JOURNAL OF NONPARAMETRIC STATISTICS
2019, VOL. 31, NO. 3, 612–628
https://doi.org/10.1080/10485252.2019.1609672

Improved robust model selection methods for a Lévy nonparametric regression in continuous time

E. A. Pchelintsev [a], V. A. Pchelintsev [b] and S. M. Pergamenshchikov [c,d]

[a]Department of Mathematical Analysis and Theory of Functions, Tomsk State University, Tomsk, Russia; [b]Division for Mathematics and Computer Sciences, Tomsk Polytechnic University, Tomsk, Russia; [c]Laboratoire de Mathématiques Raphael Salem, Université de Rouen, Rouen, Cedex France; [d]International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, Tomsk, Russia

ABSTRACT
In this paper, we develop the James–Stein improved method for the estimation problem of a nonparametric periodic function observed with Lévy noises in continuous time. An adaptive model selection procedure based on the weighted improved least squares estimates is constructed. The improvement effect for nonparametric models is studied. It turns out that in non-asymptotic setting the accuracy improvement for nonparametric models is more important than for parametric ones. Moreover, sharp oracle inequalities for the robust risks have been shown and the adaptive efficiency property for the proposed procedures has been established. The numerical simulations are given.

ARTICLE HISTORY
Received 27 March 2018
Accepted 15 April 2019

KEYWORDS
Improved non-asymptotic estimation; James–Stein procedures; robust quadratic risk; nonparametric regression; Lévy process; model selection; sharp oracle inequality; adaptive estimation; asymptotic efficiency

AMS SUBJECT CLASSIFICATIONS
62G08; 62G05

1. Introduction

Consider the following nonparametric regression model in continuous time

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

where $S(\cdot)$ is an unknown 1 – periodic function, $(\xi_t)_{0\le t\le n}$ is an unobserved noise. The problem is to estimate the function S on the observations $(y_t)_{0\le t\le n}$. Note that, if $(\xi_t)_{0\le t\le n}$ is Brownian motion, then we obtain the well-known 'signal+white noise' model which is very popular in statistical radio-physics, see, for example, Kutoyants (1977, 1984), Ibragimov and Khasminskii (1981) and Pinsker (1981). In this paper, we assume that in addition to intrinsic noises in radio-electronic systems, approximated usually by the gaussian white or colour noise, the useful signal S is distorted by the impulse flow described by Lévy processes defined in the next section. The cause of a pulse stream can be, for example, either external unintended (atmospheric) or intentional impulse noises or errors in

CONTACT E. A. Pchelintsev evgen-pch@yandex.ru Department of Mathematical Analysis and Theory of Functions, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia

Supplemental data for this article can be accessed here https://doi.org/10.1080/10485252.2019.1609672

the demodulation and the channel decoding for binary information symbols. Note that, for the first time, the impulse noises for the detection signal problems have been studied by Kassam (1988) through compound Poisson processes. Later, such processes was used in Flaksman (2002), Konev and Pergamenshchikov (2012, 2015), Pchelintsev (2013) and Konev, Pergamenshchikov, and Pchelintsev (2014) for parametric and nonparametric signal estimation problems. It should be noted that such models are too limited, since the compound Poisson process can describe only the large impulses influence with a single-fixed frequency. However, the real technical (for example, telecommunication or navigation) systems work under noise impulses having different sizes and different frequencies, see, for example, Proakis (1995). To take this into account, one needs to use many (may be infinite number) different compound Poisson processes in the same observation model. This is possible to do only in a framework of Lévy processes which are natural extensions for the compound Poisson processes. Moreover, it should be noted also that Lévy models are fruitfully used in the different applied problems, see, for example, Bertoin (1996), Barndorff-Nielsen and Shephard (2001), Cont and Tankov (2004) and Comte and Genon-Catalot (2011) and the references therein. In this paper, we consider the adaptive estimation problem for the function S, i.e. when its regularity properties are unknown. 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 the nonparametric estimation through oracle inequalities which give non-asymptotic upper bounds for the quadratic risks including the minimal risk over chosen the estimators family. It will be noted that for the first time the model selection methods were proposed by Mallows (1973) and Akaike (1974) for parametric models. Then, these methods have been developed for nonparametric estimation problems by Barron, Birgé, and Massart (1999) and Fourdrinier and Pergamenshchikov (2007) for regression models in discrete time and Konev and Pergamenshchikov (2010) in continuous time. Unfortunately, the oracle inequalities obtained in these papers cannot 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. To obtain the efficiency property one has to obtain the sharp oracle inequalities, i.e. the inequalities in which the coefficient at the principal term is close to unity. To obtain such inequalities for general non-Gaussian observations one needs to use the method proposed by Konev and Pergamenshchikov (2009a, 2009b, 2012, 2015) for semimartingale models in continuous time based on the model selection tool developed by Galtchouk and Pergamenshchikov (2009a, 2009b) for heteroscedastic non-Gaussian regression models in discrete time.

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 squares estimates. This paper proposes the improved least squares estimates which enable us to improve considerably the non-asymptotic estimation accuracy. For the first time, such idea was proposed by Fourdrinier and Pergamenshchikov (2007) for regression models in discrete time and by Konev and Pergamenshchikov (2010) for Gaussian regression models in continuous time. We develop these methods for the non-Gaussian regression models in continuous time. It should be noted that generally for the conditionally Gaussian regression models we cannot use the well-known improved estimators proposed in James and Stein (1961) and Fourdrinier and Strawderman (1996) 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 Pchelintsev (2013) and Konev et al. (2014) for parametric problems. We use these estimators to construct model selection procedures for nonparametric models. Then to study the efficiency property for the proposed estimation procedure we need to obtain a lower bound for the quadratic risks. Usually, to do this one, uses the van Trees inequality. In this paper, we show the corresponding van Trees inequality for the Lévy regression models and then we derive the needed asymptotic sharp lower bound for the normalised risks, i.e. we find the Pinsker constant for the model (1). As to the upper bound, similarly to Konev and Pergamenshchikov (2009b), we use the obtained sharp oracle inequality for the weighted least squares estimators containing the efficient Pinsker procedure. Therefore, through the oracle inequality, we estimate from above the risk of the proposed procedure by the risk of the efficient Pinsker procedure up to some coefficient which goes to one. As a result, we show the asymptotic efficiency without using the smoothness information of the function S.

The rest of the paper is organised as follows. In Section 2, we describe the noise processes in (1) and define the main risks for the estimation problem. In Section 3, we construct the improved least squares estimates and study the improvement effect for the Lévy model. In Section 4, we construct the improved model selection procedure and show the sharp oracle inequalities. In Section 5, the Monte Carlo simulation results are given. The asymptotic efficiency is studied in Section 6. In Section 7, we prove the van Trees inequality for the model (1). The proofs of the main and auxiliary results are available in Appendices A and B in the supplementary online materials.

2. Noise process model

First, we assume that the noise process $(\xi_t)_{0\le t\le n}$ in (1) is defined as

$$\xi_t = \sigma_1 w_t + \sigma_2 z_t \quad \text{and} \quad z_t = x * (\mu - \tilde{\mu})_t, \tag{2}$$

where σ_1 and σ_2 are some unknown constants, $(w_t)_{t\ge 0}$ is standard Brownian motion, '$*$' denotes the stochastic integral with respect to the compensated jump measure $\mu(\mathrm{d}s\,\mathrm{d}x)$ with deterministic compensator $\tilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, i.e.

$$z_t = \int_0^t \int_{\mathbb{R}_0} x(\mu - \tilde{\mu})(\mathrm{d}s\,\mathrm{d}x).$$

Here $\Pi(\cdot)$ is a Lévy measure, i.e. some positive measure on $\mathbb{R}_0 = \mathbb{R} \setminus \{0\}$, see, for details, Jacod and Shiryaev (2002) and Cont and Tankov (2004) such that

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

We use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_0} |z|^m \Pi(\mathrm{d}z)$. Note that the Lévy measure $\Pi(\mathbb{R}_0)$ may be equal to $+\infty$. It should be noted that in all papers on the nonparametric signal estimation in the model (1) the main condition on the jumps is the finiteness of the Lévy measure, i.e. $\Pi(\mathbb{R}_0) < +\infty$.

The process (2) allows us to consider the several independent impulse noise sources with the different frequencies. Indeed, in this case, see, for example, Cont and Tankov (2004, p. 135), we introduce compound Poisson processes into the model (1) as

$$z_t = \sum_{k=1}^{M} \sum_{j=1}^{N_t^k} Y_{k,j},$$

where $(N_t^1)_{t\geq 0}, \ldots, (N_t^M)_{t\geq 0}$ are independent Poisson processes with the intensities $\lambda_1, \ldots, \lambda_M$ and the sizes of impulses $(Y_{1,j})_{j\geq 1}, \ldots, (Y_{M,j})_{j\geq 1}$ are independent i.i.d. sequences with $\mathbf{E} Y_{k,j} = 0$ and $\varsigma_k^2 = \mathbf{E} Y_{k,j}^2 < \infty$. In this case, the Lévy measure for any Borel set $\Gamma \subseteq \mathbb{R}_0$ is defined as

$$\Pi(\Gamma) = \sum_{k=1}^{M} \lambda_k \mathbf{P}(Y_{k,1} \in \Gamma).$$

Next, note, that if

$$\sum_{k\geq 1} \lambda_k \varsigma_k^2 < \infty,$$

then we can introduce the infinite number of the noise jumps setting

$$z_t = \sum_{k=1}^{\infty} \sum_{j=1}^{N_t^k} Y_{k,j}.$$

Moreover, if the total noise intensity $\sum_{k\geq 1} \lambda_k = +\infty$, then $\Pi(\mathbb{R}_0) = +\infty$, i.e. we obtain the observation model with saturated impulse noise.

In the sequel, we will denote by Q the distribution of the process $(\xi_t)_{0\leq t\leq n}$ in the Skorokhod space $\mathbf{D}[0, n]$ and by $\mathcal{Q}_n$ we denote all these distributions for which the parameters σ_1 and σ_2 satisfy the conditions

$$0 < \sigma_* \leq \sigma_1^2 \quad \text{and} \quad \sigma = \sigma_1^2 + \sigma_2^2 \leq \sigma^*, \tag{3}$$

where the bounds σ_* and σ^* are functions of n, i.e. $\sigma_* = \sigma_*(n)$ and $\sigma^* = \sigma^*(n)$, such that for any $\epsilon > 0$

$$\liminf_{n\to\infty} n^{\epsilon} \sigma_*(n) > 0 \quad \text{and} \quad \lim_{n\to\infty} n^{-\epsilon} \sigma^*(n) = 0. \tag{4}$$

We also assume that the distribution Q of the noise process $(\xi_t)_{0\leq t\leq n}$ is unknown. We know only that this distribution belongs to the distribution family $\mathcal{Q}_n$ defined in (3)–(4). By these reasons, we use the robust estimation approach developed for nonparametric problems in Galtchouk and Pergamenshchikov (2006) and Konev and Pergamenshchikov (2012, 2015). To this end, we will measure the estimation quality by the robust risk defined as

$$\mathcal{R}_n^*(\hat{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\hat{S}_n, S), \tag{5}$$

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

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

The first goal in this paper is to develop shrinkage nonparametric estimation methods for S which improve the non-asymptotic robust estimation accuracy (5) with respect to the well-known least squares estimators. The next goal is to provide non-asymptotic optimality in the sense of sharp oracle inequalities. Moreover, asymptotically, as $n \to \infty$, our goal is to show the efficiency property for the proposed shrinkage estimators for the risks (5).

3. James–Stein improved method

Let $(\phi_j)_{j\ge 1}$ 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 $t \in \mathbb{R}$. We assume the following condition for these functions.

($\mathbf{B}_1$) The functions $(\phi_j)_{j\ge 1}$ are uniformly bounded, i.e. for some $\phi^* > 0$

$$\sup_{0\le j\le n}\ \sup_{0\le t\le 1} |\phi_j(t)| \le \phi^* < \infty. \tag{7}$$

For example, we can take the trigonometric basis defined as $\mathrm{Tr}_1 \equiv 1$ and for $j \ge 2$

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

where $[a]$ denotes integer part of a.

For estimating the unknown function S in (1) we consider the Fourier expansion $S(t) = \sum_{j=1}^{\infty}\theta_j\phi_j(t)$. The corresponding Fourier coefficients

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

can be estimated as

$$\hat{\theta}_j = \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t.$$

In view of (1), we obtain

$$\hat{\theta}_j = \theta_j + \frac{1}{\sqrt{n}}\xi_j, \tag{9}$$

where

$$\xi_j = \frac{1}{\sqrt{n}} I_n(\phi_j) \quad \text{and} \quad I_n(f) = \int_0^n f(t)\,\mathrm{d}\xi_t.$$

As Konev and Pergamenshchikov (2009b), we define a class of weighted least squares estimates for $S(t)$

$$\hat{S}_\lambda = \sum_{j=1}^{n}\lambda(j)\hat{\theta}_j\phi_j, \tag{10}$$

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

$$\nu_n = \mathrm{card}(\Lambda) \quad \text{and} \quad |\Lambda|_n = \max_{\lambda\in\Lambda} L(\lambda), \tag{11}$$

where $\mathrm{card}(\Lambda)$ is the number of the vectors λ in Λ and $L(\lambda) = \sum_{j=1}^{n} \lambda(j)$. In the sequel, we assume that all vectors from Λ satisfies the following condition.

($\mathbf{B}_2$) 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 $1 \le j \le d$ for any $\lambda \in \Lambda$.

Remark 3.1: Note that the weight coefficients satisfying the condition ($\mathbf{B}_2$) was introduced by Nussbaum (1985) to construct the efficient estimation for the nonparametric regression model in discrete time.

Now we need the σ – field generated by the jumps of the process (2), i.e. we set $\mathcal{G}_n = \sigma\{z_t,\ 0 \le t \le n\}$. To construct the improved estimators, we need the following proposition.

Proposition 3.1: *For any $n \ge 1$ the random vector $\tilde{\xi}_{d,n} = (\xi_j)_{1\le j\le d}$ is the $\mathcal{G}_n$ – conditionally Gaussian in $\mathbb{R}^d$ with zero mean and the covariance matrix $\mathbf{G}_n = (\mathbf{E}\xi_i\xi_j \mid \mathcal{G}_n)_{1\le i,\, j\le d}$ such that*

$$\inf_{Q\in\mathcal{Q}_n} (tr\mathbf{G}_n - \lambda_{\max}(\mathbf{G}_n)) \ge (d-1)\sigma_*, \tag{12}$$

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

Now, for the first d Fourier coefficients in (9), we use the improved estimation method proposed for parametric models in Pchelintsev (2013). To this end, we set $\tilde{\theta}_\lambda = (\hat{\theta}_j)_{1\le j\le d}$. We recall that the parameter d is dependent of λ. In the sequel, we will use the norm $|x|_d^2 = \sum_{j=1}^{d} x_j^2$ for any vector $x = (x_j)_{1\le j\le d}$ from $\mathbb{R}^d$. Now we define the shrinkage estimators as

$$\theta^*_{\lambda,j} = \left(1 - g_\lambda(j)\right)\hat{\theta}_j \quad \text{and} \quad g_\lambda(j) = \frac{c_n}{|\tilde{\theta}_\lambda|_d}\mathbf{1}_{\{1\le j\le d\}}, \tag{13}$$

where

$$c_n = c_n(\lambda) = \frac{(d-1)\sigma_*}{\left(\mathbf{r} + \sqrt{d\sigma^*/n}\right)n}$$

and the threshold $\sigma^* > 0$ is given in the lower bound (12). The positive parameter $\mathbf{r}$ is a function of n, i.e. $\mathbf{r} = \mathbf{r}(n)$ such that

$$\lim_{n\to\infty} \mathbf{r}(n) = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathbf{r}(n)}{n^\epsilon} = 0 \tag{14}$$

for any $\epsilon > 0$.

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

$$S_\lambda^* = \sum_{j=1}^{n} \lambda(j)\theta_{\lambda,j}^* \phi_j. \tag{15}$$

We denote the difference of quadratic risks of the estimates (10) and (15) as

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

Now we study the robust accuracy comparison between these estimators, i.e. uniformly over the distribution family $\mathcal{Q}_n$.

Theorem 3.2: *Let the observed process* $(y_t)_{0\le t\le n}$ *describes by the Equations* (1)–(2). *Then for any* $n \geq 1$

$$\sup_{Q\in\mathcal{Q}_n} \sup_{\|S\|\le \mathbf{r}} \Delta_Q(S) \le -c_n^2. \tag{16}$$

Remark 3.2: The inequality (16) means that non-asymptotically, i.e. for non large $n \geq 1$, the estimate (15) outperforms in mean square robust accuracy the estimate (10). As we will see later in the efficient weight coefficients $d \approx n^\epsilon$ as $n \to \infty$ for some $\epsilon > 0$. Therefore, in view of the definition of the constant c_n in (13) and the conditions (4) and (14) $nc_n \to \infty$ as $n \to \infty$. It should be noted also, that for the parametric regression the parameter dimension d is fixed Pchelintsev (2013), i.e. the James–Stein improved method is essentially efficient for nonparametric models.

4. Model selection

In this section, we construct a model selection procedure for the estimation of S in (1) on the basis of the weighted shrinkage estimators (15). To this end, we consider the empirical squared error defined as

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

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (15). It is obvious, that the best way is to minimise the empirical squared error with respect to λ. Making use the estimate definition (15) and the Fourier transformation of S implies

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

Since the Fourier coefficients $(\theta_j)_{j\ge1}$ are unknown, the weight coefficients $(\lambda_j)_{j\ge1}$ cannot be found by minimising this quantity. To circumvent this difficulty, one needs to replace the terms $\theta^*_{\lambda,j}\theta_j$ by their estimators $\bar{\vartheta}_{\lambda,j}$ defined as

$$\bar{\vartheta}_{\lambda,j} = \theta^*_{\lambda,j}\hat{\theta}_j - \frac{\hat{\sigma}_n}{n}, \tag{17}$$

where $\hat{\sigma}_n$ is the estimate for the limiting variance of $\sigma = \mathbf{E}_Q \xi_j^2$ which we choose in the following form

$$\hat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \hat{t}_j^2 \quad \text{and} \quad \hat{t}_j = \frac{1}{n}\int_0^n \mathrm{Tr}_j(t)\,\mathrm{d}y_t. \tag{18}$$

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

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

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

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

We define the improved model selection procedure as

$$S^* = S^*_{\lambda^*} \quad \text{and} \quad \lambda^* = \mathrm{argmin}_{\lambda\in\Lambda} J_n(\lambda). \tag{19}$$

It will be noted that λ^* exists because Λ is a finite set. If the minimising sequence in (19) λ^* is not unique, one can take any minimiser. Now, to write the oracle inequality, we set

$$\Psi_{Q,n} = (1 + (\phi^*)^4)(1 + \sigma^*)(1 + c_n^*)\nu_n,$$

where $c_n^* = n \max_{\lambda\in\Lambda} c_n^2(\lambda)$. It is useful to note that in view of the first condition in (4) and the properties (14) the constant c_n^* is not large as $n \to \infty$, i.e. for any $\epsilon > 0$

$$\lim_{n\to\infty} \frac{c_n^*}{n^\epsilon} = 0.$$

First we study the non-asymptotic properties for the procedure (19).

Theorem 4.1: *There exists some constant* $\check{\mathbf{l}} > 0$ *such that for any* $n \geq 1$ *and* $0 < \delta < 1/2$, *the risk* (6) *of estimate* (19) *for* S *satisfies the oracle inequality*

$$\mathcal{R}_Q(S^*, S) \leq \frac{1+5\delta}{1-\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(S^*_\lambda, S) + \check{\mathbf{l}} \frac{\Psi_{Q,n}}{n\delta} + \frac{12|\Lambda|_n \mathbf{E}_Q |\hat{\sigma}_n - \sigma|}{n}. \tag{20}$$

In the case, when the value of σ is known, one can take $\hat{\sigma}_n = \sigma$ and

$$P_n(\lambda) = \frac{\sigma |\lambda|_n^2}{n},$$

then we can rewrite the oracle inequality (20) in the following form

$$\mathcal{R}_Q(S^*, S) \leq \frac{1+5\delta}{1-\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(S^*_\lambda, S) + \check{\mathbf{l}} \frac{\Psi_{Q,n}}{n\delta}.$$

Also we study the accuracy properties for the estimator (18).

Proposition 4.2: *Let in the model* (1) *the function* $S(\cdot)$ *is continuously differentiable. Then, there exists some constant* $\check{\mathbf{l}} > 0$ *such that for any* $n \geq 2$ *and* S

$$\mathbf{E}_Q |\hat{\sigma}_n - \sigma| \leq \check{\mathbf{l}} \frac{(1 + \|\dot{S}\|^2)}{\sqrt{n}},$$

where $\dot{S}$ *is the derivative of the function* S.

Remark 4.1: It should be noted that to estimate the parameter σ in (17) we use the equality (9) for the Fourier coefficients $t_j = (S, \mathrm{Tr}_j)$ with respect to the trigonometric basis (8), since, as is shown in Lemma A.6 in Konev and Pergamenshchikov (2009a) for any continuously differentiable function S and for any $m \geq 1$ the sum $\sum_{j \geq m} t_j^2$ can be estimated from above in an explicit form. Therefore, through the trigonometric basis we can estimate the variance σ uniformly over the functions S, when we will study the efficiency property for the proposed procedures.

To obtain the oracle inequality for the robust risk we impose the following additional conditions.

($\mathbf{C}_1$) Assume that the upper bound for the basic function ϕ^* in (7) is a function of n, i.e. $\phi^* = \phi^*(n)$, such that for any $\epsilon > 0$

$$\lim_{n \to \infty} \frac{\phi^*(n)}{n^\epsilon} = 0.$$

($\mathbf{C}_2$) Assume that the set Λ is such that for any $\epsilon > 0$

$$\lim_{n \to \infty} \frac{\nu_n}{n^\epsilon} = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{|\Lambda|_n}{n^{1/2+\epsilon}} = 0. \tag{21}$$

We note that Theorem 4.1 and Proposition 4.2 directly imply the following inequality.

Theorem 4.3: *If the conditions* ($\mathbf{C}_1$)–($\mathbf{C}_2$) *hold for the distribution* Q *of the process* ξ *in* (1), *then for any* $n \geq 2$ *and* $0 < \delta < 1/2$, *the robust risk* (5) *of estimate* (19) *for continuously differentiable function* S *satisfies the oracle inequality*

$$\mathcal{R}_n^*(S^*, S) \leq \frac{1+5\delta}{1-\delta} \min_{\lambda \in \Lambda} \mathcal{R}_n^*(S^*_\lambda, S) + \frac{B_n(1 + \|\dot{S}\|^2)}{n\delta}, \tag{22}$$

where the term B_n *is independent of* S *and for any* $\epsilon > 0$

$$\lim_{n \to \infty} \frac{B_n}{n^\epsilon} = 0.$$

Remark 4.2: Note that sharp oracle inequalities similar to (20) and (22) was obtained earlier by Konev and Pergamenshchikov (2009a, 2012, 2015) for model selection procedures based on the weighted least squares estimates (10). Unfortunately, we cannot use such oracle inequalities for the model selection procedures, based on the weighted shrinkage estimates (15) since they depend non-linearly on the coefficients λ. This is a main technical difficulty which does not allow us to use the obtained oracle inequalities. Moreover, in all these papers, the oracle inequalities are obtained under condition that the Lévy measure is finite. The inequalities (20) and (22) are obtained without conditions on the impulse noises.

Now we specify the weight coefficients $(\lambda(j))_{j\geq 1}$ in the way proposed in Galtchouk and Pergamenshchikov (2009a) for a heteroscedastic regression model in discrete time. Consider a numerical grid of the form

$$\mathcal{A}_n = \{1,\ldots,k_n\} \times \{r_1,\ldots,r_m\},$$

where $r_i = i\rho_n$ and $m = [1/\rho_n^2]$. Both parameters $k_n \geq 1$ and $0 < \rho_n \leq 1$ are the functions of n such that for any $\epsilon > 0$

$$\begin{aligned} &\lim_{n\to\infty} k_n = +\infty, \quad \lim_{n\to\infty} \frac{k_n}{\ln n} = 0, \\ &\lim_{n\to\infty} \rho_n = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\epsilon}\rho_n = +\infty. \end{aligned} \tag{23}$$

One can take, for example,

$$\rho_n = \frac{1}{\ln(n+1)} \quad \text{and} \quad k_n = k_* + \sqrt{\ln(n+1)}$$

for some fixed $k_* \geq 0$. For each $\alpha = (\beta, r) \in \mathcal{A}_n$, we introduce the weight sequence $\lambda_\alpha = (\lambda_j(\alpha))_{j\geq 1}$ as

$$\lambda_j(\alpha) = \mathbf{1}_{\{1\leq j\leq d\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{d<j\leq\omega_\alpha\}} \tag{24}$$

where $d = d(\alpha) = [\omega_\alpha / \ln(n+1)]$,

$$\omega_\alpha = \left(\tau_\beta r v_n\right)^{1/(2\beta+1)}, \quad \tau_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad v_n = \frac{n}{\sigma^*}.$$

We set

$$\Lambda = \{\lambda(\alpha),\, \alpha \in \mathcal{A}_n\}. \tag{25}$$

It will be noted that in this case $\nu_n = k_n m$. Therefore, the conditions (23) imply the first limit equality in (21). Moreover, in view of the definition (24) and taking into account that $\tau_\beta \leq 1$ for $\beta \geq 1$ the function $L(\lambda)$ defined in (11) can be estimated for any $\lambda \in \Lambda$ as

$$\max_{\lambda\in\Lambda} L(\lambda) \leq \max_{\lambda\in\Lambda} \omega_\alpha \leq v_n^{1/3}\rho_n^{-1/3}.$$

Therefore, using here the conditions (4) and (23), we get the last limit in (21), i.e. the condition $\mathbf{C}_2$) holds for the set Λ defined in (25).

Remark 4.3: It will be observed that the specific form of weights (24) was proposed by Pinsker (1981) for the filtration problem with known smoothness of the regression function observed with an additive gaussian white noise in continuous time. Nussbaum (1985) used such weights for the gaussian regression estimation problem in discrete time.

5. 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 (19). We simulate the model (1) with 1-periodic function S of the form

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

on $[0, 1]$ and the Lévy noise process ξ_t is defined as

$$\xi_t = 0.5 w_t + 0.5 z_t.$$

Here z_t is a compound Poisson process with intensity $\lambda = \Pi(x^2) = 1$ and a Gaussian $\mathcal{N}(0, 1)$ sequence $(Y_j)_{j\geq 1}$, see, for example, Konev and Pergamenshchikov (2015).

We use the model selection procedure (19) with the weights (24) in which $k_n = 100 + \sqrt{\ln(n+1)}$, $r_i = i/\ln(n+1)$, $m = [\ln^2(n+1)]$, $\sigma^* = 0.5$ and $\delta = (3+\ln n)^{-2}$. We define the empirical risk as

$$\mathcal{R}(S^*, S) = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\left(S^*_n(t_j) - S(t_j)\right)^2,$$

$$\hat{\mathbf{E}}\left(S^*_n(\cdot) - S(\cdot)\right)^2 = \frac{1}{N}\sum_{l=1}^{N}\left(S^*_{n,l}(\cdot) - S(\cdot)\right)^2,$$

where the observation frequency $p = 100{,}001$ and the expectations was taken as an average over $N = 1000$ replications.

Table 1 gives the values for the sample risks of the improved estimate (19) and the model selection procedure based on the weighted LSE (3.15) from Konev and Pergamenshchikov (2012) 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 Konev and Pergamenshchikov (2012) and its improved version for different numbers of observation period n.

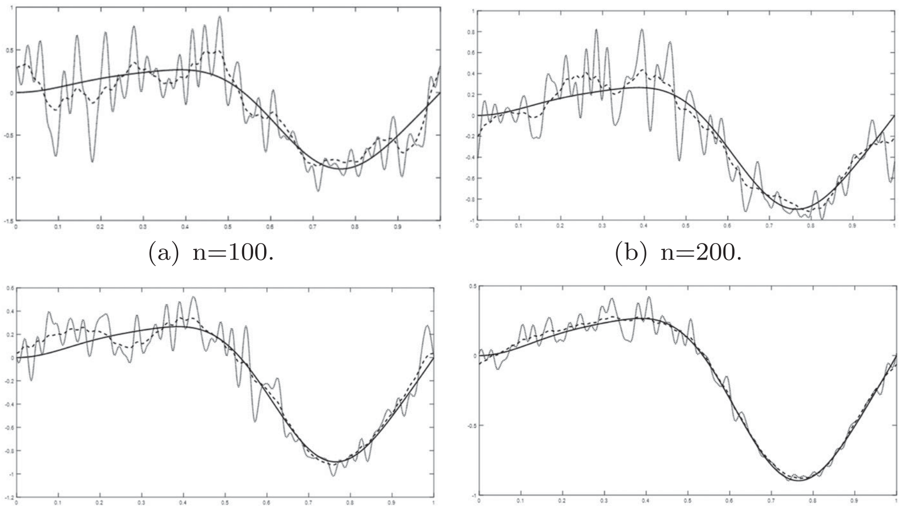
Remark 5.1: Figure 1 shows the behaviour of the procedures (10) and (19) depending on the values of observation periods n. The bold line is the function (26), the continuous line

Table 1. The sample quadratic risks for different optimal λ.

n	100	200	500	1000
$\mathcal{R}(S^*_{\lambda^*}, S)$	0.0118	0.0089	0.0031	0.0009
$\mathcal{R}(\hat{S}_{\hat{\lambda}}, S)$	0.0509	0.0203	0.0103	0.0064
$\mathcal{R}(\hat{S}_{\hat{\lambda}}, S)/\mathcal{R}(S^*_{\lambda^*}, S)$	4.3	2.3	3.3	7.1

Table 2. The sample quadratic risks for the same optimal $\hat{\lambda}$.

n	100	200	500	1000
$\mathcal{R}(S^*_{\hat{\lambda}}, S)$	0.0237	0.0103	0.0041	0.0011
$\mathcal{R}(\hat{S}_{\hat{\lambda}}, S)$	0.0509	0.0203	0.0103	0.0064
$\mathcal{R}(\hat{S}_{\hat{\lambda}}, S)/\mathcal{R}(S^*_{\hat{\lambda}}, S)$	2.1	2.2	2.5	5.8



(a) n=100. (b) n=200. (c) n=500. (d) n=1000.

Figure 1. Behaviour of the regression function and its estimates.

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 Table 2 for the same λ with various observations numbers n we can conclude that theoretical result on the improvement effect (16) is confirmed by the numerical simulations. Moreover, for the proposed shrinkage procedure, Table 1 and Figure 1, we can conclude that the benefit is considerable for non large n.

6. Asymptotic efficiency

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

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

where $r > 0$ 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 $0 \leq i \leq k-1$. In order to formulate our asymptotic results, we define the Pinsker constant which gives the lower bound for normalised asymptotic risks

$$l_k(r) = ((1+2k)r)^{1/(2k+1)} \left(\frac{k}{\pi(k+1)} \right)^{2k/(2k+1)}. \tag{27}$$

It is well known that for any $S \in W_{k,r}$ the optimal rate of convergence is $n^{-2k/(2k+1)}$, see, for example, Pinsker (1981) and Nussbaum (1985). On the basis of the model selection procedure, we construct the adaptive procedure S^* for which we obtain the following asymptotic upper bound for the quadratic risk, i.e. we show that the parameter (27) gives a lower bound for the asymptotic normalised risks. To this end we denote by Σ_n the set of all estimators $\hat{S}_n$ of S measurable with respect to the process (1), i.e. measurable with respect to σ-field $\sigma\{y_t,\ 0 \le t \le n\}$.

Theorem 6.1: *The robust risk* (5) *admits the following asymptotic lower bound*

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

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

Theorem 6.2: *The quadratic risk* (5) *for the estimating procedure S^* has the following asymptotic upper bound*

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

Theorem 6.2 follows from Theorem 3.2 and Theorem 3.1 in Konev and Pergamenshchikov (2009b). It is clear that Theorem 6.2 and Theorem 6.1 imply

Corollary 6.3: *The model selection procedure S^* is asymptotically efficient, i.e.*

$$\lim_{n\to\infty} (v_n)^{2k/(2k+1)} \sup_{S\in W_{k,r}} \mathcal{R}_n^*(S^*, S) = l_k(r). \tag{28}$$

Remark 6.1: Note that the equality (28) implies that the parameter (27) is the Pinsker constant in this case Pinsker (1981).

Remark 6.2: It should be noted that the equality (28) means that the robust efficiency holds with the convergence rate

$$(v_n)^{2k/(2k+1)}.$$

It is well known that for the simple risks the optimal (minimax) estimation convergence rate for the functions from the set $W_{k,r}$ is $n^{2k/(2k+1)}$, see, for example, Ibragimov and Khasminskii (1981), Pinsker (1981) and Nussbaum (1985). So, if the distribution upper bound $\sigma^* \to 0$ as $n \to \infty$ we obtain the more rapid rate, and if $\sigma^* \to \infty$ as $n \to \infty$ we obtain the more slow rate. In the case when σ^* is constant the robust rate is the same as the classical non-robust convergence rate.

Remark 6.3: The property (28) means that the model selection procedure (19) asymptotically has the same efficiency property as the LSE model selection, see, Galtchouk and Pergamenshchikov (2009b) and Konev and Pergamenshchikov (2009b). So, it means that the proposed shrinkage method non-asymptotically has benefit with respect to LSE and asymptotically the shrinkage methods keep the efficiency property.

7. The van Trees inequality for the Lévy processes

In this section, we consider the following continuous time parametric regression model (1) with the function S defined as

$$S(t,\theta) = \sum_{i=1}^{d} \theta_i \psi_i(t),$$

with the unknown parameters $\theta = (\theta_1, \ldots, \theta_d)'$. Here we assume that the functions $(\psi_i)_{1\le i\le d}$ are 1-periodic and orthogonal functions.

Let us denote by ν_ξ the distribution of the process $(\xi_t)_{0\le t\le n}$ on the Skorokhod space $\mathbf{D}[0,n]$. From Proposition A.4, it follows that in this space for any parameters $\theta \in \mathbb{R}^d$, the distribution $\mathbf{P}_\theta$ of the process (1) is absolutely continuous with respect to the ν_ξ and the corresponding Radon-Nikodym derivative, for any function $x = (x_t)_{0\le t\le n}$ from $\mathbf{D}[0,n]$, is defined as

$$f(x,\theta) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\nu_\xi}(x) = \exp\left\{ \int_0^n \frac{S(t,\theta)}{\sigma_1^2} \mathrm{d}x_t^c - \int_0^n \frac{S^2(t,\theta)}{2\sigma_1^2} \mathrm{d}t \right\},$$

where

$$x_t^c = x_t - \int_0^t \int_{\mathbb{R}_0} v\,(\mu_x(\mathrm{d}s,\mathrm{d}v) - \Pi(\mathrm{d}v)\,\mathrm{d}s)$$

and for any measurable set A in $\mathbb{R}$ with $0 \notin A$

$$\mu_x([0,t]\times A) = \sum_{0\le s\le t} \mathbf{1}_{\{\Delta\xi_s \in \sigma_2 A\}}.$$

Let U be a prior density on $\mathbb{R}^d$ having the following form:

$$U(\theta) = U(\theta_1, \ldots, \theta_d) = \prod_{j=1}^{d} u_j(\theta_j),$$

where u_j is some continuously differentiable density in $\mathbb{R}$. Moreover, let $g(\theta)$ be a continuously differentiable $\mathbb{R}^d \to \mathbb{R}$ function such that, for each $1 \le j \le d$,

$$\lim_{|\theta_j|\to\infty} g(\theta) u_j(\theta_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |g_j'(\theta)| U(\theta)\,\mathrm{d}\theta < \infty,$$

where $g_j'(\theta) = \partial g(\theta)/\partial\theta_j$. For any $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathbb{R}^d)$-measurable integrable function $H = H(x,\theta)$, we denote

$$\tilde{\mathbf{E}}H = \int_{\mathbb{R}^d}\int_{\mathcal{X}} H(x,\theta)\,\mathrm{d}\mathbf{P}_\theta\, U(\theta)\,\mathrm{d}\theta$$
$$= \int_{\mathbb{R}^d}\int_{\mathcal{X}} H(x,\theta) f(x,\theta) U(\theta)\,\mathrm{d}\nu_\xi(x)\,\mathrm{d}\theta,$$

where $\mathcal{X} = \mathbf{D}[0,n]$.

Lemma 7.1: *For any $\mathcal{F}_n^y$-measurable square integrable function $\hat{g}_n$ and for any $1 \le j \le d$, the following inequality holds*

$$\tilde{\mathbf{E}}(\hat{g}_n - g(\theta))^2 \ge \frac{\eta_j^2}{n\|\psi_j\|^2\sigma_1^{-2} + I_j},$$

where

$$\eta_j = \int_{\mathbb{R}^d} g_j'(\theta) U(\theta)\,\mathrm{d}\theta \quad \textit{and} \quad I_j = \int_{\mathbb{R}} \frac{\dot{u}_j^2(z)}{u_j(z)}\mathrm{d}z.$$

8. Conclusion

In the conclusion, we would like to emphasise that in this paper we develop a new model selection method based on the improved versions of the least squares estimates. It turns out that the improvement effect in the nonparametric estimation given in (16) is more important than for the parameter estimation problems since the accuracy improvement is proportional to the parameter dimension d which goes to infinity for nonparametric models. Recall that, the improved estimation methods was usually used for the parametric estimation problem only, where the parameter dimension d is always fixed, see, for example, Fourdrinier and Strawderman (1996). Therefore, the benefit in the non-asymptotic quadratic accuracy from the application of the improved estimation methods is more significant in statistical nonparametric signal processing. Moreover, for the proposed improved model selection procedures we obtain the sharp oracle inequalities. It should be emphasised that in this paper we obtain these inequalities without conditions on the jumps, i.e. without assumption that the Lévy measure is finite. To this end we developed a special analytical tool in Proposition A.2 to study the non-asymptotic properties for the corresponding stochastic integrals with respect to the process (2). Moreover, asymptotically, as n goes to infinity, we shown the adaptive efficiency for the improved model selection procedures. This is the meaning that the proposed shrinkage model selection procedures have the benefit with respect to the least squares estimator in the non-asymptotic accuracy and asymptotically they possess the same efficient properties as the least squares methods. Moreover, the behaviour of the constructed procedures is illustrated by the numerical simulations in Section 5.

Acknowledgments

The authors are grateful to the anonymous referees and to the AE for careful reading and for helpful comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Russian Science Foundation (RSF) [grant number 17-11-01049]. This work also was partially supported by the Ministry of Science and Higher Education of the Russian Federation [grants numbers 2.3208.2017/4.6 and 1.472.2016/1.4].

ORCID

E. A. Pchelintsev http://orcid.org/0000-0001-7496-2606
V. A. Pchelintsev http://orcid.org/0000-0001-7521-1177
S. M. Pergamenshchikov http://orcid.org/0000-0002-1896-4030

References

Akaike, H. (1974), 'A New Look at the Statistical Model Identification', *IEEE Transactions on Automatic Control*, 19, 716–723. doi:10.1109/TAC.1974.1100705.

Barndorff-Nielsen, O.E., and Shephard, N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics', *Journal of the Royal Statistical Society: Series B*, 63, 167–241. doi:10.1111/1467-9868.00282.

Barron, A., Birgé, L., and Massart, P. (1999), 'Risk Bounds for Model Selection Via Penalization', *Probability Theory and Related Fields*, 113, 301–413. doi:10.1007/s004400050210.

Bertoin, J. (1996), *Lévy Processes*, Cambridge: Cambridge University Press.

Comte, F., and Genon-Catalot, V. (2011), 'Estimation for Lévy Processes From High Frequency Data Within a Long Time Interval', *The Annals of Statistics*, 39, 803–837. doi:10.1214/10-AOS856.

Cont, R., and Tankov, P. (2004), *Financial Modelling with Jump Processes*, London: Chapman & Hall.

Flaksman, A.G. (2002), 'Adaptive Signal Processing in Antenna Arrays with Allowance for the Rank of the Impule-response Matrix of a Multipath Channel', *Radiophysics and Quantum Electronics*, 45, 977–988.

Fourdrinier, D., and Pergamenshchikov, S. (2007), 'Improved Model Selection Method for a Regression Function with Dependent Noise', *Annals of the Institute of Statistical Mathematics*, 59, 435–464. doi:10.1007/s10463-006-0063-7.

Fourdrinier, D., and Strawderman, W.E. (1996), 'A Paradox Concerning Shrinkage Estimators: Should a Known Scale Parameter be Replaced by an Estimated Value in the Shrinkage Factor?', *Journal of Multivariate Analysis*, 59, 109–140. doi:10.1006/jmva.1996.0056.

Galtchouk, L.I., and Pergamenshchikov, S.M. (2006), 'Asymptotically Efficient Estimates for Nonparametric Regression Models', *Statistics and Probability Letters*, 76, 852–860. doi:10.1016/j.spl.2005.10.017.

Galtchouk, L.I., and Pergamenshchikov, S.M. (2009a), 'Sharp Non-asymptotic Oracle Inequalities for Non-parametric Heteroscedastic Regression Models', *Journal of Nonparametric Statistics*, 21, 1–18. doi:10.1080/10485250802504096.

Galtchouk, L., and Pergamenshchikov, S. (2009b), 'Adaptive Asymptotically Efficient Estimation in Heteroscedastic Nonparametric Regression', *Journal of the Korean Statistical Society*, 38, 305–322. doi:j.jkss.2008.12.001.

Ibragimov, I.A., and Khasminskii, R.Z. (1981), *Statistical Estimation: Asymptotic Theory*, New York: Springer.

Jacod, J., and Shiryaev, A.N. (2002), *LimitTheorems forStochasticProcesses* (2nd ed.), Berlin: Springer.

James, W., and Stein, C. (1961), 'Estimation with quadratic loss', in *Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability*, Berkeley: University of California Press, 1, pp. 361–380.

Kassam, S.A. (1988), *SignalDetection inNon-GaussianNoise*, New York: Springer.

Konev, V.V., and Pergamenshchikov, S.M. (2009a), 'Nonparametric Estimation in a Semimartingale Regression Model. Part 1. Oracle Inequalities', *Journal of Mathematics and Mechanics of Tomsk State University*, 3, 23–41.
Konev, V.V., and Pergamenshchikov, S.M. (2009b), 'Nonparametric Estimation in a Semimartingale Regression Model. Part 2. Robust Asymptotic Efficiency', *Journal of Mathematics and Mechanics of Tomsk State University*, 4, 31–45.
Konev, V.V., and Pergamenshchikov, S.M. (2010), 'General Model Selection Estimation of a Periodic Regression with a Gaussian Noise', *Annals of the Institute of Statistical Mathematics*, 62, 1083–1111. doi:10.1007/s10463-008-0193-1.
Konev, V.V., and Pergamenshchikov, S.M. (2012), 'Efficient Robust Nonparametric Estimation in a Semimartingale Regression Model', *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48, 1217–1244. doi:10.1214/12-AIHP488.
Konev, V.V., and Pergamenshchikov, S.M. (2015), 'Robust Model Selection for a Semimartingale Continuous Time Regression From Discrete Data', *Stochastic Processes and their Applications*, 125, 294–326. doi:10.1016/j.spa.2014.08.003.
Konev, V., Pergamenshchikov, S., and Pchelintsev, E. (2014), 'Estimation of a Regression with the Pulse Type Noise From Discrete Data', *Theory of Probability & its Applications*, 58, 442–457. doi:10.1137/S0040585X9798662X.
Kutoyants, Yu.A. (1977), 'Estimation of the Signal Parameter in a Gaussian Noise', *Problems of Information Transmission*, 13, 29–36.
Kutoyants, Yu.A. (1984), *Parameter Estimation for Stochastic Processes*, Berlin: Heldeman-Verlag.
Mallows, C. (1973), 'Some Comments on C_p', *Technometrics*, 15, 661–675.
Nussbaum, M. (1985), 'Spline Smoothing in Regression Models and Asymptotic Efficiency in $\mathbf{L}_2$', *The Annals of Statistics*, 13, 984–997. doi:10.1214/aos/1176349651.
Pchelintsev, E. (2013), 'Improved Estimation in a Non-Gaussian Parametric Regression', *Statistical Inference for Stochastic Processes*, 16, 15–28. doi:10.1007/s11203-013-9075-0.
Pinsker, M.S. (1981), 'Optimal Filtration of Square Integrable Signals in Gaussian White Noise', *Problems of Transimission Information*, 17, 120–133.
Proakis, J.G. (1995), *Digital Communications*, New York: McGraw-Hill.