frac{12|\Gamma|_*\left|\sigma_Q - \hat{\sigma}_n\right|}{n(1-3\varepsilon)} + \frac{4}{n}\mathbf{L}_{1,n}(Q) \\ &+ \frac{56\left(1+c_n^*\right)\left(\sigma_Q + B_2^*\right)}{\rho n\sigma_Q} + \frac{4\left(Z^* + Z_1^*\right)}{n} + \frac{2\rho P_n(\gamma_0)}{1-\rho}.\end{aligned}$$

Taking the expectation and using the upper bound for $P_n(\gamma_0)$ in Lemma 7.1 with $\varepsilon = \rho$ yields

$$\mathcal{R}_Q\left(S^*, S\right) \le \frac{1+5\rho}{1-\rho}\mathcal{R}_Q\left(S^*_{\gamma_0}, S\right) + \frac{\breve{\mathbf{U}}_{Q,n}}{n\rho} + \frac{12|\Gamma|_*\mathbf{E}_Q\left|\sigma_Q - \hat{\sigma}_n\right|}{n},$$

where $\breve{\mathbf{U}}_{Q,n} = 4\mathbf{L}_{1,n}(Q) + 56\left(1+c_n^*\right)\left(2\mathbf{L}_{2,n}(Q)\nu + 1\right) + 2c_n^*$. Since this inequality holds for each $\gamma_0 \in \Lambda$, this implies Theorem 4.1.

7.3. Property of Penalty term

Lemma 7.1. For any $n \ge 1$, $\gamma \in \Gamma$ and $0 < \varepsilon < 1$

$$P_n(\gamma) \le \frac{\mathbf{E}\ \mathrm{Err}_n(\gamma)}{1-\varepsilon} + \frac{c_n^*}{n\varepsilon(1-\varepsilon)}. \tag{7.13}$$

Proof. By the definition of $\mathrm{Err}_n(\gamma)$ one has

$$\begin{aligned}\mathrm{Err}_n(\gamma) &= \sum_{j=1}^n \left(\gamma(j)\theta^*_{j,n} - \theta_j\right)^2 = \sum_{j=1}^n \gamma(j)\left(\left(\theta^*_{j,n} - \theta_j\right) + (\gamma(j)-1)\theta_j\right)^2 \\ &\ge \sum_{j=1}^n \gamma(j)^2\left(\theta^*_{j,n} - \theta_j\right)^2 + \sum_{j=1}^n \gamma(j)(\gamma(j)-1)\theta_j\left(\theta^*_{j,n} - \theta_j\right).\end{aligned}$$

By the condition $\mathbf{B}_2$) and the definition (3.10) we obtain that the last term in the sum can be replaced as

$$\sum_{j=1}^{n}\gamma(j)(\gamma(j)-1)\theta_j\left(\theta^*_{j,n}-\theta_j\right)=\sum_{j=1}^{n}\gamma(j)(\gamma(j)-1)\theta_j\left(\hat\theta_{j,n}-\theta_j\right),$$

i.e. $\mathbf{E}\sum_{j=1}^{n}\gamma(j)(\gamma(j)-1)\theta_j\left(\theta^*_{j,n}-\theta_j\right)=0$ and, therefore, taking into account the definition (4.12), we obtain that

$$\mathrm{Err}_n(\gamma)\ge\sum_{j=1}^{n}\gamma(j)^2\,\mathbf{E}\left(\theta^*_{j,n}-\theta_j\right)^2=\sum_{j=1}^{n}\gamma(j)^2\,\mathbf{E}\left(\frac{\xi_{j,n}}{\sqrt{n}}-g_\gamma(j)\hat\theta_j\right)^2$$

$$\ge P_n(\gamma)-\frac{2}{\sqrt{n}}\mathbf{E}\sum_{j=1}^{n}\gamma(j)^2\,g_\gamma(j)\hat\theta_{j,n}\xi_j\ge(1-\varepsilon)P_n(\gamma)-\frac{1}{\varepsilon}\mathbf{E}\sum_{j=1}^{n}g^2_\gamma(j)\hat\theta^2_j.$$

The inequality (7.6) implies the bound (7.13). Hence Lemma 7.1.

REFERENCES

1. *Barndorff-Nielsen O.E.*, *Shephard N.* Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics // J. Royal Stat. Soc. 2001. B 63. P. 167–241.
2. *Bertoin J.* Lévy Processes. Cambridge: Cambridge University Press, 1996.
3. *Cont R.*, *Tankov P.* Financial Modelling with Jump Processes. London: Chapman & Hall, 2004.
4. *Fourdrinier D.*, *Pergamenshchikov S.* Improved selection model method for the regression with dependent noise // Annals of the Institute of Statistical Mathematics. 2007. V. 59(3). P. 435–464.
5. *Galtchouk L.*, *Pergamenshchikov S.* Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models // Journal of Nonparametric Statistics. 2009. V. 21(1). P. 1–16.
6. *Galtchouk L.*, *Pergamenshchikov S.* Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression // Journal of Korean Statistical Society. 2009. V. 38(4). P. 305–322.
7. *James W.*, *Stein C.* Estimation with quadratic loss // Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability. Berkeley: University of California Press, 1961. V. 1. P. 361–380.
8. *Konev V.V.*, *Pergamenshchikov S.M.* Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities // Tomsk State University Journal of Mathematics and Mechanics. 2009. No. 3(7). P. 23–41.
9. Konev V.V., Pergamenshchikov S.M. Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency // Tomsk State University Journal of Mathematics and Mechanics. 2009. No. 4(8). P. 31–45.
10. *Konev V.V.*, *Pergamenshchikov S.M.* General model selection estimation of a periodic regression with a Gaussian noise // Annals of the Institute of Statistical Mathematics. 2010. V. 62. P. 1083–1111.
11. *Konev V.V.*, *Pergamenshchikov S.M.* Efficient robust nonparametric in a semimartingale regression model // Annals of the Institute of Henri Poincaré. Probab. and Stat. 2012. V. 48(4). P. 1217–1244.
12. *Konev V.V.*, *Pergamenshchikov S.M.* Robust model selection for a semimartingale continuous time regression from discrete data // Stochastic processes and their applications. 2015. V. 125. P. 294–326.
13. *Konev V.*, *Pergamenshchikov S. and Pchelintsev E.* Estimation of a regression with the pulse type noise from discrete data // Theory Probab. Appl. 2014. V. 58(3). P. 442–457.

14. *Pchelintsev E.* Improved estimation in a non-Gaussian parametric regression // Stat. Inference Stoch. Process. 2013. V. 16(1). P. 15–28.
15. *Pchelintsev E.*, *Pergamenshchikov S.* Oracle inequalities for the stochastic differential equations // Stat. Inference Stoch. Process. 2018. V. 21 (2). P. 469–483.
16. *Pchelintsev E.A.*, *Pchelintsev V.A.*, *Pergamenshchikov S.M.* Non asymptotic sharp oracle inequality for the improved model selection procedures for the adaptive nonparametric signal estimation problem // Communications – Scientific Letters of the University of Zilina. 2018. V. 20 (1). P. 72–76.
17. *Pchelintsev E.*, *Pergamenshchikov S.* Adaptive model selection method for a conditionally Gaussian semimartingale regression in continuous time. 2018. P. 1-50. Preprint: http://arxiv.org/abs/1811.05319.
18. *Povzun M.A.*, *Pchelintsev E.A.* Estimating parameters in a regression model with dependent noises // Tomsk State University Journal of Mathematics and Mechanics. 2017. No. 49. P. 43–51.

Received: November 13, 2018

Pchelintsev E. A., Pergamenshchikov S. M. (2019) IMPROVED MODEL SELECTION METHOD FOR AN ADAPTIVE ESTIMATION IN SEMIMARTINGALE REGRESSION MODELS. *Vestnik Tomskogo gosudarstvennogo universiteta. Matematika i mekhanika* [Tomsk State University Journal of Mathematics and Mechanics]. 58. pp. 14–31

AMS Mathematical Subject Classification: 62G08; 62G05

Пчелинцев Е.А., Пергаменщиков С.М. (2019) УЛУЧШЕННЫЙ МЕТОД ВЫБОРА МОДЕЛИ ДЛЯ АДАПТИВНОГО ОЦЕНИВАНИЯ В СЕМИМАРТИНГАЛЬНЫХ РЕГРЕССИОННЫХ МОДЕЛЯХ. *Вестник Томского государственного университета. Математика и механика*. № 58. С. 14–31

DOI 10.17223/19988621/58/2

Ключевые слова: улучшенное неасимптотическое оценивание, оценки наименьших квадратов, робастный квадратический риск, непараметрическая регрессия, семимартингальный шум, процесс Орнштейна – Уленбека – Леви, выбор модели, точное оракульное неравенство, асимптотическая эффективность.

Рассматривается задача робастного адаптивного эффективного оценивания периодической функции в непрерывной модели регрессии с зависимыми шумами, задаваемыми общим квадратично интегрируемым семимартингалом с условно-гауссовским распределением. Примером такого шума являются негауссовские процессы Орнштейна – Уленбека – Леви. Предложена адаптивная процедура выбора модели на основе улучшенных взвешенных оценок наименьших квадратов. При некоторых условиях на распределение шума доказано точно оракульное неравенство для робастного риска и установлена робастная эффективность процедуры выбора модели. Приводятся результаты численного моделирования.

Финансовая поддержка: Работа выполнена при поддержке РНФ, Грант № 17-11-01049.

Благодарности. Работа частично выполнена в рамках государственного задания Минобрнауки № 2.3208.2017/4.6. Работа второго автора частично выполнена в рамках государственного задания Минобрнауки № 1.472.2016/1.4.

PCHELINTSEV Evgeny Anatolievich, Candidate of Physics and Mathematics,
Department of Mathematics and Mechanics, National Research Tomsk State University,
Tomsk, Russia. E-mail: evgen-pch@yandex.ru

PERGAMENSHCHIKOV Serguei Markovich, Doctor of Physics and Mathematics,
Laboratory of Mathematics Raphael Salem, University of Rouen Normandy, France.
E-mail: serge.pergamenchtchikov@univ-rouen.fr