Improved model selection method for an adaptive estimation in semimartingale regression models

E A Pchelintsev, Serguei Pergamenchtchikov

Improved model selection method for an adaptive estimation in semimartingale regression models *

Pchelintsev E.A., † Pergamenshchikov S.M.‡

Abstract

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.

UDC : 519.2

*This work is supported by RSF, Grant no 17-11-01049.

†Department of Mathematical Analysis and Theory of Functions, Tomsk State University, e-mail: evgen-pch@yandex.ru

‡Laboratoire de Mathématiques Raphael Salem, Université de Rouen and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of Tomsk State University, e-mail: Serge.Pergamenchtchikov@univ-rouen.fr

1 Introduction

Consider a regression model in continuous time

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

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function, $S \in \mathbf{L}_2[0,1]$, $(\xi_t)_{t\ge 0}$ 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 $\mathbf{L}_2[0,n]$, the stochastic integral

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

has the following properties

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

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 $\mathcal{D}[0,n]$, $\varkappa_Q > 0$ is some positive constant depending on the distribution Q. The noise distribution Q is unknown and assumed to belong to some probability family $\mathcal{Q}_n$ specified below. Note that the firstly 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 $(\xi_t)_{t\ge 0}$ 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 $\widehat{S}$ we set

$$\mathcal{R}_Q(\widehat{S}, S) := \mathbf{E}_{Q,S}\,\|\widehat{S} - S\|^2 \quad \text{and} \quad \|f\|^2 := \int_0^1 f^2(t)\mathrm{d}t\,, \tag{1.4}$$

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

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

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 can not 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 problems. 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) obeys the equation

$$\mathrm{d}\xi_t = a\xi_t \mathrm{d}t + \mathrm{d}u_t\,, \quad \xi_0 = 0\,, \tag{2.1}$$

where

$$u_t = \varrho_1\, w_t + \varrho_2\, z_t \quad \text{and} \quad z_t = \int_{\mathbb{R}} x * (\mu - \widetilde{\mu})_t\,. \tag{2.2}$$

Here $(w_t)_{t\geq 0}$ is a standard Brownian motion, $\mu(\mathrm{d}s\,\mathrm{d}x)$ is the jump measure with the deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is the Lévy measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$, (see, for example in [3]), such that

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

We use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_*} |z|^m\,\Pi(\mathrm{d}z)$. Moreover, we assume that the nuisance parameters $a \leq 0$, ϱ_1 and ϱ_2 satisfy the conditions

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

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 $\check{\delta} > 0$

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

We denote by $\mathcal{Q}_n$ the family of all distributions of process (1.1) – (2.1) on the Skorokhod space $\mathbf{D}[0,n]$ satisfying the conditions (2.4) – (2.5).

It should be noted that in view of Corollary 7.2 in [16] the condition (1.3) for the process (2.1) holds with

$$\varkappa_Q = 2\underline{\varrho}_*\,. \tag{2.6}$$

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

3 Improved estimation

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

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

$$\sup_{0\leq j\leq n}\ \sup_{0\leq t\leq 1} |\phi_j(t)| \,\leq\, \phi_* < \infty\,. \tag{3.1}$$

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

$$\sup_{d\geq d_0} \frac{1}{d} \int_0^1 \Phi^*_d(v)\,\mathrm{d}v \,\leq\, \check{a}\,, \tag{3.2}$$

where $\Phi_d^*(v) = \max_{t \geq v} \left| \sum_{j=1}^d \phi_j(t)\, \phi_j(t-v) \right|$.

For example, we can take the trigonometric basis defined as $\mathrm{Tr}_1 \equiv 1$ and $\mathrm{Tr}_j(x) = \sqrt{2}\cos(\varpi_j x)$ for even j and $\mathrm{Tr}_j(x) = \sqrt{2}\cos(\varpi_j x)$ for odd $j \geq 2$, where the frequency $\varpi_j = 2\pi[j/2]$ and $[x]$ denotes integer part of x. In Lemma A1 in [16] it is shown that these functions satisfy the condition $\mathbf{B}_2$) with $d_0 = \inf\{d \geq 7 : 5 + \ln d \leq \check{a} d\}$ and $\check{a} = (1 - e^{-a_{max}})/(4a_{max})$.

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

$$\theta_j = (S, \phi_j) = \int_0^1 S(t)\, \phi_j(t)\, \mathrm{d}t \tag{3.3}$$

can be estimated as

$$\widehat{\theta}_{j,n} = \frac{1}{n} \int_0^n \phi_j(t)\, \mathrm{d}y_t\,. \tag{3.4}$$

We replace the differential $S(t)\mathrm{d}t$ by the stochastic observed differential $\mathrm{d}y_t$. In view of (1.1), one obtains

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

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

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

where the weights $\gamma = (\gamma(j))_{1 \leq j \leq n} \in \mathbb{R}^n$ belong to some finite set Γ from $[0,\, 1]^n$ for which we set

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

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.

$\mathbf{D}_1$) *Assume that for for any vector* $\gamma \in \Gamma$ *there exists some fixed integer* $d = d(\gamma)$ *such that their first* d *components equal to one, i.e.* $\gamma(j) = 1$ *for* $1 \leq j \leq d$ *for any* $\gamma \in \Gamma$.

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

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

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

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

where $\lambda_{max}(A)$ *is the maximal eigenvalue of the matrix* A.
As it is shown in Proposition 7.11 in [16] the condition $\mathbf{D}_2$) holds for the non-Gaussian Ornstein–Uhlenbeck-based model (1.1) – (2.1) with $l_n^* = \underline{\varrho}_n(d-6)/2$ and $d \geq d_0$.

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

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

where $g(j) = (\mathbf{c}_n/|\widetilde{\theta}_n|_d)\mathbf{1}_{\{1\leq j\leq d\}}$,

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

The positive parameter r_n^* is such that

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

for any $\check{\delta} > 0$.

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

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

We denote the difference of quadratic risks of the estimates (3.6) and (3.12) as

$$\Delta_Q(S) := \mathcal{R}_Q(S_\gamma^*, S) - \mathcal{R}_Q(\widehat{S}_\gamma, S)\,.$$

For this difference we obtain the following result.

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

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

Remark 3.1. *The inequality* (3.13) *means that non asymptotically, i.e. for any* $n \geq n_0$ *the estimate* (3.12) *outperforms in mean square accuracy the estimate* (3.6).

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 $(S_\gamma^*)_{\gamma\in\Gamma}$.

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

$$\mathrm{Err}_n(\gamma) = \|S_\gamma^* - S\|^2.$$

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

$$\mathrm{Err}_n(\gamma) \;=\; \sum_{j=1}^{n} \gamma^2(j)(\theta_{j,n}^*)^2 \;-\, 2\sum_{j=1}^{n} \gamma(j)\theta_{j,n}^*\,\theta_j \;+\; \sum_{j=1}^{n} \theta_j^2 \,. \tag{4.1}$$

Since the Fourier coefficients $(\theta_j)_{j\geq 1}$ are unknown, the weight coefficients $(\gamma_j)_{j\geq 1}$ can not be found by minimizing this quantity. To circumvent this difficulty one needs to replace the terms $\theta_{j,n}^*\,\theta_j$ by their estimators $\widetilde{\theta}_{j,n}$. We set

$$\widetilde{\theta}_{j,n} = \theta_{j,n}^*\,\widehat{\theta}_{j,n} - \frac{\widehat{\sigma}_n}{n} \tag{4.2}$$

where $\widehat{\sigma}_n$ is the estimate for the limiting variance of $\mathbf{E}_Q\,\xi_{j,n}^2$ which we choose in the following form

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \widehat{t}_{j,n}^2\,, \quad \widehat{t}_{j,n} = \int_0^1 \mathrm{Tr}_j(t)\mathrm{d}y_t. \tag{4.3}$$

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

$$J_n(\gamma) \;=\; \sum_{j=1}^{n} \gamma^2(j)(\theta_{j,n}^*)^2 \;-\, 2\sum_{j=1}^{n} \gamma(j)\,\widetilde{\theta}_{j,n} \;+\; \rho\,\widehat{P}_n(\gamma) \tag{4.4}$$

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

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

Substituting the weight coefficients, minimizing the cost function

$$\gamma^* = \operatorname{argmin}_{\gamma \in \Gamma} J_n(\gamma) \, , \tag{4.6}$$

in (3.6) leads to the improved model selection procedure

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

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

To prove the sharp oracle inequality, the following conditions will be needed for the family $\mathcal{Q}_n$ of distributions of the noise $(\xi_t)_{t \geq 0}$ in (1.1).

We need to impose some stability conditions for the noise Fourier transform sequence $(\xi_{j,n})_{1 \leq j \leq n}$ introduced in [15]. To this end for some parameter $\sigma_Q > 0$ we set the following function

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

In [11] the parameter σ_Q is called proxy variance.

$\mathbf{C}_1$) *There exists a proxy variance $\sigma_Q > 0$ such that for any $\epsilon > 0$*

$$\lim_{n \to \infty} \frac{\mathbf{L}_{1,n}(Q)}{n^\epsilon} = 0 \, .$$

Moreover, we define

$$\mathbf{L}_{2,n}(Q) = \sup_{|x| \leq 1} \mathbf{E}_Q \left(\sum_{j=1}^{n} x_j \, \widetilde{\xi}_{j,n} \right)^2 \quad \text{and} \quad \widetilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2 \, .$$

$\mathbf{C}_2$) *Assume that for any $\epsilon > 0$*

$$\lim_{n \to \infty} \frac{\mathbf{L}_{2,n}(Q)}{n^\epsilon} = 0 \, .$$

Theorem 4.1. *If the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) hold for the distribution Q of the process ξ in (1.1), then, for any $n \geq 1$ and $0 < \rho < 1/2$, the risk (1.4) of estimate (4.7) for S satisfies the oracle inequality*

$$\mathcal{R}_Q(S^*, S) \, \leq \, \frac{1 + 5\rho}{1 - \rho} \min_{\gamma \in \Gamma} \mathcal{R}_Q(S^*_\gamma, S) + \frac{\mathbf{B}_n(Q)}{\rho n} \, , \tag{4.9}$$

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

$$\lim_{n\to\infty} \frac{\mathbf{U}_n(Q)}{n^\epsilon} = 0\,. \tag{4.10}$$

In the case, when the value of σ_Q in $\mathbf{C}_1$) is known, one can take $\widehat{\sigma}_n = \sigma_Q$ and

$$P_n(\gamma) = \frac{\sigma_Q\,|\gamma|_n^2}{n} \tag{4.11}$$

and then we can rewrite the oracle inequality (4.9) with $\mathbf{B}_n(Q) = \mathbf{U}_n(Q)$. Now we study the estimate (4.3).

Proposition 4.2. *Let in the model* (1.1) *the function* $S(\cdot)$ *is continuously differentiable. Then, for any* $n \geq 2$,

$$\mathbf{E}_Q|\widehat{\sigma}_n - \sigma_Q| \leq \frac{\varkappa_n(Q)(1 + \|\dot{S}\|^2)}{\sqrt{n}}\,,$$

where the term $\varkappa_n(Q)$ *possesses the property* (4.10) *and* $\dot{S}$ *is the derivative of the function* S.

To obtain the oracle inequality for the robust risk (1.5) we need some additional condition on the distribution family $\mathcal{Q}_n$. We set

$$\varsigma^* = \varsigma_n^* = \sup_{Q\in\mathcal{Q}_n} \sigma_Q \quad \text{and} \quad \mathbf{L}_n^* = \sup_{Q\in\mathcal{Q}_n} \left(\mathbf{L}_{1,n}(Q) + \mathbf{L}_{2,n}(Q)\right). \tag{4.12}$$

$\mathbf{C}_1^*$) *Assume that the conditions* $\mathbf{C}_1$)–$\mathbf{C}_2$) *hold and for any* $\epsilon > 0$

$$\lim_{n\to\infty} \frac{\mathbf{L}_n^* + \varsigma_n^*}{n^\epsilon} = 0\,.$$

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

$\mathbf{C}_2^*$) *Assume that the set* Γ *is such that for any* $\epsilon > 0$

$$\lim_{n\to\infty} \frac{\nu}{n^\epsilon} = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{|\Gamma|_*}{n^{1/2+\epsilon}} = 0\,.$$

As is shown in [16], both the 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 [16] we can obtain the following result

Theorem 4.3. *Assume that the conditions* $\mathbf{C}_1^*$)–$\mathbf{C}_2^*$) *hold. Then the robust risk* (1.5) *of the estimate* (4.7) *for continuously differentiable function* $S(t)$ *satisfies for any* $n \geq 2$ *and* $0 < \rho < 1/2$ *the oracle inequality*

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

where the term $\mathbf{B}_n^*$ *satisfies the property* (4.10).

Now we specify the weight coefficients $(\gamma(j))_{j\geq 1}$ as it is 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

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

where $r_i = i\varepsilon$ and $m = [1/\varepsilon^2]$. Both parameters $k^* \geq 1$ and $0 < \varepsilon \leq 1$ are assumed to be functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that for any $\delta > 0$

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0\,, \\ \lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\delta}\varepsilon(n) = +\infty\,. \end{cases}$$

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 $\gamma_\alpha = (\gamma_\alpha(j))_{j\geq 1}$ as

$$\gamma_\alpha(j) = \mathbf{1}_{\{1\leq j\leq d\}} + \left(1 - (j/\omega_\alpha)^{\beta}\right) \mathbf{1}_{\{d<j\leq\omega_\alpha\}} \tag{4.13}$$

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

$$\tau_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\,.$$

We set

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

It will be noted that such weight coefficients satisfy the condition $\mathbf{D}_1$) and in this case the cardinal of the set Γ is $\nu = k^*m$. Moreover, taking into account that $\tau_\beta < 1$ for $\beta \geq 1$ we obtain for the set (4.14)

$$|\Gamma|_* \;\leq\; 1 + \sup_{\alpha\in\mathcal{A}} \omega_\alpha \leq 1 + (v_n/\varepsilon)^{1/3}\,.$$

Remark 4.1. *These weight coefficients are used in [11, 12] to show the asymptotic efficiency for model selection procedures.*

5 Asymptotic efficiency

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

$$W_{k,\mathbf{r}} = \{f \in \mathbf{C}^k_p[0,1] \,:\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq \mathbf{r}\}\,,$$

where $\mathbf{r} > 0$ and $k \geq 1$ are some unknown parameters, $\mathbf{C}_p^k[0,1]$ is the space of k times differentiable 1 - periodic $\mathbb{R} \to \mathbb{R}$ functions such that $f^{(i)}(0) = f^{(i)}(1)$ for any $0 \leq i \leq k-1$. To study the asymptotic efficiency we denote by Σ_n all estimators $\widehat{S}_n$ i.e. any $\sigma\{y_t\,,\, 0 \leq t \leq n\}$ mesurable functions. In the sequel we denote by Q^* the distribution of the process $(y_t)_{0\leq t\leq n}$ 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) admits the following lower bound*

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

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

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

Theorem 5.2. *The quadratic risk (1.2) for the estimating procedure (4.7) has the following asymptotic upper bound*

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

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

Corollary 5.3. *The model selection procedure (4.7) is efficient, i.e.*

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

Theorem 5.2 and Theorem 5.1 are shown by 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 (4.6). We simulate the model (1.1) with 1-periodic function S of the form

$$S(t) = t\,\sin(2\pi t) + t^2(1-t)\cos(4\pi t) \tag{6.1}$$

on [0, 1] and the 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 with intensity $\lambda = 1$ and $(Y_j)_{j\geq 1}$ is i.i.d. $\mathcal{N}(0,\,1)$ sequence (see, for example, [12]).

We use the model selection procedure (4.6) with the weights (4.13) 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

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

where $\Delta_n(t) = S_n^*(t) - S(t)$ and $\Delta_{n,l}(t) = S_{n,l}^*(t) - S(t)$ is the deviation for the lth replication. In this example we take $p = 100001$ and $N = 1000$.

Table 1: The sample quadratic risks for different optimal γ

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

Table 2: The sample quadratic risks for the same optimal $\widehat{\gamma}$

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

Table 1 gives the values for the sample risks of the improved estimate (4.6) 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 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.

Remark 6.1. *Figures 1–2 show the behavior of the procedures (3.6) and (4.6) 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 $\widehat{S}$ and the dashed line is the improved model selection procedure S^*. From the Table 2 for the same γ with various observations numbers n we 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 benefit is considerable for non large n.*

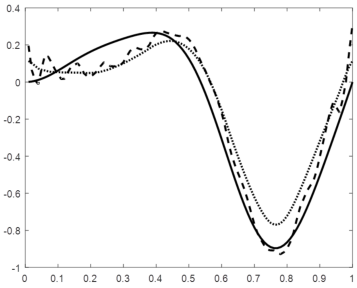

Figure 1: Behavior of the regression function and its estimates for $n = 500$.

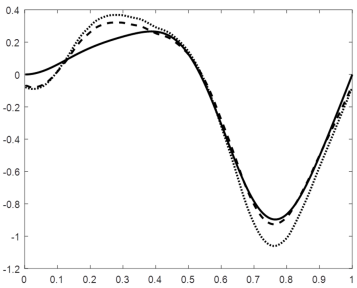

Figure 2: Behavior of the regression function and its estimates for $n = 1000$.

7 Proofs

7.1 Proof of Theorem 3.1

Consider the quadratic error of the estimate (3.12)

$$\begin{aligned}\|S^*_\gamma - S\|^2 &= \sum_{j=1}^{n}(\gamma(j)\theta^*_{j,n} - \theta_j)^2 = \sum_{j=1}^{d}(\gamma(j)\theta^*_{j,n} - \theta_j)^2 + \sum_{j=d+1}^{n}(\gamma(j)\widehat{\theta}_{j,n} - \theta_j)^2 \\ &= \sum_{j=1}^{n}(\gamma(j)\widehat{\theta}_{j,n} - \theta_j)^2 + \mathbf{c}_n^2 - 2\mathbf{c}_n\sum_{j=1}^{d}(\widehat{\theta}_{j,n} - \theta_j)\frac{\widehat{\theta}_{j,n}}{\|\widetilde{\theta}_n\|_d} \\ &= \|\widehat{S}_\gamma - S\|^2 + \mathbf{c}_n^2 - 2\mathbf{c}_n\sum_{j=1}^{d}(\widehat{\theta}_{j,n} - \theta_j)\iota_j(\widetilde{\theta}_n)\,,\end{aligned}$$

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

$$\mathcal{R}_Q(S^*_\gamma, S) = \mathcal{R}_Q(\widehat{S}_\gamma, S) + \mathbf{c}_n^2 - 2\mathbf{c}_n\,\mathbf{E}_{Q,S}\sum_{j=1}^{d}(\widehat{\theta}_{j,n} - \theta_j)\,I_{j,n}\,,$$

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

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

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

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

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

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

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

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

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

$$\Delta_Q(S) = \mathbf{c}_n^2 - 2\mathbf{c}_n n^{-1}\mathbf{E}_{Q,S}\left(\frac{\mathrm{tr}\mathbf{G}_n}{\|\widetilde{\theta}_n\|_d} - \frac{\widetilde{\theta}_n'\mathbf{G}_n\widetilde{\theta}_n}{\|\widetilde{\theta}_n\|^3}\right)$$

$$\le \mathbf{c}_n^2 - 2\mathbf{c}_n\, l_n^* n^{-1}\mathbf{E}_{Q,S}\,\frac{1}{\|\widetilde{\theta}_n\|_d}\,.$$

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

$$\mathbf{E}_{Q,S}\,(\|\widetilde{\theta}_n\|_d)^{-1} = \mathbf{E}_{Q,S}\,(\|\widetilde{\theta}+n^{-1/2}\widetilde{\xi}_n\|_d)^{-1} \ge\ (\|\theta\|_d + n^{-1/2}\mathbf{E}_{Q,S}\|\widetilde{\xi}_n\|_d)^{-1}\,.$$

Note now that the condition through the inequality (1.3) we obtain

$$\mathbf{E}_{Q,S}\|\widetilde{\xi}_n\|_d^2 \le \varkappa_Q\, d\,.$$

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

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

and, therefore,

$$\Delta_Q(S) \le \mathbf{c}_n^2 - 2\mathbf{c}_n\frac{l_n^*}{\left(r_n^* + \sqrt{d\varkappa_*/n}\right)n} = -\mathbf{c}_n^2\,.$$

Hence Theorem 3.1. $\square$

7.2 Proof of Theorem 4.1

Substituting (4.4) in (4.1) yields for any $\gamma \in \Gamma$

$$\mathrm{Err}_n(\gamma) \;=\; J_n(\gamma) + 2\sum_{j=1}^{n} \gamma(j)\left(\theta^*_{j,n}\,\widehat{\theta}_{j,n} - \frac{\widehat{\sigma}_n}{n} - \theta^*_{j,n}\,\theta_j\right)$$

$$+\; \|S\|^2 - \rho\widehat{P}_n(\gamma)\,. \tag{7.2}$$

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

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

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

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

$$\mathrm{Err}_n(\gamma) \;=\; J_n(\gamma) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n}L(\gamma) + 2\,M(\gamma) + \frac{2}{n}B_{1,n}(\gamma)$$

$$+ 2\sqrt{P_n(\gamma)}\frac{B_{2,n}(\overline{\gamma})}{\sqrt{\sigma_Q n}} - 2B_{3,n}(\gamma) \;+\; \|S\|^2 - \rho\widehat{P}_n(\gamma) \tag{7.3}$$

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

$$\mathrm{Err}_n(\gamma^*) - \mathrm{Err}_n(\gamma_0) \le 2\frac{\sigma_Q - \widehat{\sigma}_n}{n}L(x) + 2M(x) + \frac{2}{n}B_{1,n}(x)$$

$$+ 2\sqrt{P_n(\gamma^*)}\frac{B_{2,n}(\overline{\gamma^*})}{\sqrt{\sigma_Q n}} - 2\sqrt{P_n(\gamma_0)}\frac{B_{2,n}(\overline{\gamma_0})}{\sqrt{\sigma_Q n}}$$

$$- 2B_{3,n}(\gamma^*) + 2B_{3,n}(\gamma_0) - \rho\widehat{P}_n(\gamma^*) + \rho\widehat{P}_n(\gamma_0)\,,$$

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

$$2\sqrt{P_n(\gamma)}\frac{B_{2,n}(\overline{\gamma})}{\sqrt{\sigma_Q n}} \le \varepsilon P_n(\gamma) + \frac{B^2_{2,n}(\overline{\gamma})}{\varepsilon\sigma_Q n} \le \varepsilon P_n(\gamma) + \frac{B^*_2}{\varepsilon\sigma n}\,,$$

where

$$B^*_2 = \max_{\gamma\in\Gamma}\left(B^2_{2,n}(\overline{\gamma}) + B^2_{2,n}(\overline{\gamma}^2)\right)$$

with $\gamma^2 = (\gamma_j^2)_{1\le j\le n}$. Note that from definition the function $\mathbf{L}_{2,n}(Q)$ in the condition $\mathbf{C}_2$) we obtain 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 method we estimate the term $B_{3,n}$. Note that

$$\sum_{j=1}^{n} g_\gamma^2(j)\,\widehat{\theta}_j^2 = \mathbf{c}_n^2 \le \frac{\mathbf{c}_n^*}{n}\,, \tag{7.6}$$

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

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

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

$$2|B_{3,n}(\gamma)| \le \varepsilon P_n(\gamma) + \frac{\mathbf{c}_n^*}{\varepsilon\sigma_Q n}(\sigma_Q + B_2^*)\,.$$

Using the bounds above, one has

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

The setting $\varepsilon = \rho/4$ and the estimating where this is possible ρ by 1 in this inequality imply

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

Moreover, taking into account here that

$$|\widehat{P}_n(\gamma_0) - P_n(\gamma_0)| \le \frac{|\Gamma|_n|\widehat{\sigma}_n - \sigma_Q|}{n}$$

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

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

Now we examine the third term in the right-hand side of this inequality. Firstly we note that

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

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

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

We remind that the set $\Gamma_1 = \Gamma - \gamma_0$. Using Proposition **??** we can obtain that for any fixed $x = (x_j)_{1\le j\le n} \in \mathbb{R}^n$

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

and, therefore,

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

Moreover, the norm $\|S^*_{\gamma^*} - S^*_{\gamma_0}\|$ can be estimated from below as

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

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

$$\|S_x\|^2 - \|S^*_\gamma - S^*_{\gamma_0}\|^2 \le \|S_x\|^2 - \|\widehat{S}_x\|^2 - 2\sum_{j=1}^n x(j)\beta(j)\widehat{\theta}_j^2$$

$$\le -2M(x^2) - 2\sum_{j=1}^n x(j)\beta(j)\widehat{\theta}_j\theta_j - \frac{2}{\sqrt{n}}\Upsilon(x)\,,$$

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

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

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

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

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

$$\|S_x\|^2 \le \frac{1}{1-\varepsilon}\left(\|S_\gamma^* - S_{\gamma_0}^*\|^2 + \frac{Z_1^*}{n\varepsilon}\right.$$
$$\left. -2\sum_{j=1}^{n} x(j)\beta(j)\widehat{\theta}_j\theta_j - \frac{2\Upsilon(x)}{\sqrt{n}}\right)\,. \tag{7.11}$$

Moreover, note now that the property (7.6) yields

$$\sum_{j=1}^{n} \beta^2(j)\widehat{\theta}_j^2 \le 2\sum_{j=1}^{n} g_\gamma^2(j)\,\widehat{\theta}_j^2 + 2\sum_{j=1}^{n} g_{\gamma_0}^2(j)\,\widehat{\theta}_j^2 \le \frac{4\mathbf{c}^*}{\varepsilon n}\,. \tag{7.12}$$

Taking into account that $|x(j)| \le 1$ and using the inequality (7.4), we get that for any $\varepsilon > 0$

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

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

$$\frac{2}{\sqrt{n}}|\Upsilon(\gamma)| \le \frac{2|\gamma|_n}{\sqrt{n}}\left(\sum_{j=1}^{n} \beta^2(j)\widehat{\theta}_j^2\right)^{1/2}\left(\sum_{j=1}^{n} \bar{\gamma}^2(j)\,\xi_j^2\right)^{1/2}$$
$$\le \varepsilon P_n(\gamma) + \frac{\mathbf{c}^*}{n\varepsilon\sigma_Q}\sum_{j=1}^{n} \bar{\gamma}^2(j)\,\xi_j^2 \le \varepsilon P_n(\gamma) + \frac{\mathbf{c}^*(\sigma_Q + B_2^*)}{n\varepsilon\sigma_Q}\,.$$

Therefore,

$$\frac{2}{\sqrt{n}}|\Upsilon(x)| \le \frac{2}{\sqrt{n}}|\Upsilon(\gamma^*)| + \frac{2}{\sqrt{n}}|\Upsilon(\gamma_0)|$$
$$\le \varepsilon P_n(\gamma^*) + \varepsilon P_n(\gamma_0) + \frac{2\mathbf{c}^*(\sigma_Q + B_2^*)}{n\varepsilon\sigma_Q}\,.$$

So, using all these bounds in (7.11), we obtain that

$$\|S_x\|^2 \le \frac{1}{(1-\varepsilon)}\Bigg(\frac{Z_1^*}{n\varepsilon} + \|S^*_{\gamma^*} - S^*_{\gamma_0}\|^2 + \frac{6\mathbf{c}_n^*(\sigma + B_2^*)}{n\sigma\varepsilon} + \varepsilon P_n(\gamma^*) + \varepsilon P_n(\gamma_0)\Bigg).$$

Using in the inequality (7.8) this bound and the estimate

$$\|S^*_{\gamma^*} - S^*_{\gamma_0}\|^2 \le 2(\mathrm{Err}_n(\gamma^*) + \mathrm{Err}_n(\gamma_0)),$$

we obtain

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

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

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

From here and (7.7), it follows that

$$\mathrm{Err}_n(\gamma^*) \le \frac{1+\varepsilon}{1-3\varepsilon}\mathrm{Err}_n(\gamma_0) + \frac{6|\Gamma|_n|\widehat{\sigma}_n - \sigma_Q|}{n(1-3\varepsilon)} + \frac{2}{n(1-3\varepsilon)}\mathbf{L}_{1,n}(Q) + \frac{28(1+\mathbf{c}_n^*)(B_2^* + \sigma_Q)}{\rho(1-3\varepsilon)n\sigma_Q} + \frac{2(Z^* + Z_1^*)}{n(1-3\varepsilon)} + \frac{2\rho P_n(\gamma_0)}{1-3\varepsilon}.$$

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

$$\mathrm{Err}_n(\gamma^*) \le \frac{1+\rho/3}{1-\rho}\mathrm{Err}_n(\gamma_0) + \frac{12|\Gamma|_n|\widehat{\sigma}_n - \sigma_Q|}{n} + \frac{4}{n}\mathbf{L}_{1,n}(Q) + \frac{56(1+\mathbf{c}_n^*)(B_2^* + \sigma_Q)}{\rho n\sigma_Q} + \frac{4(Z^* + Z_1^*)}{n} + \frac{2\rho P_n(\gamma_0)}{1-\rho}.$$

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

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

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

7.3 Property of Penalty term

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

$$P_n(\gamma) \leq \frac{\mathbf{E}\, Err_n(\gamma)}{1-\varepsilon} + \frac{\mathbf{c}^*_n}{n\varepsilon(1-\varepsilon)}\,. \tag{7.13}$$

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

$$\mathrm{Err}_n(\gamma) = \sum_{j=1}^{n} (\gamma(j)\theta^*_{j,n} - \theta_j)^2 = \sum_{j=1}^{n} \Big(\gamma(j)(\theta^*_{j,n} - \theta_j) + (\gamma(j)-1)\theta_j\Big)^2$$

$$\geq \sum_{j=1}^{n} \gamma(j)^2(\theta^*_{j,n} - \theta_j)^2 + 2\sum_{j=1}^{n} \gamma(j)(\gamma(j)-1)\theta_j(\theta^*_{j,n} - \theta_j).$$

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

$$\sum_{j=1}^{n} \gamma(j)(\gamma(j)-1)\theta_j(\theta^*_{j,n} - \theta_j) = \sum_{j=1}^{n} \gamma(j)(\gamma(j)-1)\theta_j(\widehat{\theta}_{j,n} - \theta_j)\,,$$

i.e. $\mathbf{E}\sum_{j=1}^{n} \gamma(j)(\gamma(j)-1)\theta_j(\theta^*_{j,n} - \theta_j) = 0$ and, therefore, taking into account the definition (4.11) we obtain that

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

$$\geq P_n(\gamma) - \frac{2}{\sqrt{n}}\mathbf{E}\sum_{j=1}^{n} \gamma(j)^2 g_\gamma(j)\widehat{\theta}_{j,n}\xi_j$$

$$\geq (1-\varepsilon)\,P_n(\gamma) - \frac{1}{\varepsilon}\mathbf{E}\sum_{j=1}^{n} g^2_\gamma(j)\widehat{\theta}^2_j\,.$$

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

Acknownledgements. This work was partially supported by the research project no. 2.3208.2017/4.6 (the Ministry of Education and Science of the Russian Federation) and RFBR Grant 16-01-00121. The work of the second author was partially supported by the Russian Federal Professor program (project no. 1.472.2016/1.4, the Ministry of Education and Science of the Russian Federation).

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. **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. **21**(1). P. 1–16.

[6] Galtchouk L., Pergamenshchikov S. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression // *Journal of Korean Statistical Society*. 2009. **38**(4). P. 305–322.

[7] James W., Stein C. Estimation with quadratic loss // *In Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability*. 1961. **1**. Berkeley: University of California Press. P. 361–380.

[8] Konev V.V., Pergamenshchikov S.M. Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities // *Vestnik Tomskogo Universiteta. Mathematics and Mechanics*. 2009. **3**(7). P. 23–41.

[9] Konev V.V., Pergamenshchikov S.M. Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency // *Vestnik Tomskogo Universiteta. Mathematics and Mechanics*. 2009. **4**(8). P. 31–45 (2009).

[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. **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. **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. **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. **58**(3). P. 442–457.

[14] Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // *Stat. Inference Stoch. Process.* 2013. **16**(1). P. 15–28.

[15] Pchelintsev E., Pergamenshchikov S. Oracle inequalities for the stochastic differential equations. *Stat. Inference Stoch. Process.* 2018. **21** (2). P. 469–483.

[16] Pchelintsev E., Pergamenshchikov S. Adaptive model selection method for a conditionally Gaussian semimartingale regression in continuous time. - Preprint, 2018