

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Международная лаборатория статистики случайных
процессов и количественного финансового анализа

**Международная научная
конференция
«Робастная статистика и
финансовая математика – 2017»**

(03–05 июля 2017 г.)

Сборник статей

Под редакцией
д-ра физ.-мат. наук, профессора С.М. Пергаменщикова,
канд. физ.-мат. наук, доцента Е.А. Пчелинцева

Томск
Издательский Дом Томского государственного университета
2017

Improved estimation of a function in continuous regression with semimartingale noise^{*}

Pchelintsev E. A., Pchelintsev V. A., Pergamenshchikov S. M.

Tomsk State University, Tomsk

e-mail: evgen-pch@yandex.ru

Abstract

In this paper, we consider the robust adaptive non parametric estimation problem for the periodic function observed with the semimartingale noises in continuous time. An adaptive model selection procedure, based on the improved weighted least square estimates, is proposed. Sharp oracle inequalities for the robust risks have been obtained.

Key words: Improved non-asymptotic estimation, Weighted least squares estimates, Robust quadratic risk, Non-parametric regression, Model selection procedure, Oracle inequality.

Introduction. Consider a regression model in continuous time

$$dy_t = S(t)dt + d\xi_t, \quad 0 \leq t \leq n, \quad (1)$$

where S is an unknown 1-periodic $\mathbb{R} \rightarrow \mathbb{R}$ function, $S \in \mathbf{L}_2[0, 1]$; $(\xi_t)_{t \geq 0}$ is an unobservable conditionally gaussian semimartingale with the values in the Skorokhod space $\mathbf{D}[0, n]$ such that, for any cadlag $[0, n] \rightarrow \mathbb{R}$ function f from $\mathbf{L}_2[0, n]$, the stochastic integral

$$I_n(f) = \int_0^n f(s)d\xi_s \quad (2)$$

is well defined and has the following properties

$$\mathbf{E}_Q I_n(f) = 0 \quad \text{and} \quad \mathbf{E}_Q I_n^2(f) \leq \varkappa_Q \int_0^n f^2(s)ds. \quad (3)$$

Here \mathbf{E}_Q denotes the expectation with respect to the distribution Q in $\mathcal{D}[0, n]$ of the process $(\xi_t)_{0 \leq t \leq n}$, which is assumed to belong to some probability family \mathcal{Q}_n^* specified below; $\varkappa_Q > 0$ is some positive constant depending on the distribution Q .

The problem is to estimate the unknown function S in the model (1) on the basis of observations $(y_t)_{0 \leq t \leq n}$.

The class of the disturbances ξ satisfying conditions (3) is rather wide and comprises, in particular, the Lévy processes which are used in different applied problems. The models (1) with the Lévy's type

^{*}This study was supported by RSF grant, project no 17-11-01049

noise naturally arise in the nonparametric functional statistics problems. Moreover, non-Gaussian Ornstein–Uhlenbeck-based models, enter this class.

We consider the estimation problem in the adaptive setting, i.e. when the regularity of S is unknown. Since the distribution Q of the noise process $(\xi_t)_{0 \leq t \leq n}$ is unknown we use the robust estimation approach developed for nonparametric problems in [5]. To this end we define the robust risk as

$$\mathcal{R}^*(\widehat{S}_n, S) = \sup_{Q \in \mathcal{Q}_n^*} \mathcal{R}_Q(\widehat{S}_n, S) \quad (4)$$

where \widehat{S}_n is an estimation, i.e. any function of $(y_t)_{0 \leq t \leq n}$, $\mathcal{R}_Q(\cdot, \cdot)$ is the usual quadratic risk defined as

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

In this paper, we consider a minimax optimisation criteria which aims to minimize the robust risk (4). To do this we use the model selection methods. The interest to such statistical procedures is explained by the fact that they provide adaptive solutions for a nonparametric estimation through oracle inequalities which give a non-asymptotic upper bound for a quadratic risk including a minimal risk over chosen family of estimators. It should be noted that the model selection methods for parametric models were proposed, for the first time, by Akaike [1]. Then, these methods had been developed by Barron, Birgé and Massart [2] and Fourdrinier and Pergamenschikov [3] for the nonparametric estimation and oracle inequalities for the quadratic risks. Unfortunately, the oracle inequalities obtained in these papers can not provide the efficient estimation in the adaptive setting, since the upper bounds in these inequalities have some fixed coefficients in the main terms which are more than one. In order to obtain the efficiency property for estimation procedures, one has to obtain the sharp oracle inequalities, i.e. in which the factor at the principal term on the right-hand side of the inequality is close to unity. For this reason, one needs to use the general semi - martingale approach for the robust adaptive efficient estimation of the nonparametric signals in continuous time proposed by Konev and Pergamenschikov in [5]. The goal of this paper is to develop a new sharp model selection method for estimating the unknown signal S using the improved estimation approach. Usually, the model selection procedures are based on the least square estimators. However, in this paper, we propose to use the improved least square estimators which enable us to considerably improve the

non asymptotic estimation accuracy. Such idea was proposed, for the first time, in [3]. Our goal is to develop these methods for non gaussian regression models in continuous time and to obtain the sharp oracle inequalities. It should be noted that to apply the improved estimation methods to the non gaussian regression models in continuous time one needs to modify the well known James - Stein procedure introduced in [4] in the way proposed in [6, 7]. So, by using these estimators we construct the improved model selection procedure and we show that the constructed estimation procedure is optimal in the sense of the sharp non asymptotic oracle inequalities for the robust risks (4).

Example of noise processes. Let the useful signal S is distorted by the impulse flow described by the Lévy process, i.e. we assume that the noise process $(\xi_t)_{0 \leq t \leq n}$ is defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \tilde{\mu})_t, \quad (5)$$

where, ϱ_1 and ϱ_2 are some unknown constants, $(w_t)_{t \geq 0}$ 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 $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$, such that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^6) < \infty.$$

Here we use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_*} |z|^m \Pi(dz)$. Note that the Lévy measure $\Pi(\mathbb{R}_*)$ could be equal to $+\infty$. Then the class \mathcal{Q}_n^* of distributions of the process $(\xi_t)_{0 \leq t \leq n}$ includes all distributions for which the parameters $\varkappa_Q = \varrho_1 \geq \varsigma_*$ and $\varrho_1^2 + \varrho_2^2 \leq \varsigma^*$, where ς_* and ς^* are some fixed positive bounds.

Improved estimation. 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 $t \in \mathbb{R}$. For estimating the unknown function S in (1) we consider it's Fourier expansion

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j) = \int_0^1 S(t) \phi_j(t) dt.$$

The corresponding Fourier coefficients can be estimated as

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

In view of (1), one obtains

$$\hat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}} \xi_{j,n}, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j), \quad (6)$$

where $I_n(\phi_j)$ is given in (2).

We define a class of weighted least squares estimates for $S(t)$ as

$$\widehat{S}_\lambda = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \phi_j, \quad (7)$$

where the weights $\lambda \in \mathbb{R}^n$ belong to some finite set Λ from $[0, 1]^n$.

Now, for the first d Fourier coefficients in (7) we use the improved estimation method proposed for parametric models in [7]. To this end we set $\widehat{\theta}_n = (\widehat{\theta}_{j,n})_{1 \leq j \leq d}$. 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

$$\theta_{j,n}^* = (1 - g(j)) \widehat{\theta}_{j,n}, \quad (8)$$

where $g(j) = (\mathbf{c}_n / |\widehat{\theta}_n|_d) \mathbf{1}_{\{1 \leq j \leq d\}}$, and \mathbf{c}_n is some known parameter such that $\mathbf{c}_n \approx d/n$ as $n \rightarrow \infty$. Now we introduce a class of shrinkage weighted least squares estimates for S as

$$S_\lambda^* = \sum_{j=1}^n \lambda(j) \theta_{j,n}^* \phi_j. \quad (9)$$

We denote the difference of quadratic risks of the estimates (7) and (9) as $\Delta_Q(S) := \mathcal{R}_Q(S_\lambda^*, S) - \mathcal{R}_Q(\widehat{S}_\lambda, S)$. Now for this deviation we obtain the following result.

Theorem 1. *Assume that for any vector $\lambda \in \Lambda$ there exists some fixed integer $d = d(\lambda)$ such that their first d components equal to one, i.e. $\lambda(j) = 1$ for $1 \leq j \leq d$ for any $\lambda \in \Lambda$. Then for any $n \geq 1$*

$$\sup_{Q \in \mathcal{Q}_n^*} \sup_{\|S\| \leq \mathbf{r}} \Delta_Q(S) < -\mathbf{c}_n^2. \quad (10)$$

Remark. The inequality (10) means that non asymptotically, i.e. for any $n \geq 1$ the estimate (9) outperforms in mean square accuracy the estimate (7). Moreover, as we will see below, $n\mathbf{c}_n \rightarrow \infty$ as $d \rightarrow \infty$. This means that improvement is considerable may better than for the parametric regression (see, [7]).

Model selection procedure. This Section gives the construction of a model selection procedure for estimating a function S in (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) will be constructed on the basis of a family of estimates $(S_\lambda^*)_{\lambda \in \Lambda}$.

The performance of any estimate S_λ^* will be measured by the em-

empirical squared error

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

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (9). It is obvious, that the best way is to minimise the empirical squared error with respect to λ . Making use the estimate definition (9) and the Fourier transformation of S implies

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j)(\theta_{j,n}^*)^2 - 2 \sum_{j=1}^n \lambda(j)\theta_{j,n}^* \theta_j + \sum_{j=1}^n \theta_j^2.$$

Since the Fourier coefficients $(\theta_j)_{j \geq 1}$ are unknown, the weight coefficients $(\lambda_j)_{j \geq 1}$ 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 $\tilde{\theta}_{j,n}$. We set

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

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

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

Here we denoted by $(\text{Tr}_j)_{j \geq 1}$ the trigonometric basis in $\mathbf{L}_2[0, 1]$. For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j)(\theta_{j,n}^*)^2 - 2 \sum_{j=1}^n \lambda(j) \tilde{\theta}_{j,n} + \delta \hat{P}_n(\lambda),$$

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

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

Substituting the weight coefficients, minimizing the cost function

$$\lambda^* = \text{argmin}_{\lambda \in \Lambda} J_n(\lambda), \quad (11)$$

in (7) leads to the improved model selection procedure

$$S^* = S_{\lambda^*}^*. \quad (12)$$

It will be noted that λ^* exists because Λ is a finite set. If the minimizing sequence in (11) λ^* is not unique, one can take any minimizer. In the case, when the value of σ_Q is known, one can take $\hat{\sigma}_n = \sigma_Q$ and $P_n(\lambda) = \sigma_Q |\lambda|_n^2/n$.

Theorem 2. *For any $n \geq 2$ and $0 < \delta < 1/3$, the robust risks (4) of estimate (12) for continuously differentiable function S satisfies the*

oracle inequality

$$\mathcal{R}^*(S_{\lambda^*}^*, S) \leq \frac{1 + 3\delta}{1 - 3\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(S_\lambda^*, S) + \frac{B_n^*}{n\delta}, \quad (13)$$

where the rest term is such that $B_n^*/n^\epsilon \rightarrow 0$ as $n \rightarrow \infty$ for any $\epsilon > 0$.

Remark. The inequality (13) means that the procedure (12) is optimal in the oracle inequalities sense. This property enables to provide asymptotic efficiency in the adaptive setting, i.e. when information about the function regularity is unknown.

References

1. Akaike H. A new look at the statistical model identification // IEEE Trans. on Automatic Control. 1974. V. 19. P. 716–723.
2. Barron A., Birgé L. and Massart P. Risk bounds for model selection via penalization. // Probab. Theory Relat. Fields. 1999. V. 113. P. 301–415.
3. Fourdrinier D., Pergamenschikov S. Improved selection model method for the regression with dependent noise // Annals of the Institute of Statistical Mathematics. 2007. V.59. No. 9. P. 435–464.
4. James W., Stein C. (1961). Estimation with quadratic loss // In Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability. University of California Press, Berkeley. 1961. V. 1. P. 361–380.
5. Konev V. V., Pergamenschikov 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.
6. Konev V. (2014) Estimation of a regression with the pulse type noise from discrete data / Konev V., Pergamenschikov S., Pchelintsev E. // Theory Probab. Appl. 2014. V. 58. No. 3. P. 442–457.
7. Pchelintsev E. Improved estimation in a non-Gaussian parametric regression // Stat. Inference Stoch. Process. 2013. V. 16. No. 1. P. 15 – 28.