Annals of the Institute of Statistical Mathematics (2020) 72:1205–1235
https://doi.org/10.1007/s10463-019-00726-2

Model selection for the robust efficient signal processing observed with small Lévy noise

Slim Beltaief[1] **· Oleg Chernoyarov**[2] **· Serguei Pergamenchtchikov**[3]

Received: 23 April 2018 / Revised: 29 April 2019 / Published online: 19 July 2019

Abstract
We develop a new model selection method for an adaptive robust efficient nonparametric signal estimation observed with impulse noise which is defined by a general non-Gaussian Lévy process. On the basis of the developed method, we construct estimation procedures which are analyzed in two settings: in non-asymptotic and in asymptotic ones. For the first time for such models, we show non-asymptotic sharp oracle inequalities for quadratic and robust risks, i.e., we show that the constructed procedures are optimal in the sense of sharp oracle inequalities. Next, by making use of the obtained oracle inequalities, we provide an asymptotic efficiency property for the developed estimation methods in an adaptive setting when the signal/noise ratio goes to infinity. We apply the developed model selection methods for the signal number detection problem in multi-path information transmission.

Keywords Model selection · Non-asymptotic estimation · Robust estimation · Oracle inequalities · Efficient estimation · Statistical signal processing techniques and analysis

1 Introduction

In this paper, we consider a signal estimation problem on the basis of observations defined by a nonparametric regression model in continuous time with impulse noises of small intensity, i.e.,

This work was supported by the RSF Grant 17-11-01049 (National Research Tomsk State University).
The online version of this article contains supplementary material.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-019-00726-2) contains supplementary material, which is available to authorized users.

✉ Serguei Pergamenchtchikov
serge.pergamenchtchikov@univ-rouen.fr

Extended author information available on the last page of the article

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

where $S(\cdot)$ is an unknown deterministic signal (i.e., $[0, 1] \to \mathbb{R}$ is a non-random function), $(\xi_t)_{0\le t\le 1}$ is an unobserved noise and $\varepsilon > 0$ is the noise intensity. The problem is to estimate the function S based on the observations $(y_t)_{0\le t\le 1}$ when $\varepsilon \to 0$. Note that if $(\xi_t)_{0\le t\le 1}$ is a Brownian motion, then we obtain a "signal+white noise" model which is very popular in statistical radio-physics and is well studied by many authors: Ibragimov and Khasminskii (1981), Pinsker (1981) and Kutoyants (1984, 1994). The condition $\varepsilon \to 0$ means that the signal/noise ratio goes to infinity. In this paper, we assume that in addition to the intrinsic noise in the radio-electronic system, approximated usually by Gaussian white noise, the useful signal S is distorted by an impulse noise flow defined by Lévy process with jumps introduced in the next section. The cause of the appearance of the impulse stream in the radio-electronic systems can be, for example, either external unintended (atmospheric) noises, intentional impulse noises or errors in the demodulation and channel decoding for the binary information symbols. Note that the impulse noises for the signal detection problems have been introduced for the first time by Kassam (1988) on the basis of the compound Poisson processes. Later, Pchelintsev (2013) and Konev et al. (2014) used the compound Poisson processes for parametric regression models and Konev and Pergamenshchikov (2012, 2015) used these processes for nonparametric signal estimation problems. However, the compound Poisson process can describe only the large impulses influence of small frequencies. It should be noted that in telecommunication systems, noise impulses are without limitations on frequencies, and therefore, the compound Poisson models are too restricted for practical applications. To include all possible impulse noises, we propose to use general non-Gaussian Lévy processes in the observation model (1). In this paper, we consider a nonparametric estimation problem in the adaptive setting, i.e., when the regularity of the signal S is unknown. Moreover, we also assume that the distribution Q of the noise process $(\xi_t)_{0\le t\le 1}$ is unknown. It is only known that this distribution belongs to the distribution family $\mathcal{Q}^*_\varepsilon$ defined in the next section. By these reasons, we use a robust estimation approach proposed for nonparametric problems by Galtchouk and Pergamenshchikov (2006) and Konev and Pergamenshchikov (2012, 2015). We set the robust risk as

$$\mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) = \sup_{Q\in\mathcal{Q}^*_\varepsilon} \mathcal{R}_Q(\widehat{S}_\varepsilon, S), \tag{2}$$

where $\widehat{S}_\varepsilon$ is an estimator (i.e., any measurable function of $(y_t)_{0\le t\le 1}$),

$$\mathcal{R}_Q(\widehat{S}_\varepsilon, S) := \mathbf{E}_{Q,S}\, \|\widehat{S}_\varepsilon - S\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t)\mathrm{d}t. \tag{3}$$

In this paper, we develop a sharp model selection method to estimate the unknown signal S. The interest in such statistical procedures can be explained by the fact that they provide adaptive solutions for a nonparametric estimation through non-asymptotic oracle inequalities which give a non-asymptotic upper bound for quadratic risks including the minimal risk over a chosen family of estimators with a coefficient tending to one. Such inequalities were obtained, for example, by Galtchouk and Pergamenshchikov

(2009a) for non-Gaussian regression models in discrete time and by Konev and Pergamenshchikov (2009a) for general regression semimartingale models in continuous time. It should be noted that model selection methods were proposed, for the first time, by Akaike (1974) and Mallows (1973) for parametric models. Then, by using oracle inequalities approach, these methods had been developed for nonparametric estimation by Barron et al. (1999), for Gaussian regression models and by Fourdrinier and Pergamenshchikov (2007) for non-Gaussian models. It is known that an oracle inequality yields upper bounds for risks via minimal risk corresponding to a chosen estimators family. Unfortunately, the oracle inequalities obtained in these papers cannot be used for 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 provide efficiency properties for model selection procedures, one needs to obtain the sharp oracle inequalities, i.e., in which the coefficient at the principal term on the right-hand side of the inequality is close to one. To obtain such inequalities for general non-Gaussian observations, one needs to use the model selection method based on the weighted least square estimators proposed by Galtchouk and Pergamenshchikov (2009a, b) for the heteroscedastic regression models in discrete time and developed then by Konev and Pergamenshchikov (2009a, b, 2012, 2015) for semimartingale models in continuous time, i.e., when the observation process is given by the following stochastic differential equation

$$\mathrm{d}x_t = S(t)\mathrm{d}t + \mathrm{d}\eta_t, \quad 0 \le t \le n, \quad (n \to \infty), \tag{4}$$

where S is an unknown 1—periodic signal and the unobserved noise $(\eta_t)_{t\ge 0}$ is square integrated semimartingale. Note that, for any $0 < t < 1$, setting $\check{x}_t = n^{-1}\sum_{j=0}^{n-1}(x_{t+j} - x_j)$, we can represent this model as a model with small parameter of the (1)

$$\mathrm{d}\check{x}_t = S(t)\mathrm{d}t + \varepsilon\,\mathrm{d}\check{\eta}_t, \tag{5}$$

where $\varepsilon = n^{-1/2}$ and $\check{\eta}_t = n^{-1/2}\sum_{j=0}^{n-1}(\eta_{t+j} - \eta_j)$. If $(\eta_t)_{t\ge 0}$ is a Lévy process, then $\check{\eta}_t$ is a Lévy process as well. But the main difference between models (1) and (5) is that the jumps in the last one are small, i.e.,

$$\Delta\check{\eta}_t = \check{\eta}_t - \check{\eta}_{t-} = \mathrm{O}(n^{-1/2}) = \mathrm{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0. \tag{6}$$

But there is no such property in model (1). It should be noted that property (6) is crucial in the non-asymptotic analysis for observations on large time intervals, i.e., the methods developed for model (4) cannot be used for problem (1). Moreover, it should be emphasized that the selection model methods proposed by Konev and Pergameshchikov for model (4) provide the adaptive efficient estimation only for the case when the Lévy measure is finite. This condition considerably reduces their applications in practical problems. So, the main goal of this paper is to develop a new model selection method for the adaptive efficient signal estimation problem in model (1) with general Lévy noises without limitations on the jumps. First, we construct some model selection procedures and we show sharp non-asymptotic oracle inequalities for risks (2) and (3). To do this in Proposition 2, we develop a special analytical tool to study the non-asymptotic behavior of jumps in model (1) with infinite (or finite) Lévy measure. Moreover, to study the efficiency, we develop the Van Trees method for general Lévy processes and we obtain in Proposition 4 a new lower bound for quadratic risks. Then, by making use of this lower bound, we find the Pinsker constant. As to the upper bound, similarly to Konev and Pergamenshchikov (2009b), we use the obtained sharp oracle inequality for weight least square estimators containing the efficient Pinsker procedure. Therefore, through oracle inequalities, we estimate from above the risk for the constructed model selection procedure by the efficient risk up to some coefficient which goes to one. As a result, we provide the robust efficiency property for the constructed procedure in the adaptive setting. As an application of the developed model selection method, in this paper, we consider the signal number detection problem in the observations model (1). In many areas of science and technology, the problem arises how to select the number of freedom degrees for a statistical model that most adequately describes phenomena under studies (see, for example, Akaike 1974). An important class of such problems is the detection problem of signal number with unknown parameters observed in multi-path information transmission with noises. For example, in the signal multi-path information transmission, there is a detection problem for the number of rays in the multi-path channel. This problem is often reduced to the detection of the number of signals. As a result, the effective detection algorithms can significantly improve the noise immunity in data transmission over a multi-path channel (see, for example, the papers of Flaksman 2002; Manelis 2007; El-Behery and Macphie 1978; Trifonov and Kharin 2013; Trifonov and Shinakov 1986; Trifonov et al. 2015; Trifonov and Kharin 2015a, b). Such problems for signals with unknown amplitudes are discussed in Trifonov and Kharin (2013). The signal amplitude is an energy parameter because it affects the signal energy. At the same time, quite often, such as in radars studied by El-Behery and Macphie (1978), it is necessary to detect the number of signals, which besides unknown amplitudes contain non-energy parameters such as frequencies and initial phases. Trifonov et al. (2015) considered this problem with unknown initial phases, in Trifonov and Kharin (2015a) with unknown amplitudes and phases and in Trifonov and Kharin (2015b) the detection signal number problem is studied for orthogonal signals with arbitrary non-energy parameters. In all these papers, the detection problems are considered only for observations with the Gaussian white nose. In this paper, we consider this problem for non-Gaussian impulse noises.

The rest of the paper is organized as follows. In Sect. 2, we give the main conditions which will be assumed for model (1). In Sect. 3, we transform the observation model to delete large jumps and we develop an analytical tool, which provides to study non-asymptotic concentration properties for the squares of stochastic integrals with respect to non-Gaussian Lévy processes. In Sect. 4, we construct a model selection procedure. In Sect. 5, we state our main results on sharp oracle inequalities and the adaptive efficiency. In Sect. 6, we obtain the van Trees inequality for general Lévy processes. In Sect. 7, we study a signal number detection problem through the developed model selection method. In Sect. 8, we give simulations results. Section 9 contains the main proofs. In Appendix (Supplementary Material), we bring all auxiliary results.

2 Main conditions

In this section, we assume that the noise process $(\xi_t)_{0\le t\le 1}$ is defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \widetilde{\mu})_t, \tag{7}$$

where ϱ_1 and ϱ_2 are some unknown constants, $(w_t)_{0\le t\le 1}$ is a standard Brownian motion, "*" denotes the stochastic integral with respect to the compensated jump measure (see, for example, in Jacod and Shiryaev 2002 or Cont and Tankov 2004 for details), $\mu(\mathrm{d}s\,\mathrm{d}x)$ is a jump measure with deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\varPi(\mathrm{d}x)$, $\varPi(\cdot)$ is the unknown Lévy measure, i.e., some nonnegative measure on $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ for which

$$\int_{\mathbb{R}_*} (z^2 \wedge 1)\, \varPi(\mathrm{d}z) < \infty,$$

where $a \wedge b = \min(a, b)$. In addition, we impose the following condition.
$(\mathbf{A}_1)$ *The Lévy measure $\varPi(\cdot)$ is assumed to satisfy the following moment conditions*

$$\varPi(x^2) = 1 \quad \text{and} \quad \varPi(x^4) < \infty, \tag{8}$$

where $\varPi(|x|^m) = \int_{\mathbb{R}_*} |z|^m\, \varPi(\mathrm{d}z)$.

Note that the measure $\varPi(\mathbb{R}_*)$ could be equal to $+\infty$. In the sequel, we will denote by Q the distribution of the process $(\xi_t)_{0\le t\le 1}$ and by $\mathcal{Q}^*_\varepsilon$ the family of such distributions in the Skorokhod space $\mathbf{D}[0, 1]$ for which

$$0 < \varsigma_* \le \varrho_1^2 \quad \text{and} \quad \varkappa_Q = \varrho_1^2 + \varrho_2^2 \le \varsigma^*, \tag{9}$$

where $0 < \varsigma_* \le \varsigma^*$ are unknown parameters which can be represented as functions of ε satisfying the following additional condition.
$(\mathbf{A}_2)$ *The bounds ς_* and ς^* are functions of ε, i.e., $\varsigma_* = \varsigma_*(\varepsilon)$ and $\varsigma^* = \varsigma^*(\varepsilon)$, such that for any $\mathbf{b} > 0$*

$$\liminf_{\varepsilon\to 0}\, \varepsilon^{-\mathbf{b}}\, \varsigma_*(\varepsilon) > 0 \quad \text{and} \quad \lim_{\varepsilon\to 0}\, \varepsilon^{\mathbf{b}}\, \varsigma^*(\varepsilon) = 0. \tag{10}$$

It is clear that in condition $(\mathbf{A}_2)$ the bounds $\varsigma_* \le \varsigma^*$ may be any positive constants.

3 Transformation of the observations

First of all, we need to eliminate the large jumps in the observations (1), i.e., we transform this model as

$$\check{y}_t = y_t - \sum_{0\le s\le t} \Delta y_s\, \mathbf{1}_{\{|\Delta y_s| > \overline{a}\}}. \tag{11}$$

The parameter $\overline{a} = \overline{a}_\varepsilon > 0$ will be chosen later. So, we obtain that

$$\mathrm{d}\check{y}_t = S(t)\mathrm{d}t + \varepsilon \mathrm{d}\check{\xi}_t - \varepsilon\, \varrho_2\, \varPi(\overline{h}_\varepsilon)\, \mathrm{d}t, \tag{12}$$

where $\check{\xi}_t = \varrho_1 w_t + \varrho_2 \check{z}_t$ and $\check{z}_t = h_\varepsilon * (\mu - \widetilde{\mu})_t$. The functions $h_\varepsilon(x) = x\mathbf{1}_{\{|x|\le\check{\upsilon}_\varepsilon\}}$ and $\overline{h}_\varepsilon(x) = x\mathbf{1}_{\{|x|>\check{\upsilon}_\varepsilon\}}$ with the truncation threshold $\check{\upsilon}_\varepsilon = \overline{a}/|\varrho_2|\varepsilon$.

Remark 1 It should be noted that the sum in the transformation (11) is finite since the cadlag process has only finite number of jumps more than some positive threshold in absolute value.

Let $(\phi_j)_{j\ge1}$ be an orthonormal basis in $\mathbf{L}_2[0,1]$ with $\phi_1 \equiv 1$. We assume that this basis is uniformly bounded, i.e., for some constant $\phi^* > 0$, which may be dependent on $\varepsilon > 0$,

$$\sup_{0\le j\le n}\ \sup_{0\le t\le 1} |\phi_j(t)| \le \phi^* < \infty, \tag{13}$$

where $n = n_\varepsilon = [1/\varepsilon^2]$ and $[x]$ denotes integer part of x. For example, we can take the trigonometric basis defined as $\mathrm{Tr}_1 \equiv 1$ and for $j \ge 2$

$$\mathrm{Tr}_j(x) = \sqrt{2}\begin{cases} \cos(2\pi[j/2]x) & \text{for even} \quad j\,; \\ \sin(2\pi[j/2]x) & \text{for odd} \quad j. \end{cases} \tag{14}$$

Moreover, note that for any $[0,1] \to \mathbb{R}$ function f from $\mathbf{L}_2[0,1]$ and for any $0 \le t \le 1$ the integrals

$$I_t(f) = \int_0^t f(s)\mathrm{d}\xi_s \quad \text{and} \quad \check{I}_t(f) = \int_0^t f(s)\mathrm{d}\check{\xi}_s \tag{15}$$

are well defined with $\mathbf{E}\, I_t(f) = 0$, $\mathbf{E}\, \check{I}_t(f) = 0$,

$$\mathbf{E}\, I_t^2(f) = \varkappa_Q \, \|f\|_t^2 \quad \text{and} \quad \mathbf{E}\, \check{I}_t^2(f) = \check{\varkappa}_Q \, \|f\|_t^2, \tag{16}$$

where $\|f\|_t^2 = \int_0^t f^2(s)\mathrm{d}s$ and $\check{\varkappa}_Q = \varrho_1^2 + \varrho_2^2 \Pi(h_\varepsilon^2)$. In the sequel, we denote by

$$(f,g)_t = \int_0^t f(s)g(s)\,\mathrm{d}s \quad \text{and} \quad (f,g) = \int_0^1 f(s)g(s)\,\mathrm{d}s.$$

To estimate the function S, we use the following Fourier series

$$S(t) = \sum_{j\ge1} \theta_j\, \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j). \tag{17}$$

These coefficients can be estimated by the following way. First we estimate as

$$\widehat{\theta}_{1,\varepsilon} = \int_0^1 \phi_1(t)\mathrm{d}\, y_t = \theta_1 + \varepsilon\xi_1$$

and for $j \ge 2$

$$\widehat{\theta}_{j,\varepsilon} = \int_0^1 \phi_j(t)\mathrm{d}\, \check{y}_t. \tag{18}$$

Taking into account here that for $j \geq 2$ the integral $\int_0^1 \phi_j(t)\mathrm{d}t = 0$, we obtain from (12) that these Fourier coefficients can be represented as

$$\widehat{\theta}_{j,\varepsilon} = \theta_j + \varepsilon\,\overline{\xi}_j \quad \text{and} \quad \overline{\xi}_j = \check{I}_1(\phi_j).$$

Setting $\overline{\xi}_1 = \xi_1$, we obtain that for any $j \geq 1$

$$\widehat{\theta}_{j,\varepsilon} = \theta_j + \varepsilon\,\overline{\xi}_j. \tag{19}$$

Now, according to the model selection approach developed by Konev and Pergamenshchikov (2009a, b) we need to study for any $u \in \mathbb{R}^n$ the following functions

$$B_{1,\varepsilon}(u) = \sum_{j=1}^{n} u_j\,(\mathbf{E}_Q\,\overline{\xi}_j^2 - \check{\varkappa}_Q) \quad \text{and} \quad B_{2,\varepsilon}(u) = \sum_{j=1}^{n} u_j\,\widetilde{\xi}_j, \tag{20}$$

where $\widetilde{\xi}_j = \overline{\xi}_j^2 - \mathbf{E}_Q\,\overline{\xi}_j^2$.

Proposition 1 *Assume that condition* $(\mathbf{A}_1)$ *holds. Then*

$$\sup_{u\in[0,1]^n} \left|B_{1,\varepsilon}(u)\right| \leq \varkappa_Q. \tag{21}$$

Proof Note that $|\mathbf{E}_Q\,\overline{\xi}_1^2 - \check{\varkappa}_Q| = |\mathbf{E}_Q\,\xi_1^2 - \check{\varkappa}_Q| = \varkappa_Q - \check{\varkappa}_Q \leq \varkappa_Q$ and $\mathbf{E}_Q\,\overline{\xi}_j^2 = \check{\varkappa}_Q$ for $j \geq 2$. So, from this we immediately obtain the upper bound (21). □

Now, for any $u \in \mathbb{R}^n$ we set

$$|u|^2 = \sum_{j=1}^{n} u_j^2 \quad \text{and} \quad \#(u) = \sum_{j=1}^{n} \mathbf{1}_{\{u_j \neq 0\}}. \tag{22}$$

Now we study the concentration term $B_{2,\varepsilon}(u)$.

Proposition 2 *Assume that condition* $(\mathbf{A}_1)$ *holds. Then for any fixed truncation parameter* $\overline{a} > 0$ *and for any vector* $u \in \mathbb{R}^n$ *with* $|u| \leq 1$

$$\mathbf{E}_Q\,B_{2,\varepsilon}^2(u) \leq U_Q + 4\,\varkappa_Q \left(\frac{\overline{a}}{\varepsilon}\right)^2 \#(u)\,(\phi^*)^4, \tag{23}$$

where $U_Q = 2(25 + 16\mathbf{E}\,z_1^4)\varkappa_Q^2$.

Remark 2 It should be noted that the last term in the concentration inequality (23) is related to the influence of jumps in the observations (1). We will use the upper bounds (21) and (23) to obtain non-asymptotic sharp oracle inequalities.

4 Model selection

We estimate the function $S(x)$ for $x \in [0,1]$ by the weighted least squares estimator

$$\widehat{S}_\lambda(x) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,\varepsilon}\phi_j(x), \tag{24}$$

where $n = [1/\varepsilon^2]$, the weights $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$, $\widehat{\theta}_{j,\varepsilon}$ is defined in (18) and ϕ_j in (14). Now we set

$$\iota = \mathrm{card}(\Lambda) \quad \text{and} \quad |\Lambda|_* = \max_{\lambda\in\Lambda} \sum_{j=1}^{n} \mathbf{1}_{\{\lambda_j>0\}}, \tag{25}$$

where $\mathrm{card}(\Lambda)$ is the number of the vectors in Λ. In the sequel, we assume that ι is a function of $\varepsilon > 0$, i.e., $\iota = \iota(\varepsilon)$, such that for any $\mathbf{b} > 0$

$$\lim_{\varepsilon\to 0} \varepsilon^{\mathbf{b}} \iota(\varepsilon) = 0. \tag{26}$$

Now we chose the truncation parameter in (11) $\overline{a}$ as

$$\overline{a} = \overline{a}_\varepsilon = \frac{\varepsilon}{\sqrt{|\Lambda|_*}}. \tag{27}$$

To choose a weight sequence λ in the set Λ, we use the empirical quadratic risk, defined as

$$\mathrm{Err}_\varepsilon(\lambda) = \| \widehat{S}_\lambda - S \|^2,$$

which in our case is equal to

$$\mathrm{Err}_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,\varepsilon}\theta_j + \sum_{j=1}^{\infty} \theta_j^2. \tag{28}$$

Since the Fourier coefficients $(\theta_j)_{j\ge 1}$ are unknown, we replace the terms $\widehat{\theta}_{j,\varepsilon}\theta_j$ by

$$\widetilde{\theta}_{j,\varepsilon} = \widehat{\theta}^2_{j,\varepsilon} - \varepsilon^2 \widehat{\varkappa}_\varepsilon, \tag{29}$$

where $\widehat{\varkappa}_\varepsilon$ is an estimate for the variance parameter $\check{\varkappa}_Q$ from (16). If it is known, we set $\widehat{\varkappa}_\varepsilon = \check{\varkappa}_Q$; otherwise, this estimator will be prescribed later.

Remark 3 To understand estimate (29), note that the natural way is to replace in the production $\widehat{\theta}_j\theta_j$ the unknown coefficient θ_j with its estimator $\widehat{\theta}_j$, so we obtain $\widehat{\theta}^2_j$. But this is not a good estimator for the production since in view of (19) we obtain $\mathbf{E}_Q\,\widehat{\theta}_j\theta_j = \theta_j^2$, but $\mathbf{E}_Q\,\widehat{\theta}^2_j = \theta_j^2 + \varepsilon^2\mathbf{E}_Q\overline{\xi}^2_j$. Therefore, to obtain unbiased estimator for the production $\widehat{\theta}_j\theta_j$ for $j \ge 2$ one needs to subtract the variance $\varepsilon^2\check{\varkappa}_Q$ if $\check{\varkappa}_Q$ is known and its estimate if it is not known. This gives the (29). It should be noted also that we do not take into account the first term, i.e., the case $j = 1$. But only one term has not sufficient influence in the total sum, i.e., it is negligible in the empiric risk (28).

Finally, to choose the weights we will minimize the following cost function

$$J_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,\varepsilon} + \delta\,\widehat{P}_\varepsilon(\lambda), \tag{30}$$

where $\delta > 0$ is some threshold which will be specified later and the penalty term

$$\widehat{P}_\varepsilon(\lambda) = \varepsilon^2 \widehat{\varkappa}_\varepsilon |\lambda|^2 \quad \text{and} \quad |\lambda|^2 = \sum_{j=1}^{n} \lambda_j^2. \tag{31}$$

Note that if the $\check{\varkappa}_Q$ is known then the penalty term is defined as

$$P_\varepsilon(\lambda) = \varepsilon^2\, \check{\varkappa}_Q |\lambda|^2. \tag{32}$$

As to the penalty term, we can show the following upper bound.

Lemma 1 *Assume that condition* $(\mathbf{A}_1)$ *holds. Then for any* $0 < \varepsilon < 1$ *and* $\lambda \in \Lambda$,

$$P_\varepsilon(\lambda) \le \mathcal{R}(\widehat{S}_\lambda, S) + \varepsilon^2\, \varkappa_Q.$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}} \quad \text{and} \quad \widehat{\lambda} = \text{argmin}_{\lambda\in\Lambda} J_\varepsilon(\lambda). \tag{33}$$

We recall that the set Λ is finite so $\hat{\lambda}$ exists. In the case when $\hat{\lambda}$ is not unique, we take one of them.

Now we estimate the variance parameter $\check{\varkappa}_Q$ defined in (16). To this end for any $0 < \varepsilon \le 1/\sqrt{3}$, we set

$$\widehat{\varkappa}_\varepsilon = \sum_{j=[1/\varepsilon]+1}^{n} \widehat{\tau}^2_{j,\varepsilon}, \quad n = [1/\varepsilon^2], \tag{34}$$

where $\widehat{\tau}_{j,\varepsilon}$ are the estimators for the Fourier coefficients τ_j with respect to the trigonometric basis (14), i.e.,

$$\widehat{\tau}_{j,\varepsilon} = \int_0^1 \text{Tr}_j(t)\text{d}\check{y}_t \quad \text{and} \quad \tau_j = \int_0^1 S(t)\text{Tr}_j(t)\text{d}t. \tag{35}$$

We study this estimator.

Proposition 3 *Assume that in model (1) the unknown function* $S(\cdot)$ *is continuously differentiable. Then, for any* $0 < \varepsilon \le 1/\sqrt{3}$

$$\mathbf{E}_Q\, |\widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q| \le \varepsilon \Upsilon_Q(S), \tag{36}$$

where $\Upsilon_Q(S) = 4(\|\dot{S}\| + 1)^2 \left(1 + 2\sqrt{\varkappa_Q} + \varkappa_Q + \sqrt{U_Q}\right)$ *and* $\dot{S}$ *is the derivative of the function* S.

Remark 4 It should be noted that to estimate the parameter $\check{\varkappa}_Q$ we use equality (19) for the Fourier coefficients $(\tau_j)_{j\geq 1}$ with respect to the trigonometric basis (14). Moreover, 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 $\sum_{j\geq m} \tau_j^2$ can be estimated from above explicitly. So, taking this into account and properties (23) we obtain the upper bound (36).

Now, we specify the weight coefficients $(\lambda(j))_{1\leq j\leq n}$. Consider a numerical grid of the form

$$\mathcal{A} = \{1, \ldots, k^*\} \times \{r_1, \ldots, r_{\mathbf{m}}\}, \tag{37}$$

where $r_i = i\,\varpi$ and $\mathbf{m} = [1/\varpi^2]$. We assume that both the parameters $k^* \geq 1$ and $0 < \varpi < 1$ are functions of ε, i.e., $k^* = k^*_\varepsilon$ and $\varpi = \varpi_\varepsilon$ such that

$$\lim_{\varepsilon\to 0} \left(\frac{1}{k^*_\varepsilon} + \frac{k^*_\varepsilon}{|\ln \varepsilon|}\right) = 0 \quad \text{and} \quad \lim_{\varepsilon\to 0} \left(\varpi_\varepsilon + \frac{\varepsilon^{\mathbf{b}}}{\varpi_\varepsilon}\right) = 0 \tag{38}$$

for any $\mathbf{b} > 0$. One can take, for example, for $0 < \varepsilon < 1$

$$\varpi_\varepsilon = |\ln \varepsilon|^{-1} \quad \text{and} \quad k^*_\varepsilon = k^*_0 + \sqrt{|\ln \varepsilon|}, \tag{39}$$

where $k^*_0 \geq 0$ is some fixed constant. For each $\alpha = (\beta, r) \in \mathcal{A}$, we introduce the weights $\lambda_\alpha = (\lambda_\alpha(j))_{1\leq j\leq n}$ from $\mathbb{R}^n$ as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\leq j<j_*\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{j_*\leq j\leq \omega_\alpha\}}, \tag{40}$$

where $j_* = j_*(\alpha) = \left[\omega_\alpha/|\ln \varepsilon|\right]$, $\omega_\alpha = \mathrm{d}_\beta\,(r\,\upsilon_\varepsilon)^{1/(2\beta+1)}$,

$$\mathrm{d}_\beta = \left(\frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\right)^{1/(2\beta+1)}, \qquad \upsilon_\varepsilon = \frac{1}{\varepsilon^2\,\varsigma^*}, \tag{41}$$

and the threshold ς^* is introduced in (9). Now we define the set Λ as

$$\Lambda = \{\lambda_\alpha,\ \alpha \in \mathcal{A}\}. \tag{42}$$

Note that in this case $\iota = k^*\mathbf{m}$ and conditions (38) imply directly property (26). Moreover, from (40) we find that for any $\alpha \in \mathcal{A}$

$$\sum_{j=1}^n \lambda_\alpha(j) \leq \omega_\alpha \leq d_*\, r_{\mathbf{m}}^{1/3}\, \upsilon_\varepsilon^{1/3} \quad \text{and} \quad d_* = \sup_{\beta\geq 1} \mathrm{d}_\beta.$$

Therefore, conditions (38) imply that for any $\mathbf{b} > 0$

$$\lim_{\varepsilon \to 0} \varepsilon^{2/3+\mathbf{b}} |\Lambda|_* = 0. \tag{43}$$

Remark 5 The parameters β and r are defined by the regularity of the unknown function S (see, for details, Remark 8). It should be emphasized that the weight coefficients defined by set (42) are used by Konev and Pergamenshchikov (2012, 2015) for continuous-time regression models to show the asymptotic efficiency.

5 Main results

5.1 Oracle inequalities

First we set the following constant which will be used to describe the rest term in the oracle inequalities. We set

$$\Psi_{Q,\varepsilon} = (1 + (\phi^*)^4) \left(1 + \varkappa_Q^2 + \frac{1}{\check{\varkappa}_Q} \right) \iota. \tag{44}$$

We start with the sharp oracle inequalities.

Theorem 1 *Assume that condition* $(\mathbf{A}_1)$ *holds. Then there exists a constant* $\mathbf{l}_* > 0$ *such that for any* $\varepsilon > 0$ *and* $0 < \delta < 1/6$, *the model selection procedure* (33) *with the truncation parameter* (27) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \varepsilon^2 \, \mathbf{l}_* \, \frac{\Psi_{Q,\varepsilon} + |\Lambda|_* \, \mathbf{E}_S |\widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q|}{\delta}. \tag{45}$$

If the parameter $\check{\varkappa}_Q$ is known, we can simplify this inequality.

Corollary 1 *If the variance parameter* $\check{\varkappa}_Q$ *is known and condition* $(\mathbf{A}_1)$ *holds, then there exists a constant* $\mathbf{l}_* > 0$ *such that for any* $\varepsilon > 0$ *and for any* $0 < \delta < 1/6$, *the model selection procedure* (33) *with the truncation parameter* (27) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \varepsilon^2 \, \mathbf{l}_* \, \frac{\Psi_{Q,\varepsilon}}{\delta}. \tag{46}$$

Remark 6 It should be noted that in the classical "signal+white noise" model, i.e., when in the process (7) the parameter $\varrho_1 = 1$ and the Lévy measure $\Pi = 0$, we obtain $\check{\varkappa}_Q = 1$. Therefore, we can use inequality (46).

Using Proposition 3, we can obtain the following inequality.

Theorem 2 *Assume that condition* $(\mathbf{A}_1)$ *holds and the unknown signal* $S(\cdot)$ *is continuously differentiable* $[0, 1] \to \mathbb{R}$ *function. Then there exists a constant* $\mathbf{l}_* > 0$ *such* *that for any* $0 < \delta < 1/6$ *and for any* $0 < \varepsilon \le 1/\sqrt{3}$, *for which* $|\Lambda|_* \le 1/\varepsilon$, *the estimation procedure* (33) *with the truncation parameter* (27) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \varepsilon^2 \, \mathbf{l}_* \, \frac{\Psi_{Q,\varepsilon}(\|\dot{S}\|+1)^2}{\delta}. \tag{47}$$

Now we study the robust risks defined in (2) for procedure (33). To do this, we will use the following condition on the basis functions $(\phi_j)_{j\ge1}$.
$(\mathbf{A}_3)$ *The upper bound for the basis functions in (13) is a function of* $\varepsilon > 0$, *i.e.,* $\phi_* = \phi_*(\varepsilon)$, *such that for any* $\mathbf{b} > 0$

$$\lim_{n\to\infty} \varepsilon^{\mathbf{b}} \, \phi_*(\varepsilon) = 0. \tag{48}$$

Theorem 3 *Assume that conditions* $(\mathbf{A}_1)$–$(\mathbf{A}_3)$ *hold and the unknown function* $S(\cdot)$ *is continuously differentiable. Then for any* $0 < \delta < 1/6$ *and for any* $0 < \varepsilon \le 1/\sqrt{3}$ *for which* $|\Lambda|_* \le 1/\varepsilon$, *the robust risks for the estimation procedure* (33) *with the truncation parameter* (27) *satisfies the following oracle inequality*

$$\mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*_\varepsilon(\widehat{S}_\lambda, S) + \varepsilon^2 \, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta}, \tag{49}$$

where the term $\mathbf{U}^*_\varepsilon(S) > 0$ *is such that under conditions* (26) *and* (48) *for any* $r > 0$ *and* $\mathbf{b} > 0$

$$\lim_{\varepsilon\to0} \varepsilon^{\mathbf{b}} \sup_{\|\dot{S}\|\le r} \mathbf{U}^*_\varepsilon(S) = 0. \tag{50}$$

Now taking into account property (43), we can deduce the following theorem for procedure (33) with the weight coefficients (42).

Theorem 4 *Assume that conditions* $(\mathbf{A}_1)$–$(\mathbf{A}_3)$ *hold and the unknown function* $S(\cdot)$ *is continuously differentiable. Then the model selection procedure* (33) *constructed through the weight coefficients* (42) *satisfies the oracle inequality* (49) *with property* (50).

***Remark* 7** Note that the similar sharp oracle inequalities were obtained by Galtchouk and Pergamenshchikov (2009a) and Konev and Pergamenshchikov (2012) for the model selection procedures based on the trigonometric basis functions (14). In this paper, we obtain these inequalities for the model selection procedures based on any arbitrary orthogonal basic function in $\mathbf{L}_2[0, 1]$. We use the trigonometric functions only to estimate the noise parameter $\check{\varkappa}_Q$.

5.2 Adaptive robust efficiency

Now we study the asymptotically efficiency properties for procedure (33), (40) with respect to the robust risks (2) defined by the distribution family (9)–(10). To this end, we assume that the unknown function (17) belongs to the following ellipsoid in l_2

$$W_{\mathbf{r}}^{k}=\{S\in\mathbf{L}_2[0,1]\,:\,\sum_{j=1}^{\infty}a_j\,\theta_j^2\,\leq\mathbf{r}\},\tag{51}$$

where $a_j=\sum_{i=0}^{k}(2\pi[j/2])^{2i}$.

It is easy to see that in the case when the functions $(\phi_j)_{j\geq 1}$ are trigonometric (14), then this set coincides with the Sobolev ball

$$W_{\mathbf{r}}^{k}=\{f\in\mathbf{C}_{per}^{k}[0,1]\,:\,\sum_{j=0}^{k}\|f^{(j)}\|^2\leq\mathbf{r}\},\tag{52}$$

where $\mathbf{r}>0$ and $k\geq 1$ are some parameters, $\mathbf{C}_{per}^{k}[0,1]$ is the set of k times continuously differentiable $f\,:\,[0,1]\to\mathbb{R}$ functions such that $f^{(i)}(0)=f^{(i)}(1)$ for all $0\leq i\leq k$. Similarly to Konev and Pergamenshchikov (2012, 2015), we will show here that the asymptotic sharp lower bound for the robust risk (2) is given by

$$l_*(\mathbf{r})=\,((2k+1)\mathbf{r})^{1/(2k+1)}\left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)}.\tag{53}$$

Note that this is the well-known Pinsker constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, Pinsker 1981). Let $\mathcal{S}_\varepsilon$ be the set of all estimators $\widehat{S}_\varepsilon$ measurable with respect to the sigma algebra $\sigma\{y_t,\,0\leq t\leq 1\}$ generated by the process (1).

Theorem 5 *Assume that conditions* $(\mathbf{A}_1)$–$(\mathbf{A}_2)$ *hold. Then the robust risk* (2) *admits the following lower bound*

$$\liminf_{\varepsilon\to 0}\,\upsilon_\varepsilon^{2k/(2k+1)}\,\inf_{\widehat{S}_\varepsilon\in\mathcal{S}_\varepsilon}\,\sup_{S\in W_r^k}\,\mathcal{R}_\varepsilon^*(\widehat{S}_\varepsilon,S)\geq\,l_*(\mathbf{r}),\tag{54}$$

where the rate υ_ε *is given in* (41), *i.e.,* $\upsilon_\varepsilon=\left(\varepsilon^2\,\varsigma^*\right)^{-1}$.

To study the upper bound for the model selection procedure (33), we need to assume the following condition on the penalization parameter δ in (30).

$(\mathbf{A}_4)$ *The parameter* δ *in* (30) *is a function of* ε, *i.e.,* $\delta=\delta_\varepsilon$ *is such that*

$$\lim_{\varepsilon\to 0}\,\delta_\varepsilon=0\quad\text{and}\quad\lim_{\varepsilon\to 0}\,\varepsilon^{-\mathbf{b}}\,\delta_\varepsilon=+\infty\tag{55}$$

for any $\mathbf{b}>0$.

For example, we can take $\delta_\varepsilon=(6+|\ln\varepsilon|)^{-1}$.

Theorem 6 *Assume that conditions* $(\mathbf{A}_1)$–$(\mathbf{A}_4)$ *hold. Then the model selection procedure* (33) *constructed through the weight coefficients* (42) *admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to 0}\,\upsilon_\varepsilon^{2k/(2k+1)}\,\sup_{S\in W_r^k}\,\mathcal{R}_\varepsilon^*(\widehat{S}_*,S)\leq l_*(\mathbf{r}).\tag{56}$$

Moreover, Theorems 5 and 6 imply the following result.

Corollary 2 *Under the conditions Theorem 6*

$$\lim_{\varepsilon\to 0} \upsilon_\varepsilon^{2k/(2k+1)} \inf_{\widehat{S}_\varepsilon \in \mathcal{S}_\varepsilon} \sup_{S\in W_r^k} \mathcal{R}_\varepsilon^*(\widehat{S}_\varepsilon, S) = l_*(\mathbf{r}). \tag{57}$$

***Remark* 8** It should be noted (see, for example, Pinsker 1981). that if the parameters k and r of the Sobolev ball (51) are known, then to obtain the efficient estimation it suffice to chose the weight least square estimator (24) with the weights (40) and $\alpha = (k, r)$. In the adaptive estimation case, i.e., when these parameters are unknown we propose to use the selection model procedure for the family $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ which contains the efficient estimator. Then, the efficiency property is provided through the sharp oracle inequalities. Moreover, note also that the optimal (minimax) risk convergence rate for the Sobolev ball W_r^k is $\varepsilon^{-4k/(2k+1)}$. We see here that the efficient robust rate is $\upsilon_\varepsilon^{2k/(2k+1)}$, i.e., if the distribution upper bound $\varsigma^* \to 0$ as $n \to \infty$ we obtain the more rapid rate with respect to $\varepsilon^{-4k/(2k+1)}$, and if $\varsigma^* \to \infty$ as $\varepsilon \to 0$ 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.

6 The van Trees inequality for Lévy processes

In this section, we consider the following continuous-time parametric regression model

$$\mathrm{d}y_t = S(t,\theta)\mathrm{d}t + \mathrm{d}\xi_t, \quad 0 \le t \le 1, \tag{58}$$

where $S(t,\theta) = \sum_{i=1}^d \theta_i\,\phi_i(t)$ with the unknown parameters $\theta = (\theta_1, \ldots, \theta_d)'$ and the process $(\xi_t)_{0\le t\le 1}$ is defined in (7). Note now that according to Proposition A1 from Appendix the distribution $\mathbf{P}_\theta$ of the process (58) is absolutely continuous with respect to the $\mathbf{P}_\xi$ on $\mathbf{D}[0,1]$ and the corresponding Radon–Nikodym derivative is

$$f(x,\theta) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mathbf{P}_\xi}(x) = \exp\left\{\int_0^1 \frac{S(t,\theta)}{\varrho_1^2}\,\mathrm{d}x_t^c - \int_0^1 \frac{S^2(t,\theta)}{2\varrho_1^2}\,\mathrm{d}t\right\}, \tag{59}$$

where $(x_t^c)_{0\le t\le T}$ is the continuous part of the process $(x_t)_{0\le t\le T}$ in $\mathbf{D}[0,T]$, i.e.,

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

and for any $t > 0$ and any measurable Γ from $\mathbb{R}_*$

$$\mu_x([0,t],\Gamma) = \sum_{0\le s\le t} \mathbf{1}_{\{\Delta x_s \in \varrho_2 \Gamma\}}.$$

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

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

where φ_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 $1 \le j \le d$,

$$\lim_{|\theta_j| \to \infty} g(\theta)\, \varphi_j(\theta_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |g'_j(\theta)|\, \Phi(\theta)\, \mathrm{d}\theta < \infty, \tag{60}$$

where

$$g'_j(\theta) = \frac{\partial g(\theta)}{\partial \theta_j}.$$

For any $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathbb{R}^d)$ measurable integrable function $H = H(x, \theta)$, we denote

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

where $\mathcal{X} = \mathbf{D}[0, 1]$.

Proposition 4 *For any $\mathcal{F}^y = \sigma\{y_t\, 0 \le t \le 1\}$ measurable square integrable function $\widehat{g}$ and for any $1 \le j \le d$, the following inequality holds*

$$\widetilde{\mathbf{E}}(\widehat{g} - g(\theta))^2 \ge \frac{\Lambda_j^2}{\|\phi_j\|^2 \varrho_1^{-2} + I_j}, \tag{61}$$

where

$$\lambda_j = \int_{\mathbb{R}^d} g'_j(\theta)\, \Phi(\theta)\, \mathrm{d}\theta \quad \textit{and} \quad I_j = \int_{\mathbb{R}} \frac{\dot{\varphi}_j^2(z)}{\varphi_j(z)}\, \mathrm{d}z.$$

***Remark* 9** Note that the lower bound (61) is an extension for the van Trees inequality used for the "signal+white noise" model (see, for example, inequality (A.5) in Konev and Pergamenshchikov 2009b).

7 Signal number detection

In this section, we consider the estimation problem for the signal number in the multi-path connection channel. In the framework of the statistical radio-physics models, we study the telecommunication system in which we observe the summarized signal in the multi-path channel with noise on the time interval [0, 1]:

$$y_t = \sum_{j=1}^{q} \theta_j \phi_j(t) + \nu_t, \quad 0 \le t \le 1,$$

where $(\nu_t)_{t\ge 0}$ is the Gaussian white noise. The energetic parameters $(\theta_j)_{j\ge 1}$ and the number of signals q are unknown, and the signals $(\phi_j)_{j\ge 1}$ are known orthonormal functions, i.e., $\int_0^1 \phi_i(t)\,\phi_j(t)\,\mathrm{d}t = \mathbf{1}_{\{i\ne j\}}$. The problem is to estimate q when signal/noise ratio goes to infinity. To describe this problem in a mathematical framework, one has to use the following stochastic differential equation

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \phi_j(t) \right) \mathrm{d}t + \varepsilon \mathrm{d}w_t, \tag{62}$$

where $(w_t)_{t\ge 0}$ is the standard Brownian motion and the parameter $\varepsilon > 0$ is the noise intensity. We study this model when the signal/noise ratio goes to infinity, i.e., $\varepsilon \to 0$. The logarithm of the likelihood ratio for model (62) can be represented as

$$\ln L_\varepsilon = \frac{1}{\varepsilon^2} \sum_{j=1}^{q} \theta_j \int_0^1 \phi_j(t)\mathrm{d}y_t - \frac{1}{2\varepsilon^2} \sum_{j=1}^{q} \theta_j^2.$$

If we try to construct the maximum likelihood estimators for $(\theta_j)_{1\le j\le q}$ and q, then we obtain that

$$\max_{1\le q\le q_*} \max_{\theta_j} \ln L_\varepsilon = \frac{1}{2\varepsilon^2} \sum_{j=1}^{q_*} \left(\int_0^1 \phi_j(t)\mathrm{d}y_t \right)^2.$$

Therefore, the maximum likelihood estimation for $\widehat{q} = q^*$. So, if $q^* = \infty$ we obtain that $\widehat{q} = \infty$. Thus, this estimator gives nothing, i.e., it does not work. For these reasons, we propose to study the estimation problem for q for the process (62) in a nonparametric setting and to apply the model selection procedure (33). To this end, we consider model (1) with the unknown function S defined as

$$S(t) = \sum_{j=1}^{q} \theta_j \, \phi_j(t). \tag{63}$$

For this problem, we use the LSE family $(\widehat{S}_d)_{1\le d\le m}$ defined as

$$\widehat{S}_d(x) = \sum_{j=1}^{d} \widehat{\theta}_{j,\varepsilon} \phi_j(x). \tag{64}$$

This estimate can be obtained from (12) with the weights $\lambda_d(j) = \chi\{j \le d\}$. The number of estimators ι satisfies condition (26). As a risk for the signal number, we use

$$\mathbf{D}_\varepsilon(d, q) = \mathcal{R}^*_\varepsilon(\widehat{S}_d, S), \tag{65}$$

where the risk $\mathcal{R}^*_\varepsilon(\widehat{S}, S)$ is defined in (2) and d is an integer number (maybe random) from the set $\{1, \ldots, \iota\}$. In this case, the cost function (30) has the following form.

$$J_\varepsilon(d) = \sum_{j=1}^{d} \widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{d} \widetilde{\theta}_{j,\varepsilon} + \delta\,\widehat{P}_\varepsilon(\lambda). \tag{66}$$

So, for this problem the LSE model selection procedure is defined as

$$\widehat{q}_\varepsilon = \mathrm{argmin}_{1\le d\le \iota} J_\varepsilon(d). \tag{67}$$

Note that Theorem 3 implies that the robust risks of procedure (33) with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfy the following oracle inequality

$$\mathbf{D}_\varepsilon(\widehat{q}_\varepsilon,\, q) \le \frac{1+3\delta}{1-3\delta} \min_{1\le d\le \iota} \mathbf{D}_\varepsilon(d, q) + \varepsilon^2\, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta}, \tag{68}$$

where the last term satisfies property (50).

8 Simulations

In this section, we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (33). In (1), we chose

$$S(t) = \sum_{j=1}^{10} \frac{j}{j+1}\, \phi_j(t), \tag{69}$$

with $\phi_j(t) = \sqrt{2}\sin(2\pi l_j t)$, $l_j = [\sqrt{j}]j$. We simulate the model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon \mathrm{d}w_t.$$

The frequency of observations per period equals $p = 100000$. We use the weight sequence as proposed by Galtchouk and Pergamenshchikov (2009a) for a discrete-time model: $k^* = 100 + \sqrt{|\ln \varepsilon|}$ and $m = [|\ln \varepsilon|^2]$. We calculated the empirical quadratic risk defined as

$$\overline{\mathbf{R}} = \frac{1}{p} \sum_{j=1}^{p} \widehat{\mathbf{E}}\, \left(\widetilde{S}_\varepsilon(u_j) - S(u_j)\right)^2, \quad u_j = j/p,$$

and the relative quadratic risk

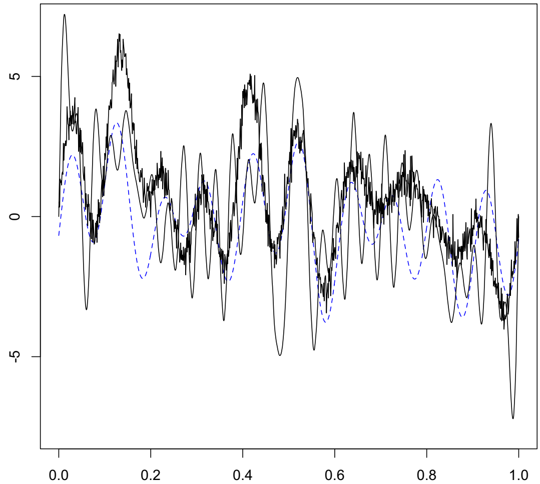
$$\overline{\mathbf{R}}_* = \overline{\mathbf{R}}/\|S\|^2_p \quad \text{and} \quad \|S\|^2_p = \frac{1}{p}\sum_{j=1}^{p} S^2(u_j).$$

The expectations was taken as an average over $N = 10000$ replications, i.e.,

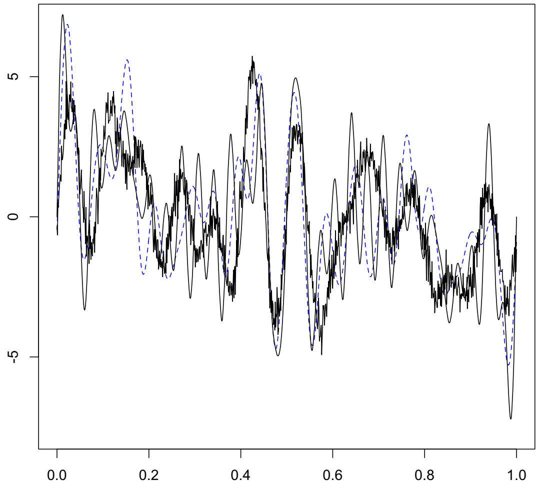
$$\widehat{\mathbf{E}}\,\left(\widetilde{S}_{\varepsilon}(\cdot) - S(\cdot)\right)^2 = \frac{1}{N}\sum_{l=1}^{N}\left(\widetilde{S}^{l}_{\varepsilon}(\cdot) - S(\cdot)\right)^2.$$

We used the cost function with $\delta = (3 + |\ln\varepsilon|)^{-2}$.
In the following graphics, the dashed line is the model selection procedure (33), the continuous line is the function (69) and the bold line is the corresponding observations (1).

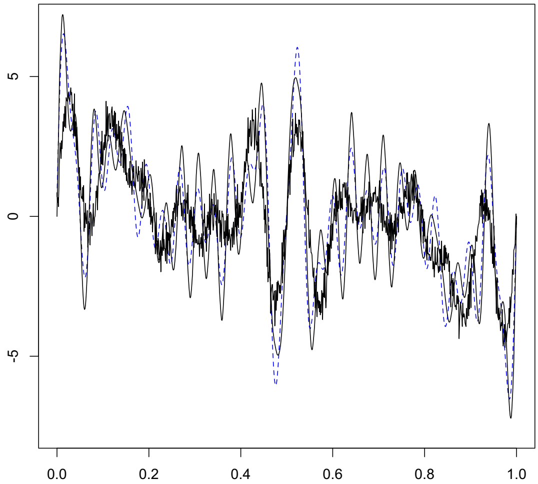
$\varepsilon = 1/\sqrt{20}$



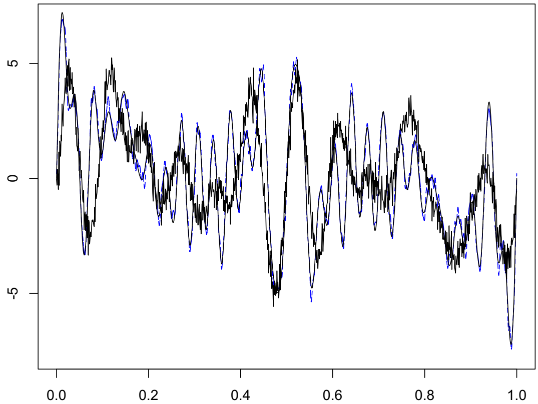
$\varepsilon = 1/\sqrt{100}$



$\varepsilon = 1/\sqrt{200}$



$\varepsilon = 1/\sqrt{1000}$



Empirical risks.

ε	$\overline{R}$	$\overline{R_*}$
$1/\sqrt{20}$	0.0158	0.307
$1/\sqrt{100}$	0.0113	0.059
$1/\sqrt{200}$	0.0076	0.04
$1/\sqrt{1000}$	0.0035	0.0185

Estimation of the number signals.

To estimate the signal number q, we use two procedures. The first $\widehat{q}_1$ is (68) with $\iota = [\ln \varepsilon^{-2}]$. The second $\widehat{q}_2$ is defined through the shrinkage approach for the model selection procedure (69).

$$\widehat{q}_2 = \inf\{j \geq 1 : |\widehat{\theta}_j| \leq \mathbf{c}_\varepsilon^*\}, \quad \mathbf{c}_\varepsilon^* = \varepsilon\sqrt{|\log \varepsilon|}.$$

ε	$\hat{q}_1$	$\hat{q}_2$
$1/\sqrt{20}$	6	5
$1/\sqrt{100}$	8	7
$1/\sqrt{200}$	9	7
$1/\sqrt{1000}$	10	9

***Remark* 10** From the simulation, we can conclude that the LSE procedure (68) is more appropriate than shrinkage method for such number detection problem.

9 Proofs

First we recall the Novikov inequalities (see Novikov 1975) also referred to as the Bichteler–Jacod inequalities (see Bichteler and Jacod 1983, Marinelli and Röckner 2014), providing upper bounds for the moments of the order $p \geq 2$ of the supremum of purely discontinuous local martingales

$$\mathbf{E} \sup_{0\leq t\leq 1} |g * (\mu - \widetilde{\mu})_t|^p \leq C_p^* \left(\mathbf{E}\left(|g|^2 * \widetilde{\mu}_1\right)^{p/2} + \mathbf{E}\left(|g|^p * \widetilde{\mu}_1\right)\right), \tag{70}$$

where C_p^* is some positive constant.

9.1 Proof of Proposition 2

First note that

$$B^2_{2,\varepsilon}(u) \leq 2\widetilde{\xi}_1^2 + 2B^2_{2,\varepsilon}(u'), \tag{71}$$

where $u' = (0, u_2, \ldots, u_n) \in \mathbb{R}^n$. It should be noted that

$$\mathbf{E}\,\widetilde{\xi}_1^2 \leq \mathbf{E}\,\xi_1^4 \leq 8\left(\varrho_1^4 \mathbf{E}\,w_1^4 + \varrho_2^4 \mathbf{E}\,z_1^4\right) = 8\left(3\varrho_1^4 + \varrho_2^4 \mathbf{E}\,z_1^4\right) \leq 8(3 + \mathbf{E}\,z_1^4)\varkappa_Q^2.$$

Note that from (70) and conditions (8) we obtain that

$$\mathbf{E}\,z_1^4 \leq \mathbf{C}^*(1 + \Pi(x^4)) \;<\; \infty.$$

To study the last term in the right-hand side of inequality (71), we set for any function f from $\mathbf{L}_2[0, 1]$

$$\check{I}_t(f) = \int_0^t f(s)\mathrm{d}\check{\xi}_s \quad \text{and} \quad \widetilde{I}_t(f) = \check{I}_t^2(f) - \mathbf{E}\,\check{I}_t^2(f).$$

Moreover, we set

$$D_t = \sum_{j=2}^{n} u_j \, \widetilde{I}_t(\phi_j).$$

Taking into account that $\widetilde{\xi}_j = \widetilde{I}_1(\phi_j)$ for $j \geq 2$, we obtain

$$B_{2,\varepsilon}(u') = \sum_{j=2}^{n} u_j \, \widetilde{I}_1(\phi_j) = D_1.$$

By the Itô formula, we can write that for any function f from $\mathbf{L}_2[0,1]$

$$\mathrm{d}\widetilde{I}_t(f) = 2\check{I}_{t-}(f) f(t) \mathrm{d}\xi_t + \varrho_2^2 \, f^2(t) \, \mathrm{d}\check{m}_t, \tag{72}$$

where $\check{m}_t = h_\varepsilon^2 * (\mu - \widetilde{\mu})_t$. So, setting

$$V_t = \sum_{j=2}^{n} u_j \, \check{I}_t(\phi_j)\phi_j(t) \quad \text{and} \quad \Psi_t = \sum_{j=2}^{n} u_j \, \phi_j^2(t),$$

we obtain that

$$\mathrm{d}D_t = 2 \, V_{t-} \, \mathrm{d}\xi_t + \varrho_2^2 \, \Psi_t \, \mathrm{d}\check{m}_t,$$

and therefore,

$$D_1^2 \leq 2 \, M_1^2 + 2\varrho_2^4 \, \check{M}_t^2, \tag{73}$$

where $M_t = \int_0^t V_{s-} \, \mathrm{d}\xi_s$ and $\check{M}_t = \int_0^t \Psi_s \, \mathrm{d}\check{m}_s$. Moreover, taking into account that for any f and g from $\mathbf{L}_2[0,1]$

$$\mathbf{E} \, \check{I}_t(f) \, \check{I}_t(g) = \check{\varkappa}_Q \int_0^t f(s)g(s) \, \mathrm{d}s,$$

we get

$$\begin{aligned}
\int_0^1 \mathbf{E} \, V_t^2 \, \mathrm{d}t &= \sum_{i,j=2}^{n} u_i \, u_j \int_0^1 \phi_i(t)\phi_j(t) \, \mathbf{E} \, \check{I}_t(\phi_i) \, \check{I}_t(\phi_j) \, \mathrm{d}t \\
&= \frac{\check{\varkappa}_Q}{2} \sum_{i,j=2}^{n} u_i \, u_j \left(\int_0^1 \phi_i(t)\phi_j(t) \, \mathrm{d}t \right)^2 = \frac{\check{\varkappa}_Q}{2} \sum_{i=2}^{n} u_i^2 = \frac{\check{\varkappa}_Q |u'|^2}{2} \leq \frac{\check{\varkappa}_Q}{2},
\end{aligned}$$

i.e.,

$$\mathbf{E} \left(\int_0^1 V_t \, \mathrm{d}w_t \right)^2 \leq \check{\varkappa}_Q / 2.$$

Now, we estimate the second term in inequality (73). To this end, we show that

$$\sup_{0\le t\le 1} \mathbf{E}\, \check{I}_t^4(f) < \infty. \tag{74}$$

Indeed, taking into account that for any non-random bounded function f the stochastic integral $\int_0^t f(s)\mathrm{d}w_s$ is $\left(0, \int_0^1 f^2(t)\mathrm{d}t\right)$ Gaussian random variable, we obtain through inequality (70) for some constant $\mathbf{C}^* > 0$

$$\begin{aligned}\mathbf{E}\, \check{I}_t^4(f) &\le 8\varrho_1^4\, \mathbf{E}\left(\int_0^t f(s)\mathrm{d}w_s\right)^4 + 8\varrho_2^4\, \mathbf{E}\left(\int_0^t f(s)\mathrm{d}\check{z}_s\right)^4 \\ &\le 24\varrho_1^4 \int_0^1 f^2(t)\mathrm{d}t + 8C^*\, \varrho_2^4 \left(\left(\Pi(h_\varepsilon^2)\int_0^1 f^2(t)\mathrm{d}t\right)^2 + \Pi(h_\varepsilon^4)\int_0^1 f^4(t)\mathrm{d}t\right).\end{aligned}$$

Therefore, we obtain inequality (74), from which using the Hölder inequality it follows that

$$\sup_{0\le t\le 1} \mathbf{E}\, V_t^4 < \infty.$$

From here using the properties of the stochastic integrals with respect to the Wiener process (see, for example, Lemma 4.12, p.125 in Liptser and Shiryaev 1977), we obtain that

$$\mathbf{E}\left(\int_0^t V_s\,\mathrm{d}w_s\right)^4 \le 36t^3 \int_0^t \mathbf{E}\, V_s^4\,\mathrm{d}s < \infty.$$

Moreover, from (70) we get

$$\sup_{0\le t\le 1} \mathbf{E}\left(\int_0^t V_{s-}\,\mathrm{d}\check{z}_s\right)^4 \le C^*\left(\left(\Pi(h_\varepsilon^2)\right)^2 + \Pi(h_\varepsilon^4)\right)\int_0^1 \mathbf{E}\, V_t^4\,\mathrm{d}t < \infty.$$

Therefore, $\sup_{0\le t\le 1} \mathbf{E}\, M_t^4 < \infty$ and

$$\int_0^1 \mathbf{E}\, M_t^2 V_t^2\,\mathrm{d}t \le \sup_{0\le t\le 1}\left(\mathbf{E}\, M_t^4\right)^{1/2}\left(\mathbf{E}\, V_t^4\right)^{1/2} < \infty.$$

This implies that $\mathbf{E}\int_0^1 M_{t-}\,\mathrm{d}M_t = 0$. So, by the Itô formula

$$\begin{aligned}\mathbf{E}\, M_1^2 &= \varrho_1^2\int_0^1 \mathbf{E}\, V_t^2\mathrm{d}t + \mathbf{E}\sum_{0\le t\le 1}(\Delta M_t)^2 \\ &= (\varrho_1^2 + \varrho_2^2\,\Pi(h_\varepsilon^2))\int_0^1 \mathbf{E}\, V_t^2\mathrm{d}t \le \check{\varkappa}_Q^2/2 \le \varkappa_Q^2/2.\end{aligned}$$

To estimate the last term in the right side of inequality (73), note that $\check{M}_t$ is a square integrated martingale with the quadratic characteristic

$$<\check{M}>_t = \Pi(h^4_\varepsilon)t,$$

i.e., $\mathbf{E}\int_0^1 \check{M}_{t-}\,\mathrm{d}\check{M}_t = 0$, and therefore, the Itô formula yields

$$\mathbf{E}\,\check{M}_1^2 = \mathbf{E}\sum_{0\le t\le 1}\left(\Psi_t\Delta\check{m}_t\right)^2 = \Pi(h^4_\varepsilon)\int_0^1 \Psi_t^2\mathrm{d}t \le \varrho_2^{-2}\,\Pi(x^2)(\overline{a}/\varepsilon)^2\left(\phi^*\right)^4\#(u).$$

Taking into account that $\Pi(x^2)=1$, we obtain that

$$\mathbf{E}\,D_1^2 \le \varkappa_Q^2 + 2\varrho_2^2\,(\phi^*)^4(\overline{a}/\varepsilon)^2\#(u) \le \varkappa_Q^2 + 2\varkappa_Q\,(\overline{a}/\varepsilon)^2\left(\phi^*\right)^4\#(u).$$

This implies the upper bound (23). □

9.2 Proof of Theorem 1

First note that we can rewrite the empirical squared error in (28) as follows

$$\mathrm{Err}_\varepsilon(\lambda) = J_\varepsilon(\lambda) + 2\sum_{j=1}^n \lambda(j)\check{\theta}_{j,\varepsilon} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda), \tag{75}$$

where $\check{\theta}_{j,\varepsilon} = \widetilde{\theta}_{j,\varepsilon} - \theta_j\widehat{\theta}_{j,\varepsilon}$. Now using the definition of $\widetilde{\theta}_{j,\varepsilon}$ in (29), we obtain that

$$\check{\theta}_{j,\varepsilon} = \varepsilon\theta_j\overline{\xi}_j + \varepsilon^2\widetilde{\xi}_j + \varepsilon^2(\mathbf{E}_Q\,\overline{\xi}^2_{j,\varepsilon} - \check{\varkappa}_Q) + \varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon),$$

where $\widetilde{\xi}_j = \overline{\xi}^2_j - \mathbf{E}_Q\,\overline{\xi}^2_j$ and $\overline{\xi}_j = \check{I}_1(\phi_j)$. Setting

$$M_\varepsilon(\lambda) = \varepsilon\sum_{j=1}^n \lambda(j)\theta_j\overline{\xi}_j \quad \text{and} \quad L(\lambda) = \sum_{j=1}^n \lambda(j), \tag{76}$$

we can rewrite (75) as

$$\begin{aligned} Er_\varepsilon(\lambda) &= J_\varepsilon(\lambda) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,L(\lambda) + 2M_\varepsilon(\lambda) + 2\varepsilon^2 B_{1,\varepsilon}(\lambda)\\ &\quad + 2\varepsilon\sqrt{P_\varepsilon(\lambda)}\frac{B_{2,\varepsilon}(u_\lambda)}{\sqrt{\check{\varkappa}_Q}} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda), \end{aligned} \tag{77}$$

where $u_\lambda = \lambda/|\lambda|$, the exact penalization is defined in (32) and the functions $B_{1,\varepsilon}(\cdot)$ and $B_{2,\varepsilon}(\cdot)$ are defined in (20). It should be noted that for the truncation parameter (27) bound (23) implies

$$\sup_{\lambda\in\Lambda}\mathbf{E}_Q\left|B^2_{2,\varepsilon}(u_\lambda)\right| \le U_Q + 4\,\varkappa_Q\left(\frac{\overline{a}}{\varepsilon}\right)^2 |\Lambda|_*\,(\phi^*)^4 := U_{1,Q}. \tag{78}$$

Using here the definition of the threshold $\overline{a} = \overline{a}_\varepsilon$ in(27), we obtain that this upper bound can be represented as $U_{1,Q} = U_Q + 4\varkappa_Q\,(\phi^*)^4$. Let $\lambda_0 = (\lambda_0(j))_{1\le j\le n}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be as in (33). Substituting λ_0 and $\widehat{\lambda}$ in equation (77), we obtain

$$\begin{aligned}\mathrm{Err}_\varepsilon(\widehat{\lambda}) - \mathrm{Err}_\varepsilon(\lambda_0) = {} & J(\widehat{\lambda}) - J(\lambda_0) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,L(\varpi) + 2\varepsilon^2 B_{1,\varepsilon}(\varpi)\\ & + 2M_\varepsilon(\varpi) + 2\varepsilon\sqrt{P_\varepsilon(\widehat{\lambda})}\frac{B_{2,\varepsilon}(\widehat{u})}{\sqrt{\check{\varkappa}_Q}} - 2\varepsilon\sqrt{P_\varepsilon(\lambda_0)}\frac{B_{2,\varepsilon}(u_0)}{\sqrt{\check{\varkappa}_Q}}\\ & - \delta\widehat{P}_\varepsilon(\widehat{\lambda}) + \delta\widehat{P}_\varepsilon(\lambda_0),\end{aligned} \tag{79}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{u} = u_{\widehat{\lambda}}$ and $u_0 = u_{\lambda_0}$. Note that by (25)

$$|L(\varpi)| \le\; L(\hat{\lambda}) + L(\lambda) \le 2|\Lambda|_*.$$

The inequality

$$2|ab| \le \delta a^2 + \delta^{-1}b^2 \tag{80}$$

implies that for any $\lambda \in \Lambda$

$$2\varepsilon\sqrt{P_\varepsilon(\lambda)}\frac{|B_{2,\varepsilon}(u_\lambda)|}{\sqrt{\check{\varkappa}_Q}} \le\; \delta P_\varepsilon(\lambda) + \varepsilon^2\frac{B^2_{2,\varepsilon}(u_\lambda)}{\delta\check{\varkappa}_Q}.$$

From bound (21), it follows that for $0 < \delta < 1$

$$\begin{aligned}\mathrm{Err}_\varepsilon(\hat{\lambda}) \le\; & \mathrm{Err}_\varepsilon(\lambda_0) + 2M_\varepsilon(\varpi) + 2\varepsilon^2\frac{B^*_{2,\varepsilon}}{\delta\check{\varkappa}_Q} + 2\varepsilon^2\,\check{\varkappa}_Q\\ & + \varepsilon^2|\widehat{\varkappa} - \check{\varkappa}_Q|(|\widehat{\lambda}|^2 + |\lambda_0|^2 + 4|\Lambda|_*) + 2\delta P_\varepsilon(\lambda_0),\end{aligned}$$

where $B^*_{2,\varepsilon} = \sup_{\lambda\in\Lambda} B^2_{2,\varepsilon}(u_\lambda)$. It should be noted that through (78) we can estimate this term as

$$\mathbf{E}_Q\,B^*_{2,\varepsilon} \le \sum_{\lambda\in\Lambda}\mathbf{E}_Q\,B^2_{2,\varepsilon}(u_\lambda) \le \iota U_{1,Q}. \tag{81}$$

Taking into account that $\sup_{\lambda\in\Lambda}|\lambda|^2 \le |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned}\mathrm{Err}_\varepsilon(\widehat{\lambda}) \le {} & \mathrm{Err}_\varepsilon(\lambda_0) + 2M_\varepsilon(\varpi) + 2\varepsilon^2\frac{B^*_{2,\varepsilon}}{\delta\check{\varkappa}_Q} + 2\varepsilon^2\,\check{\varkappa}_Q\\ & + \frac{6\varepsilon^2|\Lambda|_*}{n}|\widehat{\varkappa} - \check{\varkappa}_Q| + 2\delta P_\varepsilon(\lambda_0).\end{aligned} \tag{82}$$

To estimate the second term in the right-hand side of this inequality, we introduce

$$S_{\upsilon}=\sum_{j=1}^{n}\upsilon(j)\theta_{j}\phi_{j},\quad \upsilon=(\upsilon(j))_{1\le j\le n}\in\mathbb{R}^{n}.$$

Moreover, note that

$$M_{\varepsilon}^{2}(\upsilon)\le 2\varepsilon^{2}\left(\upsilon^{2}(1)\,\theta_{1}^{2}\xi_{1}^{2}+\check{I}_{1}(\Phi)\right),$$

where $\Phi(t)=\sum_{j=2}^{n}\upsilon(j)\theta_{j}\phi_{j}(t)$. Therefore, thanks to (16) we obtain that for any non-random $\upsilon\in\mathbb{R}^{n}$

$$\mathbf{E}M_{\varepsilon}^{2}(\upsilon)\le 2\,\varkappa_{Q}\varepsilon^{2}\upsilon^{2}(1)\theta_{1}^{2}+2\check{\varkappa}_{Q}\varepsilon^{2}\sum_{j=2}^{n}\upsilon^{2}(j)\theta_{j}^{2}\le 2\,\varkappa_{Q}\varepsilon^{2}||S_{\upsilon}||^{2}. \qquad (83)$$

To estimate this function for a random vector, we set

$$M_{\varepsilon}^{*}=\sup_{\upsilon\in\Lambda_{1}}\frac{M^{2}(\upsilon)}{\varepsilon^{2}||S_{\upsilon}||^{2}}\quad\text{and}\quad \Lambda_{1}=\Lambda-\lambda_{0}.$$

So, through inequality (80)

$$2|M_{\varepsilon}(\upsilon)|\le\delta||S_{\upsilon}||^{2}+\varepsilon^{2}\frac{M_{\varepsilon}^{*}}{\delta}. \qquad (84)$$

It is clear that the last term here can be estimated as

$$\mathbf{E}\,M_{\varepsilon}^{*}\le\sum_{\upsilon\in\Lambda_{1}}\frac{\mathbf{E}\,M_{\varepsilon}^{2}(\upsilon)}{\varepsilon^{2}||S_{\upsilon}||^{2}}\le 2\sum_{\upsilon\in\Lambda_{1}}\varkappa_{Q}=2\,\varkappa_{Q}\,\iota, \qquad (85)$$

where $\iota=\#(\Lambda)$. Moreover, note that, for any $\upsilon\in\Lambda_{1}$,

$$||S_{\upsilon}||^{2}-||\widehat{S}_{\upsilon}||^{2}=\sum_{j=1}^{n}\upsilon^{2}(j)(\theta_{j}^{2}-\widehat{\theta}_{j}^{2})\le 2|M_{\varepsilon}(\upsilon^{2})|,$$

where $\upsilon^{2}=(\upsilon^{2}(j))_{1\le j\le n}$. Taking into account now that for any $x\in\Lambda_{1}$ the components $|\upsilon(j)|\le 1$, we can estimate the last term as in (83), i.e.,

$$\mathbf{E}\,M_{\varepsilon}^{2}(\upsilon^{2})\le 2\varepsilon^{2}\varkappa_{Q}\,||S_{\upsilon}||^{2}.$$

Similarly, setting

$$M_{1,\varepsilon}^{*}=\sup_{\upsilon\varepsilon\Lambda_{1}}\frac{M_{\varepsilon}^{2}(\upsilon^{2})}{\varepsilon^{2}||S_{\upsilon}||^{2}}$$

we obtain

$$\mathbf{E}_Q \, M^*_{1,\varepsilon} \le 2\varkappa_Q \, \iota. \tag{86}$$

By the same way, we find that

$$2|M_\varepsilon(\upsilon^2)| \le \delta||S_\upsilon||^2 + \frac{M^*_{1,\varepsilon}}{n\delta},$$

and for any $0 < \delta < 1$,

$$||S_\upsilon||^2 \le \frac{||\widehat{S}_\upsilon||^2}{1-\delta} + \frac{\varepsilon^2 M^*_{1,\varepsilon}}{\delta(1-\delta)}.$$

So, from (84) we get

$$2M(\upsilon) \le \frac{\delta||\widehat{S}_\upsilon||^2}{1-\delta} + \frac{\varepsilon^2(M^*_\varepsilon + M^*_{1,\varepsilon})}{\delta(1-\delta)}.$$

Therefore, taking into account that $\|\widehat{S}_\varpi\|^2 \le 2\,(\mathrm{Err}_\varepsilon(\widehat{\lambda}) + \mathrm{Err}_\varepsilon(\lambda_0))$, the term $M_\varepsilon(\varpi)$ can be estimated as

$$2M_\varepsilon(\varpi) \le \frac{2\delta(\mathrm{Err}_\varepsilon(\widehat{\lambda}) + \mathrm{Err}_\varepsilon(\lambda_0))}{1-\delta} + \frac{\varepsilon^2(M^*_\varepsilon + M^*_{1,\varepsilon})}{\delta(1-\delta)}.$$

Using this bound in (82), we obtain

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) \le\ & \frac{1+\delta}{1-3\delta}\mathrm{Err}_\varepsilon(\lambda_0) + \frac{\varepsilon^2(M^*_\varepsilon + M^*_{1,\varepsilon})}{\delta(1-3\delta)} + \frac{2\varepsilon^2 B^*_{2,\varepsilon}}{\delta(1-3\delta)\check{\varkappa}_Q} \\ & + \frac{2\varepsilon^2\,\varkappa_Q}{1-3\delta} + \frac{6\varepsilon^2\,|\Lambda|_*}{(1-3\delta)}|\widehat{\varkappa} - \check{\varkappa}_Q| + \frac{2\delta}{(1-3\delta)}P_\varepsilon(\lambda_0).\end{aligned}$$

Moreover, for $0 < \delta < 1/6$ we can rewrite this inequality as

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) \le\ & \frac{1+\delta}{1-3\delta}\mathrm{Err}_\varepsilon(\lambda_0) + \frac{2\varepsilon^2(M^*_\varepsilon + M^*_{1,\varepsilon})}{\delta} + \frac{4\varepsilon^2 B^*_{2,\varepsilon}}{\delta\check{\varkappa}_Q} \\ & + 4\varepsilon^2\,\varkappa_Q + 12\varepsilon^2\,|\Lambda|_*|\widehat{\varkappa} - \check{\varkappa}_Q| + 4\delta\,P_\varepsilon(\lambda_0).\end{aligned}$$

Using here bounds (81), (85), (86) and taking into account that $\check{\varkappa}_Q \le \varkappa_Q$, we obtain

$$\begin{aligned}\mathcal{R}(\widehat{S}_*, S) \le\ & \frac{1+\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{4\varepsilon^2\,\varkappa_Q(2\iota+\delta)}{\delta} + \frac{4\varepsilon^2 U_{1,Q}\iota}{\delta\check{\varkappa}_Q} \\ & + 12\varepsilon^2\,|\Lambda|_*\mathbf{E}_Q\,|\widehat{\varkappa} - \check{\varkappa}_Q| + \frac{2\delta}{1-3\delta}\,P_\varepsilon(\lambda_0).\end{aligned}$$

Now, from Lemma 1 it follows that

$$\mathcal{R}(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{4\varepsilon^2 \varkappa_Q (2\iota + \delta)}{\delta} + \frac{4\varepsilon^2 U_{1,Q}\iota}{\delta \check{\varkappa}_Q}$$
$$+ 12\varepsilon^2 \, |\Lambda|_* \mathbf{E}_Q \, |\widehat{\varkappa} - \check{\varkappa}_Q| + \varepsilon^2 \frac{2\delta}{1-3\delta} \, \varkappa_Q.$$

Taking into account here that $2\delta/(1-3\delta) \le 1$ for $0 < \delta < 1/6$ and using the function (44), we obtain inequality (45) for some constant $\mathbf{l}_* > 0$ which depends on $\Pi(x^4)$. Hence Theorem 1 is shown. □

9.3 Proof of Theorem 5

Firstly, note that for any fixed $Q \in \mathcal{Q}^*_\varepsilon$

$$\sup_{S \in W^k_r} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) \ge \sup_{S \in W^k_r} \mathcal{R}_Q(\widehat{S}_\varepsilon, S). \tag{87}$$

Now for any fixed $0 < \check{\gamma} < 1$, we set

$$d = d_\varepsilon = \left[\frac{k+1}{k} \upsilon_\varepsilon^{1/(2k+1)} \, l_*(r_0)\right] \quad \text{and} \quad \mathbf{r}_0 = (1-\check{\gamma})\mathbf{r}. \tag{88}$$

Using this definition, we introduce the parametric family $(S_z)_{z \in \mathbb{R}^d}$ as

$$S_z(x) = \sum_{j=1}^{d} z_j \, \phi_j(x). \tag{89}$$

To define the Bayesian risk, we choose a prior distribution on $\mathbb{R}^d$ as

$$\kappa = (\kappa_j)_{1 \le j \le d} \quad \text{and} \quad \kappa_j = s_j \, \eta_j, \tag{90}$$

where η_j are i.i.d. Gaussian $\mathcal{N}(0, 1)$ random variables and the coefficients

$$s_j = \sqrt{\frac{s^*_j}{\upsilon_\varepsilon}} \quad \text{and} \quad s^*_j = \left(\frac{d}{j}\right)^k - 1.$$

Denoting by μ_κ the distribution of the random variables $(\kappa_j)_{1 \le j \le d}$ on $\mathbb{R}^d$, we introduce the Bayesian risk as

$$\widetilde{\mathcal{R}}_Q(\widehat{S}) = \int_{\mathbb{R}^d} \mathcal{R}_Q(\widehat{S}, S_z) \, \mu_\kappa(\mathrm{d}z). \tag{91}$$

Furthermore, for any function $f \in \mathbf{L}_2[0,1]$, we denote by $\mathbf{p}(f)$ its projection in $\mathbf{L}_2[0,1]$ onto $W_{k,r}$, i.e., $\|f-\mathbf{p}(f)\| = \inf_{h\in W_r^k}\|f-h\|$. Since W_r^k is a convex and closed set in $\mathbf{L}_2[0,1]$, this projector exists and is unique for any function $f\in\mathbf{L}_2[0,1]$, and moreover, $\|f-h\|^2 \ge \|\mathbf{p}(f)-h\|^2$ for any $h\in W_r^k$. So, setting $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{S})$, we obtain that

$$\sup_{S\in W_r^k} \mathcal{R}(\widehat{S},S) \ge \int_{\{z\in\mathbb{R}^d\,:\,S_z\in W_r^k\}} \mathbf{E}_{S_z}\|\widehat{\mathbf{p}}-S_z\|^2\,\mu_\kappa(\mathrm{d}z).$$

Taking into account now that $\|\widehat{\mathbf{p}}\|^2 \le \mathbf{r}$, we obtain

$$\sup_{S\in W_r^k} \mathcal{R}_Q(\widehat{S},S) \ge \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon \tag{92}$$

and $\Delta_\varepsilon = \int_{\{z\in\mathbb{R}^d\,:\,S_z\notin W_{k,\mathbf{r}}\}}(\mathbf{r}+\|S_z\|^2)\,\mu_\kappa(\mathrm{d}z)$. Therefore, in view of (87),

$$\sup_{S\in W_{k,\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon,S) \ge \sup_{Q\in\mathcal{Q}^*_\varepsilon} \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon. \tag{93}$$

As to the last term in this inequality, in Appendix we show that for any $\mathbf{b}>0$

$$\lim_{\varepsilon\to 0} \varepsilon^{-\mathbf{b}}\,\Delta_\varepsilon = 0. \tag{94}$$

Now it is easy to see that $\|\widehat{\mathbf{p}}-S_z\|^2 \ge \sum_{j=1}^d(\widehat{z}_j-z_j)^2$, where $\widehat{z}_j = \int_0^1\widehat{\mathbf{p}}(t)\,\phi_j(t)\mathrm{d}t$. So, in view of Proposition 4 and reminding that $\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*$ we obtain

$$\sup_{Q\in\mathcal{Q}^*_\varepsilon}\widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) \ge \sup_{0<\varrho_1^2\le\varsigma^*}\sum_{j=1}^d \frac{1}{\varepsilon^{-2}\,\varrho_1^{-2}+\upsilon_\varepsilon\,(s_j^*)^{-1}}$$
$$= \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d\frac{s_j^*}{s_j^*+1} = \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d\left(1-\frac{j^k}{d_\varepsilon^k}\right).$$

Therefore, using now definition (88), inequality (93) and limit (94), we obtain that

$$\liminf_{n\to\infty}\inf_{\widehat{S}\in\Pi_\varepsilon}\upsilon_\varepsilon^{\frac{2k}{2k+1}}\sup_{S\in W_{k,r}}\mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon,S) \ge (1-\check{\gamma})^{\frac{1}{2k+1}}\,l_*(\mathbf{r}).$$

Taking here limit as $\check{\gamma}\to 0$ implies Theorem 5. □

9.4 Proof of Theorem 6

First we suppose that the parameters $k\ge 1$, $\mathbf{r}>0$ in (52) and ς^* in (9) are known. Let the family of admissible weighted least squares estimates $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ given by (42).

Consider the pair $\check{\alpha} = (k, \check{r})$ and $\check{r} = \varpi\,[\mathbf{r}/\varpi]$. Denote the corresponding estimate as

$$\check{S} = \widehat{S}_{\check{\lambda}} \quad \text{and} \quad \check{\lambda} = \lambda_{\check{\alpha}}. \tag{95}$$

Note that for sufficiently small ε the pair $\check{\alpha}$ belongs to set (37).

Proposition 5 *Assume that conditions* $(\mathbf{A}_1)$–$(\mathbf{A}_2)$ *hold. Then the estimator* $\check{S}$ *admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to 0}\, \upsilon_\varepsilon^{2k/(2k+1)} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\check{S}, S) \le l_*(\mathbf{r}). \tag{96}$$

Combining Theorem 4 and Proposition 5 yields Theorem 6. □

Acknowledgements The second and last authors are partly supported by the Ministry of Education and Science of the Russian Federation in the framework of the research Project No. 2.3208.2017/4.6. The last author is partly supported by the Russian Federal Professor program (Project No. 1.472.2016/1.4, Ministry of Education and Science).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC – 19*, 716–723.

Barron, A., Birgé, L., Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, *113*, 301–415.

Bichteler, K., Jacod, J. (1983). Calcul de Malliavin pour les diffusions avec sauts: Existence d'une densité dans le cas unidimensionnel. In *Séminaire de probabilité. Lecture notes in mathematics, XVII* (Vol. 986, pp. 132–157). Berlin: Springer.

Cont, R., Tankov, P. (2004). *Financial modelling with jump processes*. *Financial mathematics series*. Boca Raton: Chapman & Hall/CRC.

El-Behery, I. N., Macphie, R. H. (1978). Maximum likelihood estimation of the number, directions and strengths of point radio sources from variable baseline interferometer data. *IEEE Transactions on Antennas and Propagation*, *26*(2), 294–301.

Flaksman, A. G. (2002). Adaptive signal processing in antenna arrays with allowance for the rank of the impule-response matrix of a multipath channel. *Izvestiya Vysshikh Uchebnykh Zavedenij, Radiofizika*, *45*(12), 1064–1077.

Fourdrinier, D., Pergamenshchikov, S. (2007). Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*, *59*(3), 435–464.

Galtchouk, L. I., Pergamenshchikov, S. M. (2006). Asymptotically efficient estimates for non parametric regression models. *Statistics and Probability Letters*, *76*(8), 852–860.

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

Galtchouk, L. I., Pergamenshchikov, S. M. (2009a). Sharp non asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, *21*(1), 1–16.

Ibragimov, I. A., Khasminskii, R. Z. (1981). *Statistical estimation: Asymptotic theory*. New York: Springer.

Jacod, J., Shiryaev, A. N. (2002). *Limit theorems for stochastic processes*. 2nd ed. Berlin: Springer.

Kassam, S. A. (1988). *Signal detection in non-Gaussian noise*. *IX*. New York: Springer.

Konev, V. V., Pergamenshchikov, S. M. (2009a). Nonparametric estimation in a semimartingale regression model. Part 1. Oracle inequalities. *Tomsk State University Journal of Mathematics and Mechanics*, *3*, 23–41.

Konev, V. V., Pergamenshchikov, S. M. (2009b). Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Tomsk State University Journal of Mathematics and Mechanics*, *4*, 31–45.
Konev, V. V., Pergamenshchikov, S. M. (2012). Efficient robust nonparametric estimation in a semimartingale regression model. *Annales de l'institut Henri Poincare (B) Probability and Statistics*, *48*(4), 1217–1244.
Konev, V. V., Pergamenshchikov, S. M. (2015). Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic Processes and Their Applications*, *125*, 294–326.
Konev, V., Pergamenshchikov, S., Pchelintsev, E. (2014). Estimation of a regression with the impulse type noise from discrete data. *Theory of Probability and Its Applications*, *58*(3), 442–457.
Kutoyants, Yu A. (1984). *Parameter estimation for stochastic processes*. Berlin: Heldeman-Verlag.
Kutoyants, Y. A. (1994). *Identification of dynamical systems with small noise*. *Mathematics and its applications*, Vol. 300. Dordrecht: Kluwer Academic Publishers Group.
Liptser, R. S., Shiryaev, A. N. (1977). *Statistics of random processes*. *General theory*. New York: Springer.
Mallows, C. (1973). Some comments on C_p. *Technometrics*, *15*, 661–675.
Manelis, B. V. (2007). Algorithms for tracking and demodulation of the mobile communication signal in the conditions of the Indiscriminating multipath propagation. *Radiotekhnika*, *4*, 16–21. **(in Russian)**.
Marinelli, C., Röckner, M. (2014). On maximal inequalities for purely discontinuous martingales in infinite dimensions. In *Séminaire de Probabilités*. *Lecture notes in mathematics, XLVI* (pp. 293–315). Berlin: Springer.
Novikov, A. A. (1975). On discontinuous martingales. *Theory of Probability and Its Applications*, *20*(1), 11–26.
Pchelintsev, E. (2013). Improved estimation in a non-Gaussian parametric regression. *Statistical Inference for Stochastic Processes*, *16*(1), 15–28.
Pinsker, M. S. (1981). Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission information*, *17*, 120–133.
Trifonov, A. P., Kharin, A. V. (2013). Estimating the number of signals with unknown amplitudes. *Nelineinyi Mir*, *11*(12), 853–866. **(in Russian)**.
Trifonov, A. P., Kharin, A. V. (2015a). Estimation of the number of radio signals with unknown amplitudes and phases. *Radiophysics and Quantum Electronics*, *58*(1), 57–70.
Trifonov, A. P., Kharin, A. V. (2015b). Estimation of the number of orthogonal signals with unknown non-energy parameters. *Radioelectronics and Communications Systems*, *58*(8), 362–370.
Trifonov, A. P., Kharin, A. V., Chernoyarov, O. V., Kalashnikov, K. S. (2015). Determining the number of radio signals with unknown phases. *International Journal on Communications Antenna and Propagation*, *5*(6), 367–374.
Trifonov, A. P., Shinakov, Yu S. (1986). *Joint distinction of signals and estimation of their parameters on the background noise*. Moscow: Radio i Svyaz'. **(in Russian)**.

Affiliations

Slim Beltaief[1] · Oleg Chernoyarov[2] · Serguei Pergamenchtchikov[3]

Slim Beltaief
beltaiefslim@hotmail.fr

Oleg Chernoyarov
chernoyarovov@mpei.ru

1 Laboratoire de Mathématiques Raphael Salem, Université de Rouen Normandie, Avenue de l'Université, BP.12, Technopôle du Madrillet, Saint-Étienne-du-Rouvray 76801, France

[2] International Laboratory SSP&QF of Tomsk State University and Department of Radio Engineering Devices and Antenna Systems, National Research University “Moscow Power Engineering Institute”, Krasnokazarmennaya 14, Moscow 111250, Russia

[3] International Laboratory SSP&QF of Tomsk State University and Laboratoire de Mathématiques Raphael Salem, Université de Rouen Normandie, Avenue de l’Université, BP.12, Technopôle du Madrillet, Saint-Étienne-du-Rouvray 76801, France