Adaptive efficient estimation for generalized semi-Markov Big Data models

Vlad Stefan Barbu, Slim Beltaief, Serguei M Pergamenchtchikov

Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Adaptive efficient estimation for generalized semi-Markov Big Data models

Vlad Stefan Barbu · Slim Beltaief · Serguei Pergamenchtchikov

Received: date / Revised: date

Abstract In this paper we study generalized semi-Markov high dimension regression models in continuous time observed in fixed discrete time moments. The generalized semi-Markov process has dependent jumps and, therefore, it is an extension of the semi-Markov regression introduced in Barbu, Beltaief and Pergamenshchikov (2019a). For such models we consider estimation problems in nonparametric setting. To this end we develop model selection procedures for which sharp non-asymptotic oracle inequalities for the robust risks are obtained. Moreover, we give constructive sufficient conditions which provide through the obtained oracle inequalities the adaptive robust efficiency property in minimax sense. It should be noted also that for these results we do not use either sparse conditions or the parameter dimension in the model. As examples, it is considered regression models constructed through spherical symmetric noise impulses and truncated fractional Poisson processes. Numeric Monte-Carlo simulations confirming the theoretical results are given in the supplementary materials.

This research was supported by RSF, project no 20-61-47043 (National Research Tomsk State University, Russia)

V. S. Barbu
Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen Normandie, France
E-mail: vladstefan.barbu@univ-rouen.fr

S. Beltaief
ALTEN de Toulouse (*https://www.alten.fr*), France
E-mail: beltaiefslim@hotmail.fr

S. M. Pergamenchtchikov
Laboratoire de Mathématiques Raphael Salem, Université de Rouen, Avenue de l'Université, BP.12, 76801 Saint-Etienne-du-Rouvray, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, Tomsk, Russia
E-mail: Serge.Pergamenshchikov@univ-rouen.fr

Keywords Regression model · Generalized semi-Markov processes · fractional Poisson processes · Non-asymptotic estimation · Robust estimation · Model selection · Sharp oracle inequalities · Asymptotic efficiency · Adaptive estimation

1 Introduction

1.1 Motivations

In this paper we study the following linear regression model in continuous time

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \beta_j \mathbf{u}_j(t) \right) \mathrm{d}t + \mathrm{d}\xi_t \,, \quad 0 \le t \le T \,, \tag{1}$$

where the functions $(\mathbf{u}_j)_{1\le j\le q}$ are known linear independent 1-periodic $\mathbb{R} \to \mathbb{R}$ functions, the duration of observations T is an integer number and $(\xi_t)_{t\ge 0}$ is an unobservable noise process defined in Section 2. The process (1) is observed only at the fixed time moments

$$(y_{t_j})_{0\le j\le n} \,, \quad t_j = \frac{j}{p} \quad \text{and} \quad n = pT \,, \tag{2}$$

where the observations frequency p is some fixed integer number. We consider the model (1) in the case when the parameter dimension is grater than the number of observations, i.e., $q > n$. Such models are called big data or high dimension regression in continuous time (see, for example, in Fujimori (2019) for diffusion processes). The problem is to estimate the unknown parameters $(\beta_j)_{1\le j\le q}$ on the basis of the observations (2). Usually for such problems one uses either the Lasso algorithm or the Dantzig selector method. It should be emphasized that to apply these methods one needs to assume sparsity conditions which provide the non large ("reasonable") number of the non-zero unknown parameters and, moreover, the parameter dimension q must be known (see, for example, in Hasttie, Friedman and Tibshirani (2008)). It should be noted also that the case of unknown parameter dimension q is one of the crucial points in important practical problems such as, for example, signal and image statistical processing (see, for example, Beltaief, Chernoyarov and Pergamenshchikov (2020) and the references therein). In this paper we study the model (1) in the nonparametric setting as the estimation problem for the function

$$S(t) = \sum_{j=1}^{q} \beta_j \mathbf{u}_j(t) \,,$$

i.e.

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t \,, \quad 0 \le t \le T \,, \tag{3}$$

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function from $\mathcal{L}_2[0,1]$. Here, we assume neither sparsity conditions, nor the condition that the parameter dimension is known, i.e., in particular, we can assume that $q = +\infty$. Now the problem is to estimate the unknown function S in the model (3) on the basis of observations (2). Originally, such problems were considered in the framework "signal+white noise" models (see, for example, Ibragimov and Khasminskii (1981); Kutoyants (1994); Pinsker (1981)). Later, it was extended to the "color noise" models defined through non Gaussian Ornstein-Uhlenbeck processes Barndorff-Nielsen and Shephard (2001); Konev and Pergamenshchikov (2012, 2015). The problem here is that the dependence defined on the basis of the Ornstein-Uhlenbeck processes disappears very fast, at a geometric rate. This means that such models are asymptotically equivalent to models with independent observations. To keep the dependence in the observations for large time periods for the estimation problem on the complete data in the paper Barbu, Beltaief and Pergamenshchikov (2019a) it is proposed to define the model (3) through semi-Markov processes with jumps. Such models considerably extend the potential applications of statistical results in many important practical fields such as finance, insurance, signals and image processing, reliability, biology (see, for example, Barbu, Beltaief and Pergamenshchikov (2019a); Barbu and Limnios (2008) and the references therein). In this paper we extend the semi-Markov regression models to the generalized semi-Markov processes by introducing an additional dependence in jump sizes of $(\xi_t)_{t\geq 0}$.

1.2 Methods

In this paper, in order to estimate the function S, we develop model selection methods using the quadratic risks defined as

$$\mathcal{R}_Q(\widehat{S}_T, S) = \mathbf{E}_{Q,S} \|\widehat{S}_T - S\|^2 , \quad \|f\|^2 = \int_0^1 f^2(s)\mathrm{d}s , \tag{4}$$

where $\widehat{S}_T(\cdot)$ is some estimate (i.e. any periodical function measurable with respect to the observations $\sigma\{y_{t_0}, \dots y_{t_n}\}$) and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (3) corresponding to the unknown noise distribution Q in the Skorokhod space $\mathcal{D}[0,T]$ and to the function S. We assume that this distribution belongs to some distribution family $\mathcal{Q}_T$ specified in Section 2. To study the properties of the estimators uniformly over the noise distribution (what is really needed in practice), we use the robust risk defined as

$$\mathcal{R}^*_T(\widehat{S}_T, S) = \sup_{Q\in\mathcal{Q}_T} \mathcal{R}_Q(\widehat{S}_T, S) . \tag{5}$$

It should be noted that statistical procedures which are optimal in the sense of this risk possess stable mean square accuracy uniformly over all possible admissible noise distributions in the model (3). This means that the corresponding

statistical optimal algorithms have high noise immunity and, therefore, significantly improve the quality and reliability of statistical inferences obtained on their basis.

To construct model selection procedures on the basis of the discrete data (2) we use the approach proposed in Konev and Pergamenshchikov (2015). It should be noted that the main analytic tool in this paper is based on the exponential decrease rate of the dependence in Ornstein-Uhlenbeck models, and, therefore, we cannot apply these methods to semi-Markov models, which can retain a dependence in noises for a long time. So, in this paper, to study the estimation problem on the discrete observations (2) for the model (3) with noises defined through semi-Markov processes we develop new methods based on the special renewal theory from Barbu, Beltaief and Pergamenshchikov (2019a); based on these techniques we can analyse the approximation errors in the discrete observations and obtain non asymptotic sharp oracle inequalities. Moreover, as a consequence, we found constructive sufficient conditions on the observations frequency which provide the robust efficiency for proposed model selection procedures in adaptive setting, i.e. in the case when the regularity properties of the function S are unknown.

1.3 Main contributions of this paper

In this paper we use for the first time nonparametric adaptive methods for estimation problems in the framework of the big data generalized semi-Markov regression models. To this end we develop model selection procedures and corresponding analytical tools providing, under some constructive sufficient conditions, the optimality in the sharp oracle inequality sense and the robust adaptive efficiency in the minimax sense for the proposed estimators. It turns out that these conditions hold true for important practical cases such as, for example, regression models constructed through truncated fractional Poisson processes introduced in Barbu, Beltaief and Pergamenshchikov (2019b). Moreover, in this paper, we extend for the first time the model from Barbu, Beltaief and Pergamenshchikov (2019a) using the generalized semi-Markov models obtained by introducing a dependence structure in the sizes of the jumps. As an example, we use spherically symmetric random variables, which play very important role in many practical applications (see, for example, Fourdrinier and Pergamenshchikov (2007) and the references therein).

1.4 Organization of the paper

The rest of the paper is organized as follows. In Section 2 we state the main conditions under which we consider the model (3). In Section 3 we represent fractional Poisson processes and its main properties. In Section 4 we construct model selection procedures on the basis of weighted least squares estimates. In Section 5 we state the main results. In section 6 we develop the stochastic

calculus for the generalized semi-Markov processes. Section 7 gives the proofs of the main results. Some auxiliary tools are given in an Appendix.

2 Main conditions

First, we assume that the noise process $(\xi_t)_{t\geq 0}$ in the model (3) is defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 L_t + \varrho_3 z_t\,, \tag{6}$$

where ϱ_1, ϱ_2 and ϱ_3 are unknown coefficients, $(w_t)_{t\geq 0}$ is a standard Brownian motion, $L_t = \int_0^t \int_{\mathbb{R}_*} x(\mu(\mathrm{d}s,\mathrm{d}x) - \widetilde{\mu}(\mathrm{d}s,\mathrm{d}x))$, $\mu(\mathrm{d}s\,\mathrm{d}x)$ is the jump measure with deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is the Lévy measure on $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$ (see, for example Liptser and Shiryaev (1989) for details), with

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^8) < \infty\,. \tag{7}$$

Here we use the usual notations for $\Pi(|x|^m) = \int_{\mathbb{R}} |z|^m\,\Pi(\mathrm{d}z)$. Note that $\Pi(|x|)$ may be equal to $+\infty$. In this paper we assume that the "dependent part" in the noise (6) is modelled by the *generalized semi-Markov process* $(z_t)_{t\geq 0}$ defined as

$$z_t = \sum_{i=1}^{N_t} \zeta_i, \tag{8}$$

where $(\zeta_i)_{i\geq 1}$ are random variables satisfying the following conditions:

$\mathbf{C}_1$) $\forall i \geq 1$ *the expectations* $\mathbf{E}\,\zeta_i = 0$, $\mathbf{E}\,\zeta_i^2 = 1$ *and* $\sup_{l\geq 1}\mathbf{E}\,\zeta_l^4 < \infty$*;*

$\mathbf{C}_2$) $\mathbf{E}\,\zeta_i\,\zeta_j = 0$ *for any* $i \neq j$*;*

$\mathbf{C}_3$) *For any* $1 \leq k_1 < k_2 < k_3 < k_4$ *the random variables* $(\zeta_{k_i})_{1\leq i\leq 4}$ *are such that* $\mathbf{E}\,\zeta_{k_1}^{\iota_1}\,\zeta_{k_2}^{\iota_2}\,\zeta_{k_3}^{\iota_3}\,\zeta_{k_4}^{\iota_4} = 0$ *for any* $\iota_1,\ldots,\iota_4 \in \{0,1,2,3\}$ *for which* $3 \leq \sum_{i=1}^4 \iota_i \leq 4$ *and at least one among them is equal to one.*

Now we give some examples for the correlation conditions $\mathbf{C}_1$) – $\mathbf{C}_3$). To this end, we first remind the definition of spherically symmetric distribution (see, for example, in Fourdrinier and Pergamenshchikov (2007)). A random vector $\zeta = (\zeta_1,\ldots,\zeta_d)'$ is called spherically symmetric if its density in $\mathbb{R}^d$ has the form $\mathbf{g}(|\cdot|^2)$ for some nonnegative function $\mathbf{g}$. Here the prime denotes the transposition. Note that there is a very important particular case of the spherically symmetric vectors represented by Gaussian mixture distributions. The vector $\zeta = (\zeta_1,\ldots,\zeta_d)'$ is called Gaussian mixture in $\mathbb{R}^d$ if it has the spherically symmetric distribution with

$$\mathbf{g}(t) = \mathbf{E}\,\frac{1}{(2\pi\mathbf{s})^{d/2}}\,e^{-\frac{t}{2\mathbf{s}^2}}\,, \tag{9}$$

where $\mathbf{s}$ is a non negative random variable. It should be emphasized that in radio-physics such distributions are very popular for statistical signal processing (see, for example, Middleton (1979); Kassam (1988)). Using these definitions it is easy to see that the following random variables satisfy the conditions $\mathbf{C}_1$) – $\mathbf{C}_3$):

– $(\zeta_j)_{j\geq 1}$ are i.i.d. random variables satisfying condition $\mathbf{C}_1$);
– for some $d > 1$ the random vector $(\zeta_1, \ldots, \zeta_d)'$ that has a spherically symmetric distribution in $\mathbb{R}^d$, with $\mathbf{E}\,\zeta_1^2 = 1$, $\mathbf{E}\,\zeta_1^4 < \infty$ and the random variables $(\zeta_j)_{j> d}$ are independent and satisfying condition $\mathbf{C}_1$);
– for any $d \geq 1$ a random vector $(\zeta_1, \ldots, \zeta_d)'$ is a Gaussian mixture with mixture variable $\mathbf{s}$ for which $\mathbf{E}\,\mathbf{s}^2 = 1$ and $\mathbf{E}\,\mathbf{s}^4 < \infty$.

In (8) the process N_t is a general counting process defined as

$$N_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{\sum_{l=1}^{k}\ \tau_l \leq t\}} \tag{10}$$

with $(\tau_l)_{l\geq 1}$ an i.i.d. sequence of positive integrated random variables with the distribution η and mean $\overline{\tau} = \mathbf{E}_Q\,\tau_1 > 0$. We assume that the processes $(N_t)_{t\geq 0}$, $(Y_i)_{i\geq 1}$ and $(L_t)_{t\geq 0}$ are independent. In the sequel we will use the renewal measure defined as

$$\overline{\eta} = \sum_{l=1}^{\infty} \eta^{(l)}\,, \tag{11}$$

where $\eta^{(l)}$ is the lth convolution power of the measure η.

Remark 1 Note that in the case when the random variables $(\zeta_j)_{j\geq 1}$ are i.i.d. random variables, then (8) is the semi-Markov process used in Barbu, Beltaief and Pergamenshchikov (2019a).

To use the renewal methods from Barbu, Beltaief and Pergamenshchikov (2019a) we assume that the distribution η has a density g for which the following conditions hold true.

$\mathbf{H}_1$) *Assume that, for any* $x \in \mathbb{R}$, *there exist the finite limits* $g(x-) = \lim_{z\to x-} g(z)$ *and* $g(x+) = \lim_{z\to x+} g(z)$ *and, for any* $\forall K > 0$, $\exists \delta = \delta(K) > 0$ *for which*

$$\sup_{|x|\leq K} \int_0^{\delta} \frac{|g(x+t) + g(x-t) - g(x+) - g(x-)|}{t} \mathrm{d}t \,<\, \infty.$$

$\mathbf{H}_2$) $\forall \gamma > 0$ *the upper bound* $\sup_{z\geq 0}\, z^{\gamma}|2g(z) - g(z-) - g(z+)| \,<\, \infty$.

$\mathbf{H}_3$) *There exists* $\beta > 0$ *such that* $\int_{\mathbb{R}_+}\, e^{\beta x}\, g(x)\,\mathrm{d}x < \infty$.

$\mathbf{H}_4$) $\exists t^* > 0$ *such that the Fourier transformation* $\widehat{g}(\theta - it)$ *belongs to* $\mathcal{L}_1(\mathbb{R})$ *for any* $0 \leq t \leq t^*$, *where* $\widehat{g}(z) = (2\pi)^{-1} \int_{\mathbb{R}}\, e^{izv} g(v)\mathrm{d}v$.

Moreover, to check these conditions we will use the following assumption.

$\mathbf{H}_4^*$) *The density* g *is two time continuously differentiable on* $\mathbb{R}_+$ *with* $g(0) = 0$ *and there exists* $\beta > 0$ *such that* $\int_0^{+\infty}\, e^{\beta x} \left(g(x) + |g^{'}(x)| + |g^{''}(x)|\right) \mathrm{d}x < \infty$ *and* $\lim_{x\to\infty}\, e^{\beta x} \left(g(x) + |g^{'}(x)|\right) = 0$.

It is clear that the conditions $\mathbf{H}_1$)–$\mathbf{H}_3$) hold true in this case. To obtain the condition $\mathbf{H}_4$) it suffices to calculate the integral in $\widehat{g}$, integrating by parts two times. For example, one can take gamma distribution of order $m \geq 2$

$$g(x) = \frac{\mathbf{a}^m x^{m-1}}{m!}\, e^{-\mathbf{a}x} \mathbf{1}_{\{x \geq 0\}} \quad \text{and} \quad \mathbf{a} > 0\,. \tag{12}$$

It should be noted that in view of Proposition 5.2 from Barbu, Beltaief and Pergamenshchikov (2019a), Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) imply that the renewal measure (11) has a continuous density ρ such that

$$\|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)|\, \mathrm{d}x < \infty \quad \text{and} \quad \Upsilon(x) = \rho(x) - \frac{1}{\overline{\tau}}\,. \tag{13}$$

Remark 2 It should be noted that Condition $\mathbf{H}_4$) does not hold for the exponential random variable $(\tau_j)_{j\geq 1}$ since its density is not continuous in zero. But for exponential random variables, i.e. in the case when $(N_t)_{t\geq 0}$ is a Poisson process, the renewal density can be calculated directly, i.e. $\rho(x) \equiv 1/\overline{\tau}$ and $\Upsilon \equiv 0$.

Now we describe the class of possible admissible noise distributions used in the robust risk (5). To this end we set

$$\sigma_Q = \varrho_1^2 + \varrho_2^2 + \frac{\varrho_3^2}{\overline{\tau}}\,. \tag{14}$$

As to the parameters in (6), we assume that

$$\varsigma_* \leq \sigma_Q \leq \varsigma^*\,, \tag{15}$$

where the unknown bounds $0 < \varsigma_* \leq \varsigma^*$ can be functions of T, i.e. $\varsigma_* = \varsigma_*(T)$ and $\varsigma^* = \varsigma^*(T)$, such that for any $\mathbf{b} > 0$

$$\lim_{T\to\infty} T^{\mathbf{b}}\, \varsigma_*(T) = +\infty \quad \text{and} \quad \lim_{T\to\infty} \frac{\varsigma^*(T)}{T^{\mathbf{b}}} = 0\,. \tag{16}$$

We denote by $\mathcal{Q}_T$ the family of all distributions of the process (6) in $\mathcal{D}[0,T]$ satisfying the properties (15) – (16).

Remark 3 As we will see later, the parameter (14) is the limit of the Fourier transform of the noise process (6). This limit is called variance proxy (see Konev and Pergamenshchikov (2012)).

3 Truncated fractional Poisson processes

As an example of the process (10) satisfies the conditions $\mathbf{H}_1$) – $\mathbf{H}_4$) we give the truncated fractional Poisson process introduced in Barbu, Beltaief and Pergamenshchikov (2019b). To this end, we remind the definition of the fractional Poisson process (see, for example, Biard and Saussereaur (2014); Laskin (2003)). The process (10) is called fractional Poisson process if the i.i.d. random variables (τ_j) have the Mittag-Leffler distribution which, for some $\mathbf{a} > 0$, is defined as

$$\mathbf{P}(\tau_1 > t) = \mathcal{E}_H(-\mathbf{a}t^H)\,, \tag{17}$$

where $0 < H \leq 1$ is called the Hurst index,

$$\mathcal{E}_H(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1+Hk)} \quad \text{and} \quad \Gamma(x) = \int_0^{+\infty} t^{x-1}\, e^{-t}\, \mathrm{d}t\,.$$

Note that, if $H = 1$, then we obtain the exponential distribution with parameter $\mathbf{a} > 0$ and, therefore, the process (10) is a Poisson process. If $0 < H < 1$, then the density of the distribution (17) (see, for example, Repin and Saichev (2000)) can be represented as

$$f_H(t) = \frac{\mathbf{a}\,\sin(\pi H)}{\pi} \int_0^{+\infty} \frac{x^H\, e^{-tx}}{x^{2H} + \mathbf{a}^2 + 2\mathbf{a}x^H \cos(\pi H)} \mathrm{d}x\,. \tag{18}$$

Form here we can directly obtain that

$$f_H(t) \sim t^{H-1}\,, \quad f_H^{'}(t) \sim t^{H-2}\,, \quad f_H^{''}(t) \sim t^{H-3} \quad \text{as} \quad t \to 0 \tag{19}$$

and

$$f_H(t) \sim t^{-H-1}\,, \quad f_H^{'}(t) \sim t^{-H-2}\,, \quad f_H^{''}(t) \sim t^{-H-3} \quad \text{as} \quad t \to \infty\,. \tag{20}$$

In particular, this implies that the Mittag-Leffler distribution has a heavy tail, i.e.

$$\mathbf{P}(\tau_1 > t) \quad \sim \quad t^{-H} \quad \text{as} \quad t \to \infty\,, \tag{21}$$

i.e. $\mathbf{E}\tau_1 = +\infty$. Therefore, the condition $\mathbf{H}_3$) does not hold for the distribution (17). To correct this effect, in Barbu, Beltaief and Pergamenshchikov (2019b) it is proposed to replace the Mittag-Leffler random variables in (10) with i.i.d. random variables distributed as $\tau_1^* = \min(X_*^{\mathbf{b}}, X^*)$, where X_* is a Mittag-Leffler with $0 < H < 1$, $0 < \mathbf{b} \leq H/3$ and X^* is a positive random variable satisfying the condition $\mathbf{H}_4^*$). Such processes are called truncated Poisson processes. Using the asymptotic properties (19) and (20) one can check directly that the random variable τ_1^* satisfies the condition $\mathbf{H}_4^*$) and, therefore, the conditions $\mathbf{H}_1$) – $\mathbf{H}_4$) hold true for this case.

Remark 4 It should be noted also that the process (10) with the Mittag-Leffler random variables has a "memory" in its increments (see, for example, Maheshwari and Vellaisamy (2016)) in the sense that, for any $\delta > 0$ and $s > 0$, the correlation coefficient

$$\mathrm{Corr}\left(\left(N_{s+\delta} - N_s\right),\left(N_{t+\delta} - N_t\right)\right) \quad \sim \quad t^{-\frac{3-H}{2}} \quad \text{as} \quad t \to \infty\,.$$

It should be noted that this property is very important for many practical problems and allows essentially to expand the possible applications of statistical results. Unfortunately, we can't use directly the fractional Poisson process in the regression model (3) since the impulse noise of the fractional Poisson processes will be very rare, since the time between jumps is not integrable, i.e. very large and, therefore, they have almost negligible influence in the observation models. On the contrary, the truncated process has an exponential moment, i.e. the same property as Poisson processes, and, moreover, it keeps a dependence on large time intervals.

4 Model selection

In this section we construct a model selection procedure for estimating the unknown function S given in (3) starting from the discrete-time observations (2) and we establish the oracle inequality for the associated risk. To this end, note that for any function $f : [0,T] \to \mathbb{R}$ from $\mathcal{L}_2[0,T]$, the integral

$$I_T(f) = \int_0^T f(s)\mathrm{d}\xi_s \tag{22}$$

is well defined, with $\mathbf{E}_Q\, I_T(f) = 0$. Moreover, as it is shown in Lemma 1 under the conditions $\mathbf{H}_1$)–$\mathbf{H}_4$),

$$\mathbf{E}_Q\, I_T^2(f) \le \varkappa_Q \int_0^T f_s^2 \mathrm{d}\, s \quad \text{and} \quad \varkappa_Q = \varrho_1^2 + \varrho_2^2 + \varrho_3^2\,|\rho|_* \tag{23}$$

where $|\rho|_* = \sup_{t\ge 0}|\rho(t)| < \infty$.
In this paper we will use the trigonometric basis $(\phi_j)_{j\ge 1}$ in $\mathbf{L}_2[0,1]$ defined as

$$\phi_1 = 1\,, \quad \phi_j(x) = \sqrt{2}\mathrm{Tr}_j(2\pi[j/2]x)\,, \quad j \ge 2\,, \tag{24}$$

where the function $\mathrm{Tr}_j(x) = \cos(x)$ for even j and $\mathrm{Tr}_j(x) = \sin(x)$ for odd j, $[x]$ denotes the integer part of x. Note, that these functions are orthonormal on the points $(t_j)_{1\le j\le p}$, i.e. for any $1 \le i, j \le p$

$$(\phi_i,\phi_j)_p = \frac{1}{p}\sum_{l=1}^{p}\phi_i(t_l)\phi_j(t_l) = \mathbf{1}_{\{i=j\}}\,. \tag{25}$$

In the sequel we denote by $\|x\|_p^2 = (x,x)_p$. Now note that, for any $1 \leq l \leq p$,

$$S(t_l) = \sum_{j=1}^{p} \theta_{j,p}\, \phi_j(t_l) \quad \text{and} \quad \theta_{j,p} = (S, \phi_j)_p \,. \tag{26}$$

Using the approach from Konev and Pergamenshchikov (2015), we estimate the Fourier coefficients $\theta_{j,p}$ as

$$\widehat{\theta}_{j,p} = \frac{1}{T}\int_0^T \psi_{j,p}(t)\mathrm{d}\, y_t, \quad \text{and} \quad \psi_{j,p}(t) = \sum_{l=1}^{n} \phi_j(t_l) 1_{\{t_{l-1}<t\leq t_l\}} \,. \tag{27}$$

It is clear that the functions $(\psi_{j,p})_{1\leq j\leq p}$ are orthonormal in $\mathcal{L}_2[0,1]$, i.e.

$$(\psi_{j,p}, \psi_{i,p}) = \int_0^1 \psi_{j,p}(t)\psi_{i,p}(t)\mathrm{d}\, t \ = (\phi_j, \phi_i)_p = \mathbf{1}_{\{i=j\}} \,. \tag{28}$$

The Fourier coefficients of S in the basis can be represented as

$$\overline{\theta}_{j,p} = (S, \psi_{i,p}) = \int_0^1 S(t)\psi_{i,p}(t)\mathrm{d}\, t \ = \theta_{j,p} + h_{j,p}, \tag{29}$$

where $h_{j,p}(S) = \sum_{l=1}^{p} \int_{t_{l-1}}^{t_l} \phi_j(t_l)(S(t) - S(t_l))\mathrm{d}\, t$. Therefore, (27) implies

$$\widehat{\theta}_{j,p} = \overline{\theta}_{j,p} + \frac{1}{\sqrt{T}}\xi_{j,p} \quad \text{and} \quad \xi_{j,p} = \frac{1}{\sqrt{T}} I_T(\psi_{j,p}) \,. \tag{30}$$

As in Barbu, Beltaief and Pergamenshchikov (2019a) we use the model selection procedures based on the following weighted least squares estimators

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{p} \lambda(j)\widehat{\theta}_{j,p}\psi_{j,p}(t) \,, \quad 0 \leq t \leq 1 \,, \tag{31}$$

where the weight vector $\lambda = (\lambda(1), \ldots, \lambda(p))'$ belongs to some finite set Λ from $[0,1]^p$. Here the prime $'$ denotes the transposition. Moreover, we set

$$\mathbf{m}_* = \mathrm{card}(\Lambda) \quad \text{and} \quad \Lambda_* = \max_{\lambda\in\Lambda} \sum_{j=1}^{p} \mathbf{1}_{\{\lambda(j)>0\}} \,, \tag{32}$$

where $\mathrm{card}(\Lambda)$ is the cardinal number of the set Λ. We assume that $\Lambda_* \leq n$. Now we use the same criteria as in Barbu, Beltaief and Pergamenshchikov (2019a) to chose a weight vector in Λ, i.e.we minimize the empirical error

$$\mathrm{Err}(\lambda) = \|\widehat{S}_\lambda - S\|^2 \,, \tag{33}$$

which can be represented as

$$\mathrm{Err}(\lambda) = \sum_{j=1}^{p} \lambda^2(j)\widehat{\theta}_{j,p}^2 - 2\sum_{j=1}^{p} \lambda(j)\widehat{\theta}_{j,p}\overline{\theta}_{j,p} + \|S\|^2 \,. \tag{34}$$

Note that the Fourier coefficients $(\theta_j)_{j\geq 1}$ are unknown. Therefore, using the approach from Barbu, Beltaief and Pergamenshchikov (2019a) to minimize this function we replace the terms $\widehat{\theta}_{j,p}\overline{\theta}_{j,p}$ by their estimators

$$\widetilde{\theta}_{j,p} = \widehat{\theta}^2_{j,p} - \frac{\sigma_Q}{T}\,,$$

where the proxy variance σ_Q is defined in (15). In the case when this variance is unknown we use its estimator, i.e.

$$\widetilde{\theta}_{j,p} = \widehat{\theta}^2_{j,p} - \frac{\widehat{\sigma}_T}{T} \quad \text{and} \quad \widehat{\sigma}_T = \frac{T}{p}\sum_{j=[\sqrt{T}]}^{p} \widehat{\theta}^2_{j,p}\,. \tag{35}$$

Now, using this estimator we define the penalty term as

$$\widehat{P}_T(\lambda) = \frac{\widehat{\sigma}_T|\lambda|^2}{T} \quad \text{and} \quad |\lambda|^2 = \sum_{j=1}^{p}\lambda^2(j)\,. \tag{36}$$

In the case, when the variance σ_Q is known we set

$$P_T(\lambda) = \frac{\sigma_Q|\lambda|^2}{T}\,. \tag{37}$$

Finally, we define the cost function as

$$J_T(\lambda) = \sum_{j=1}^{p}\lambda^2(j)\widehat{\theta}^2_{j,T} - 2\sum_{j=1}^{p}\lambda(j)\widetilde{\theta}_{j,T} + \delta\,\widehat{P}_T(\lambda), \tag{38}$$

where $\delta > 0$ is some threshold which will be specified later. Now we set the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}} \quad \text{and} \quad \widehat{\lambda} = \operatorname{argmin}_{\lambda\in\Lambda} J_T(\lambda)\,. \tag{39}$$

In the case when $\widehat{\lambda}$ is not unique we take one of them.

5 Main results

5.1 Oracle inequalities

Firstly, we obtain the non asymptotic oracle inequality for the model selection procedure (39). To this end we need a condition for the observations frequency.

H$_5$) *Assume that the frequency p is a function of T, i.e. $p = p_T$, such that*

$$\liminf_{T\to\infty}\frac{p_T}{T^{5/6}} > 0 \quad \text{and} \quad \limsup_{T\to\infty}\frac{p_T}{T} < \infty\,. \tag{40}$$

Theorem 1 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold true. Then, there exists some constant* $\mathbf{c}^* > 0$ *such that for any* $T \geq 1$ *and any noise distribution* $Q \in \mathcal{Q}_T$ *and* $0 < \delta \leq 1/6$, *the procedure* (39) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \mathbf{c}^* \frac{\mathbf{m}_*}{\delta T} \left(1 + \sigma_Q^2 + \Lambda_* \, \mathbf{E}_Q |\widehat{\sigma}_T - \sigma_Q|\right) . \tag{41}$$

In the case when σ_Q is known the inequality (41) has the following form.

Corollary 1 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold true and that the proxy variance* σ_Q *is known. Then there exists some constant* $\mathbf{c}^* > 0$ *such that for any* $T \geq 1$ *and for any noise distribution* $Q \in \mathcal{Q}_T$ *and* $0 < \delta \leq 1/6$, *the procedure* (39) *with* $\widehat{\sigma}_T = \sigma_Q$, *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \mathbf{c}^* \frac{(1+\sigma_Q^2)\mathbf{m}_*}{\delta T} .$$

Now we study the estimator $\widehat{\sigma}_T$ defined in (35).

Proposition 1 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$) – $\mathbf{H}_5$) *hold true for the model* (3) *and that* $S(\cdot)$ *is continuously differentiable. Then, there exists a constant* $\mathbf{c}^* > 0$ *such that for any* $T \geq 2$, $Q \in \mathcal{Q}_T$ *and* $p > \sqrt{T}$,

$$\mathbf{E}_{Q,S} |\widehat{\sigma}_T - \sigma_Q| \leq \mathbf{c}^* \, (1 + |\dot{S}|^2)(1+\sigma_Q)^2 \, \mathbf{g}^*_{T,p} , \tag{42}$$

where $\mathbf{g}^*_{T,p} = \sqrt{T}/p + 1/\sqrt{p}$.

Now Theorem 1 and this proposition imply directly the following result.

Theorem 2 *Assume that the function* S *is continuously differentiable and that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold true. Then there exists some constant* $\mathbf{c}^* > 0$ *such that for any continuously differentiable function* S *for any* $T \geq 2$, *for any noise distribution* $Q \in \mathcal{Q}_T$, $p > \sqrt{T}$ *and* $0 < \delta \leq 1/6$,

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \mathbf{c}^* \frac{\mathbf{m}_*}{\delta T} (1+\sigma_Q)^2 \left(1 + |\dot{S}|^2\right) \left(1 + \Lambda_* \mathbf{g}^*_{T,p}\right) .$$

To study robust properties of the procedure (39) we need a condition for weights.

$\mathbf{H}_6$) *The parameters* $\mathbf{m}_*$ *and* Λ_* *defined in* (32) *can be functions of* T, *i.e.* $\mathbf{m}_* = \mathbf{m}_*(T)$ *and* $\Lambda_* = \Lambda_*(T)$, *such that for any* $\mathbf{b} > 0$ $\lim_{T\to\infty} T^{-\mathbf{b}} \mathbf{m}_*(T) = 0$ *and* $\lim_{T\to\infty} T^{-1/3-\mathbf{b}} \Lambda_*(T) = 0$.

Now, Theorem 2 implies the following oracle inequality.

Theorem 3 *Assume that the function S is continuously differentiable, the conditions the conditions $\mathbf{C}_1$) – $\mathbf{C}_3$) and $\mathbf{H}_1$)–$\mathbf{H}_6$) hold true. Then, the procedure (39) for any $T \geq 2$, $p > \sqrt{T}$ and $0 < \delta < 1/6$ satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{U}_T^*}{T\delta}\,,$$

where the term $\mathbf{U}_T^* > 0$ is such that for any $r > 0$ and $\mathbf{b} > 0$,

$$\lim_{T\to\infty} \sup_{\|\dot{S}\| \leq r} T^{-\mathbf{b}} \mathbf{U}_T^* = 0\,. \tag{43}$$

In order to obtain the efficiency property, we specify the weight coefficients in the procedure (39). Consider, for some fixed $0 < \varepsilon < 1$, a numerical grid of the form

$$\mathcal{A} = \{1, \ldots, k^*\} \times \{\varepsilon, \ldots, m\varepsilon\}\,, \quad m = [1/\varepsilon^2]\,, \tag{44}$$

where $k^* \geq 1$ and ε are functions of T, i.e. $k^* = k^*(T)$ and $\varepsilon = \varepsilon(T)$, such that

$$\begin{cases} \lim_{T\to\infty} k^*(T) = +\infty\,, \quad \lim_{T\to\infty} \dfrac{k^*(T)}{\ln T} = 0\,, \\ \lim_{T\to\infty} \varepsilon(T) = 0 \quad \text{and} \quad \lim_{T\to\infty} T^{\mathbf{b}} \varepsilon(T) = +\infty \end{cases} \tag{45}$$

for any $\mathbf{b} > 0$. One can take, for example, for $T \geq 2$

$$\varepsilon(T) = \frac{1}{\ln T} \quad \text{and} \quad k^*(T) = k_0^* + \sqrt{\ln T}\,,$$

where $k_0^* \geq 0$ is a fixed constant. For each $\alpha = (k, r) \in \mathcal{A}$, we set the vector

$$\lambda_\alpha = (\lambda_\alpha(j))_{1 \leq j \leq p}$$

through its components which are defined as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j < \ln T\}} + \left(1 - (j/\omega_\alpha)^k\right) \mathbf{1}_{\{\ln T \leq j \leq \omega_\alpha\}}\,,$$

where

$$\omega_\alpha = \left(\frac{(k+1)(2k+1)}{\pi^{2k} k}\, r \upsilon_T\right)^{1/(2k+1)}\,, \quad \upsilon_T = T/\varsigma^*$$

and ς^* is introduced in (15). Now we define the set Λ as

$$\Lambda = \{\lambda_\alpha\,,\, \alpha \in \mathcal{A}\}\,. \tag{46}$$

These weight coefficients are used in Konev Pergamenshchikov (2009a); Konev and Pergamenshchikov (2012, 2015) for continuous time regression models to show the asymptotic efficiency. Note also that in this case the cardinal of the set Λ is $\mathbf{m}_* = k^* m$. Moreover, taking into account that for $k \geq 1$ the coefficient $\omega_\alpha < (r\upsilon_T)^{1/(2k+1)}$, we obtain that the norm of the set Λ defined in (32) can be bounded as $\Lambda_* \leq \sup_{\alpha\in\mathcal{A}} \omega_\alpha \leq (\upsilon_T/\varepsilon)^{1/3}$. Therefore, the properties (45) imply the condition $\mathbf{H}_6$).

5.2 Robust asymptotic efficiency

Now we study the asymptotic efficiency properties for the procedure (39), (46) with respect to the robust risks (5) defined by the distribution family (15) – (16). To this end, we assume that the unknown function S in the model (3) belongs to the Sobolev ball

$$\mathcal{W}_{\mathbf{r},\mathbf{k}} = \left\{ f \in \mathcal{C}^{\mathbf{k}}_{per}[0,1] \, : \, \sum_{j=0}^{\mathbf{k}} \|f^{(j)}\|^2 \leq \mathbf{r} \right\} , \tag{47}$$

where $\mathbf{r} > 0, k \geq 1$ are some unknown parameters, $\mathcal{C}^{\mathbf{k}}_{per}[0,1]$ is the set of $\mathbf{k}$ times continuously differentiable functions $f : [0,1] \to \mathbb{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \leq i \leq \mathbf{k}$. Note, that the class (47) is an ellipsoid, i.e.

$$\mathcal{W}_{\mathbf{r},\mathbf{k}} = \left\{ f = \sum_{j \geq 1} \theta_j \phi_j \, : \, \sum_{j=1}^{\infty} a_j \, \theta_j^2 \, \leq \mathbf{r} \right\} \tag{48}$$

where $a_j = \sum_{i=0}^{\mathbf{k}} (2\pi[j/2])^{2i}$. Similarly to Barbu, Beltaief and Pergamenshchikov (2019a) we will show here that the asymptotic sharp lower bound for the normalized robust risk (5) is given by the well-known Pinsker constant defined as

$$\mathbf{l}_* = \mathbf{l}_*(\mathbf{r}) = \, ((2\mathbf{k}+1)\mathbf{r})^{1/(2\mathbf{k}+1)} \left(\frac{\mathbf{k}}{(\mathbf{k}+1)\pi} \right)^{2\mathbf{k}/(2\mathbf{k}+1)} . \tag{49}$$

To study efficient properties we need to use the set Ξ_T of all possible estimators $\widehat{S}_T$ measurable with respect to the sigma-algebra $\sigma\{y_t \, , \, 0 \leq t \leq T\}$.

Theorem 4 *For the risk* (5) *with the coefficient rate* $\upsilon_T = T/\varsigma^*$

$$\liminf_{T \to \infty} \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)} \inf_{\widehat{S}_T \in \Xi_T} \sup_{S \in \mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}^*_T(\widehat{S}_T, S) \geq \mathbf{l}_* \, . \tag{50}$$

Note that, if the radius $\mathbf{r}$ and the regularity $\mathbf{k}$ are known, i.e. for the non-adaptive estimation problem on the continuous observations $(y_t)_{0 \leq t \leq T}$, in Barbu, Beltaief and Pergamenshchikov (2019a) it is proposed to use the estimate $\widehat{S}_{\lambda_0}$ defined in (31) with the weights (46)

$$\lambda_0 = \lambda_{\alpha_0} \, , \quad \alpha_0 = (\mathbf{k}, r_0) \quad \text{and} \quad r_0 = [\mathbf{r}/\varepsilon]\varepsilon \, . \tag{51}$$

Now, we show the same result for the discrete observations (2).

Proposition 2 *Assume that the conditions the conditions* $\mathbf{C}_1$) – $\mathbf{C}_2$) *and* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold true. Then*

$$\lim_{T \to \infty} \upsilon_T^{2k/(2k+1)} \sup_{S \in \mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}^*_T(\widehat{S}_{\lambda_0}, S) \leq \mathbf{l}_* \, . \tag{52}$$

For the adaptive estimation we user the model selection procedure (39) with the parameter δ defined as a function of T, i.e. $\delta = \delta_T$, such that

$$\lim_{T\longrightarrow\infty} \delta_T = 0 \quad \text{and} \quad \lim_{T\longrightarrow\infty} T^{\mathbf{b}}\, \delta_T = +\infty \tag{53}$$

for any $\mathbf{b} > 0$. For example, we can take $\delta_T = (6 + \ln T)^{-1}$.

Theorem 5 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_6$) *hold true. Then the robust risk* (5) *for the procedure* (39) *with the coefficients* (46) *and the parameter* $\delta = \delta_T$ *satisfying* (53) *has the following upper bound*

$$\limsup_{T\to\infty} \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}_T^*(\widehat{S}_*, S) \le \mathbf{l}_* \,.$$

Theorem 4 and Theorem 5 imply the following result.

Theorem 6 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_6$) *hold true. Then the procedure* (39) *with the weight coefficients* (46) *and the parameter* $\delta = \delta_T$ *satisfying* (53) *is asymptotically efficient, i.e.*

$$\lim_{T\to\infty} \frac{\inf_{\widehat{S}_T\in\Xi_T} \sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}_T^*(\widehat{S}_T, S)}{\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}_T^*(\widehat{S}_*, S)} = 1$$

and

$$\lim_{T\to\infty} \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \mathcal{R}_T^*(\widehat{S}_*, S) = \mathbf{l}_* \,.$$

Remark 5 It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball $\mathcal{W}_{\mathbf{r},\mathbf{k}}$ is $T^{2\mathbf{k}/(2\mathbf{k}+1)}$ (see, for example, Pinsker (1981); Konev Pergamenshchikov (2009b)). We see here that the efficient robust rate is $\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}$, i.e. if the distribution upper bound $\varsigma^* \to 0$ as $T \to \infty$ we obtain a faster rate with respect to $T^{2\mathbf{k}/(2\mathbf{k}+1)}$, and if $\varsigma^* \to \infty$ as $T \to \infty$ we obtain a slower rate. In the case when ς^* is constant the robust rate is the same as the classical non robust convergence rate.

5.3 Big data analysis for the model (1)

Now we consider the estimation problem for the parameters $(\beta_j)_{1\le j\le q}$ in (3) with unknown q. In this case we have to estimate the sequence $\beta = (\beta_j)_{j\ge1}$ in which $\beta_j = 0$ for $j \ge q+1$. To this end we assume that the functions $(\mathbf{u}_j)_{j\ge1}$ are orthonormal in $\mathcal{L}_2[0,1]$, i.e. $(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{1}_{\{i\ne j\}}$. Indeed, we can use always the Gram-Schmidt orthogonalization procedure to provide this property. Thus, in this case we estimate the parameters $\beta = (\beta_j)_{j\ge1}$ through the estimator (39) as $\widehat{\beta}_* = (\widehat{\beta}_{*,j})_{j\ge1}$ and $\widehat{\beta}_{*,j} = (\mathbf{u}_j, \widehat{S}_*)$. Similarly, using the weighted estimators (31) we define the basic estimators $(\widehat{\beta}_\lambda)_{\lambda\in\Lambda}$ as $\widehat{\beta}_\lambda = (\widehat{\beta}_{j,\lambda})_{j\ge1}$ and $\widehat{\beta}_{j,\lambda} = (\mathbf{u}_j, \widehat{S}_\lambda)$. Taking into account that in this case $|\widehat{\beta}_* - \beta|^2 = \sum_{j=1}^\infty (\widehat{\beta}_{*,j} - \beta_j)^2 = \|\widehat{S}_* - S\|^2$ and $|\widehat{\beta}_\lambda - \beta|^2 = \|\widehat{S}_\lambda - S\|^2$, Theorem 3 implies the following oracle inequality.

Theorem 7 *Assume that the function* (3) *is continuously differentiable and the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$), $\mathbf{H}_1$)–$\mathbf{H}_5$) *and* (15)–(16) *hold true. Then, for any* $n \geq 1$ *and* $0 < \delta < 1/6$, *the following oracle inequality holds true*

$$\sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,S}|\widehat{\beta}_* - \beta|^2 \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,S}|\widehat{\beta}_\lambda - \beta|^2 + \frac{\mathbf{U}_T^*}{T\delta}\,,$$

where the term $\mathbf{U}_T^* > 0$ *satisfies the property* (43).

Moreover, Theorem 6 implies the following efficiency property.

Theorem 8 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_6$) *hold true. Then the estimator* $\widehat{\beta}_*$ *constructed through the procedure* (39) *with the weight coefficients* (46) *and the parameter* $\delta = \delta_T$ *satisfying* (53) *is asymptotically efficient in the minimax sense, i.e.*

$$\lim_{T\to\infty} \frac{\inf_{\widehat{\beta}_T} \sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,S}|\widehat{\beta}_T - \beta|^2}{\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,S}|\widehat{\beta}_* - \beta|^2} = 1 \tag{54}$$

and

$$\lim_{T\to\infty} \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)} \sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}} \sup_{Q\in\mathcal{Q}_T} \mathbf{E}_{Q,S}|\widehat{\beta}_* - \beta|^2 = \mathbf{l}_*\,,$$

where the infimum is taken over all possible estimators $\widehat{\beta}_T$ *measurable with respect the field* $\sigma\{y_t\,,\, 0 \leq t \leq T\}$ *and the lower bound* $\mathbf{l}_*$ *is defined in* (49).

Remark 6 It should be emphasized that the efficiency properties (54) are obtained without sparse conditions on the number of non zero parameters β_j in the model (1) (see, for example, in Hasttie, Friedman and Tibshirani (2008)). Moreover, we do not use even the parameter dimension q which can be equal to $+\infty$.

6 Stochastic calculus for generalized semi-Markov processes

In this section we study some properties of the stochastic integrals (22). First, note that using the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) and the stochastic calculus developed in Barbu, Beltaief and Pergamenshchikov (2019a) for semi-Markov processes we can show the following Lemmas 1 and 2.

Lemma 1 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_2$) *and* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any non random functions* f *and* h *from* $\mathcal{L}_2[0,T]$

$$\mathbf{E}_Q\, I_t(f) I_t(h) = (\varrho_1^2 + \varrho_2^2)\,(f,h)_t \,+ \varrho_3^2\,(f, h\rho)_t\,, \tag{55}$$

where $(f,h)_t = \int_0^t f(s)\,h(s)\mathrm{d}s$ *and* ρ *is the density of the renewal measure* (11).

It should be noted that this lemma implies directly that the stochastic integral (22) satisfies the properties (23).

Lemma 2 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_2$) *and* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any bounded* $[0,\infty)\to\mathbb{R}$ *functions* f *and* h *and for any* $k\geq 1$,

$$\mathbf{E}_Q\left(I_{\mathbf{t}_k-}(f)\,I_{\mathbf{t}_k-}(h)\mid\mathcal{G}\right)=(\varrho_1^2+\varrho_2^2)(f\,,\,h)_{\mathbf{t}_k}+\varrho_3^2\sum_{l=1}^{k-1}f(\mathbf{t}_l)\,h(\mathbf{t}_l),$$

where $\mathbf{t}_k=\sum_{j=1}^{k}\tau_j$ *and* $\mathcal{G}=\sigma\{\mathbf{t}_l\,,\,l\geq 1\}$.

Lemma 3 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any nonrandom bounded* $[0,T]\to\mathbb{R}$ *functions* f *and* h, *the expectation* $\mathbf{E}_Q\int_0^T I_{t-}^2(f)I_{t-}(h)h(t)\mathrm{d}\xi_t=0$.

Proof. Setting $\check{L}_t=\varrho_1 w_t+\varrho_2 L_t$, we can represent the integral (22) as

$$I_t(f)=\check{I}_t(f)+\varrho_3 I_t^z(f)\,, \tag{56}$$

where $\check{I}_t(f)=\int_0^t f(u)\mathrm{d}\check{L}_u$ and $I_t^z(f)=\int_0^t f(u)\mathrm{d}z_u$. Note here, that using the condition (7) and the inequality for martingales from Novikov (1975) we can obtain that $\mathbf{E}_Q\sup_{0\leq t\leq T}\check{I}_t^8(f)<\infty$. Since $\check{L}_t$ and z_t are independent, we get $\mathbf{E}_Q\int_0^T I_{t-}^2(f)I_{t-}(h)h(t)d\check{L}_t=0$. Moreover, the conditions $\mathbf{C}_1$) – $\mathbf{C}_3$) yield, that for any non random $(c_{i,j})$ and $k\geq 1$ $\mathbf{E}\left(\sum_{j=1}^{k-1}c_{1,j}\zeta_j\right)^2\zeta_k=0$ and $\mathbf{E}\left(\sum_{j=1}^{k-1}c_{1,j}\zeta_j\right)^2\left(\sum_{j=1}^{k-1}c_{2,j}\zeta_j\right)\zeta_k=0$. Therefore, taking into account that the sequence $(\zeta_k)_{k\geq 1}$ does not depend on the moments $(\mathbf{t}_k)_{k\geq 1}$ and the process $(\check{L}_t)_{t\geq 0}$, and using the same method as in the proof of Lemma 8.4 from Barbu, Beltaief and Pergamenshchikov (2019a) we obtain

$$\mathbf{E}_Q\int_0^T I_{t-}^2(f)I_{t-}(h)h(t)\mathrm{d}z_t=\mathbf{E}_Q\sum_{k\geq 1}\mathbf{1}_{\{\mathbf{t}_k\leq T\}}\,I_{\mathbf{t}_k-}^2(f)I_{\mathbf{t}_k-}(h)h(\mathbf{t}_k)\zeta_k=0\,.$$

This implies Lemma 3. □

Now we study the integrals defined in (64) as functions of f.

Proposition 3 *Assume that the conditions* $\mathbf{C}_1$) – $\mathbf{C}_3$) *and* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any* $[0,\infty)\to\mathbb{R}$ *functions* f,h *such that* $|f|_*\leq 1$ *and* $|h|_*\leq 1$, *one has*

$$|\mathbf{E}_Q\widetilde{I}_T(f)\widetilde{I}_T(h)|\leq 12\sigma_Q^2(1+\overline{\tau})^2\left((f,h)_T^2+T\widetilde{\mathbf{c}}\right)\,, \tag{57}$$

where $\widetilde{\mathbf{c}}=(2+\Pi(x^4)+2|\rho|_*)(1+\|\Upsilon\|_1^2)$ *and* $|f|_*=\sup_{t\geq 0}|f(t)|$.

Proof. First of all, note that in view of the Ito formula and using the fact that for the process (6) the jumps $\Delta z_s \Delta L_s = 0$ a.s. for any $s \geq 0$, we obtain that

$$\begin{aligned} \mathrm{d}I_t^2(f) &= 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2 f^2(t)\mathrm{d}\,t \\ &\quad + \varrho_2^2 \mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta L_s)^2 + \varrho_3^2 \mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta z_s)^2\,. \end{aligned}$$

Note also that Lemma 1 yields $\mathbf{E}_Q I_t^2(f) = (\varrho_1^2+\varrho_2^2)\|f\|_t^2 + \varrho_3^2\|f\sqrt{\rho}\|_t^2$ with $\|f\|_t^2 = \int_0^t f^2(t)\mathrm{d}t$. Therefore,

$$\mathrm{d}\widetilde{I}_t(f) = 2I_{t-}(f)f(t)d\xi_t + f^2(t)\mathrm{d}\widetilde{m}_t\,, \quad \widetilde{m}_t = \varrho_2^2\check{m}_t + \varrho_3^2 m_t\,,$$

where $\check{m}_t = \sum_{0\leq s\leq t}(\Delta L_s)^2 - t$ and $m_t = \sum_{0\leq s\leq t}(\Delta z_s)^2 - \int_0^t \rho(s)\mathrm{d}s$. Thus,

$$\mathbf{E}_Q\widetilde{I}_T(f)\widetilde{I}_T(h) = \mathbf{E}_Q\int_0^T \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(h) + \mathbf{E}_Q\int_0^T \widetilde{I}_{t-}(h)\mathrm{d}\widetilde{I}_t(f) + \mathbf{E}_Q\,[\widetilde{I}(f),\widetilde{I}(h)\,]_T\,.$$

Using here Lemma 3 and, taking into account that $(\check{m}_t)_{t\geq 0}$ is a square integrated martingale, we get

$$\mathbf{E}_Q\int_0^T \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(h) = \mathbf{E}_Q\int_0^T \widetilde{I}_{t-}(f)h^2(t)\mathrm{d}\widetilde{m}_t = \rho_3^2\mathbf{E}_Q\int_0^T I_{t-}^2(f)h^2(t)\mathrm{d}m_t\,.$$

The last integral can be represented as

$$\mathbf{E}_Q\int_0^T I_{t-}^2(f)h^2(t)\mathrm{d}m_t = J_1 - J_2\,, \tag{58}$$

where $J_1 = \mathbf{E}_Q\sum_{k\geq 1} I^2_{\mathbf{t}_k-}(f)h^2(\mathbf{t}_k)\mathbf{1}_{\{\mathbf{t}_k\leq T\}}$ and $J_2 = \int_0^T \mathbf{E}_Q\, I_t^2(f)h^2(t)\rho(t)\mathrm{d}t$. By Lemma 2 we get

$$J_1 = \mathbf{E}_Q\sum_{k\geq 1}\mathbf{E}_Q\left(I^2_{\mathbf{t}_k-}(f)|\mathcal{G}\right)h^2(\mathbf{t}_k)\mathbf{1}_{\{\mathbf{t}_k\leq T\}} = (\varrho_1^2+\varrho_2^2)J_{1,1} + \varrho_3^2 J_{1,2}\,,$$

where $J_{1,1} = \mathbf{E}_Q\sum_{k\geq 1}\|f\|^2_{\mathbf{t}_k}h^2(\mathbf{t}_k)\mathbf{1}_{\{\mathbf{t}_k\leq T\}} = \int_0^T \|f\|_t^2 h^2(t)\rho(t)\mathrm{d}t$ and

$$\begin{aligned} J_{1,2} &= \mathbf{E}_Q\sum_{k\geq 1}\sum_{l=1}^{k-1} f^2(\mathbf{t}_l)\,h^2(\mathbf{t}_k)\mathbf{1}_{\{\mathbf{t}_k\leq T\}} = \mathbf{E}_Q\sum_{l\geq 1} f^2(\mathbf{t}_l)\sum_{k\geq l+1} h^2(\mathbf{t}_k)\mathbf{1}_{\{\mathbf{t}_k\leq T\}} \\ &= \int_0^T f^2(x)\left(\int_0^{T-x} h^2(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x\,. \end{aligned}$$

Moreover, using Lemma 1 for the last term in (58), we obtain that

$$J_2 = (\varrho_1^2+\varrho_2^2)\int_0^T \|f\|_t^2 h^2(t)\rho(t)\mathrm{d}t + \varrho_3^2\int_0^T \|f\sqrt{\rho}\|_t^2 h^2(t)\rho(t)\mathrm{d}t$$

and we can represent the expectation in (58) as

$$\mathbf{E}_Q \int_0^T I^2_{t-}(f)h^2(t)\mathrm{d}m_t = \varrho_3^2 \int_0^T f^2(x) \left(\int_x^T h^2(t)(\Upsilon(t-x)-\Upsilon(t))\mathrm{d}t \right) \rho(x)\mathrm{d}x\,,$$

i.e. we can estimate this as $|\mathbf{E}_Q \int_0^T I^2_{t-}(f)h^2(t)\mathrm{d}m_t| \le 2\varrho_3^2 T\|\Upsilon\|_1$. Therefore,

$$|\mathbf{E}_Q \int_0^T \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(h)| + |\mathbf{E}_Q \int_0^T \widetilde{I}_{t-}(h)\mathrm{d}\widetilde{I}_t(f)| \le 4\varrho_3^4 T\|\Upsilon\|_1\,. \tag{59}$$

Furthermore, note that

$$[\widetilde{I}(f),\widetilde{I}(h)]_T = < \widetilde{I}^c(f),\widetilde{I}^c(h) >_T + \mathbf{D}_T(f,h)\,,$$

where $\widetilde{I}^c_t(f) = 2\varrho_1 \int_0^t I_s(f)f(s)\mathrm{d}w_s$ and $\mathbf{D}_T(f,h) = \sum_{0\le t\le T} \Delta\widetilde{I}^d_t(f)\Delta\widetilde{I}^d_t(h)$. In this case $\widetilde{I}^d_t(f) = 2\int_0^t I_{s-}(f)f(s)\mathrm{d}\xi^d_s + \int_0^t f^2(s)\mathrm{d}\widetilde{m}_s$ and $\xi^d_t = \varrho_2 L_t + \varrho_3 z_t$. Therefore, in view of Lemma 1,

$$\begin{aligned}
\mathbf{E}_Q < \widetilde{I}^c(f),\widetilde{I}^c(h) >_T &= 4\rho_1^2 \int_0^T \mathbf{E}_Q(I_t(f)I_t(h))f(t)h(t)\mathrm{d}t \\
&= 4\rho_1^2(\rho_1^2+\rho_2^2)\int_0^T (f,h)_t\, f(t)h(t)\mathrm{d}t + 4\rho_1^2\rho_3^2\int_0^T (f,h\rho)_t f(t)h(t)\mathrm{d}t \\
&= 4\rho_1^2\sigma_Q\,(f,h)_T^2 + 4\rho_1^2\rho_3^2\int_0^T (f,g\Upsilon)_t f(t)h(t)\mathrm{d}t\,.
\end{aligned}$$

Since $|f|_* \le 1$ and $|h|_* \le 1$, we get $\int_0^T |(f,h\Upsilon)_t f(t)h(t)|\mathrm{d}t \le T\|\Upsilon\|_1$ and

$$\left|\mathbf{E}_Q < \widetilde{I}^c(f),\widetilde{I}^c(h) >_T\right| \le 4\sigma_Q^2\left((f,h)_T^2 + T\overline{\tau}\|\Upsilon\|_1\right)\,. \tag{60}$$

To study the process $\mathbf{D}_T(f,h)$ note that $\Delta\xi^d_t\Delta\widetilde{m}_t = \varrho_2^3(\Delta L_t)^3 + \varrho_3^3(\Delta z_t)^3$. Note also that for any $t \ge 0$ the expectation $\mathbf{E}_Q I_t(f) = 0$. Therefore, using the definition of the process L_t we obtain through the Fubini theorem, that for any bounded $[0,T] \to \mathbb{R}$ measurable non random functions V

$$\mathbf{E}_Q \sum_{0\le t\le T} V(t)\, I_{t-}(f)(\Delta L_t)^3 = \mathit{\Pi}(x^3)\int_0^T V(t)\mathbf{E}_Q\, I_t(f)\mathrm{d}t = 0\,.$$

Moreover, since the processes $(\check{L}_t)_{t\ge0}$ and $(z_t)_{t\ge0}$ are independent we get

$$\mathbf{E}_Q \sum_{0\le t\le T} V(t)\check{I}_{t-}(f)(\Delta z_t)^3 = \mathbf{E}_Q \sum_{k\ge1} V(\mathbf{t}_k)\zeta_k^3 \mathbf{E}_Q\left(\check{I}_{\mathbf{t}_k-}(f)|\mathcal{G}^z\right) = 0\,,$$

where the integral $\check{I}_t(f)$ is defined in (56) and the $\mathcal{G}^z = \sigma\{z_t\,,\, t\ge0\}$. Note that the condition $\mathbf{C}_3$) implies that for any $k \ge 1$ and nonrandom $(c_j)_{j\ge1}$ $\mathbf{E}_Q\left(\sum_{j=1}^{k-1} c_j\zeta_j\right)\zeta_k^3 = 0$. Therefore, $\mathbf{E}_Q \sum_{0\le t\le T} V(t)I^z_{t-}(f)(\Delta z_t)^3 = 0$ and

$$\mathbf{E}_Q \sum_{0\le t\le T} I_{t-}(f)f(t)h^2(t)\Delta\xi^d_t\,\Delta\widetilde{m}_t = \mathbf{E}_Q \sum_{0\le t\le T} I_{t-}(h)h(t)f^2(t)\Delta\xi^d_t\,\Delta\widetilde{m}_t = 0\,.$$

So, the expectation of $\mathbf{D}_T(f,h)$ can be represented as

$$\mathbf{E}_Q\,\mathbf{D}_T(f,h) = 4\varrho_2^2\mathbf{E}_Q\,\mathbf{D}_{1,T}(f,h) + 4\varrho_3^2\mathbf{E}_Q\,\mathbf{D}_{2,T}(f,h) + \mathbf{E}_Q\,\mathbf{D}_{3,T}(f,h)\,,$$

where $\mathbf{D}_{1,T}(f,h) = \sum_{0\le t\le T} I_{t-}(f)I_{t-}(h)f(t)h(t)(\Delta L_t)^2$,

$$\mathbf{D}_{2,T}(f,h) = \sum_{0\le t\le T} I_{t-}(f)I_{t-}(h)f(t)h(t)(\Delta z_t)^2$$

and $\mathbf{D}_{3,T}(f,h) = \sum_{0\le t\le T} f^2(t)\,h^2(t)(\Delta\widetilde{m}_t)^2$. First, since $\mathit{\Pi}(x^2)=1$, we get

$$\mathbf{E}_Q\,\mathbf{D}_{1,T}(f,h) = \int_0^T f(t)h(t)\mathbf{E}_Q\,I_t(f)I_t(h)\,\mathrm{d}t = (\varrho_1^2+\varrho_2^2)\int_0^T f(t)h(t)\,(f,h)_t\,\mathrm{d}t$$

$$+\,\varrho_3^2\int_0^T f(t)h(t)\,(f,h\rho)_t\,\mathrm{d}t = \sigma_Q(f,h)_T^2 + \varrho_3^2\int_0^T f(t)h(t)\,(f,h\Upsilon)_t\,\mathrm{d}t$$

and $|\mathbf{E}_Q\,\mathbf{D}_{1,T}(f,h)| \le \sigma_Q\left((f,h)_T^2 + T\overline{\tau}\|\Upsilon\|_1\right)$. Then, taking into account that $\mathbf{E}\,\zeta_j^2 = 1$ and using Lemma 2, we represent the expectation for $\mathbf{D}_{2,T}(f,h)$ as

$$\begin{aligned}\mathbf{E}_Q\,\mathbf{D}_{2,T}(f,h) &= \mathbf{E}\sum_{k\ge1}\mathbf{E}_Q\left(I_{\mathbf{t}_k-}(f)I_{\mathbf{t}_k-}(h)|\mathcal{G}\right)f(\mathbf{t}_k)h(\mathbf{t}_k)\,\mathbf{1}_{\{\mathbf{t}_k\le T\}}\\ &= (\varrho_1^2+\varrho_2^2)\mathbf{E}_Q\sum_{k\ge1}(f\,,h)_{\mathbf{t}_k}f(\mathbf{t}_k)h(\mathbf{t}_k)\,\mathbf{1}_{\{\mathbf{t}_k\le T\}} + \varrho_3^2\,\mathbf{E}_Q\,\mathbf{D}_{2,T}^{'}(f,h)\\ &= (\varrho_1^2+\varrho_2^2)\int_0^T(f,h)_t\,f(t)h(t)\rho(t)\mathrm{d}t + \varrho_3^2\mathbf{E}_Q\,\mathbf{D}_{2,T}^{'}(f,h)\,,\end{aligned}$$

where $\mathbf{D}_{2,T}^{'}(f,h) = \sum_{k\ge1}\sum_{l=1}^{k-1} f(\mathbf{t}_l)\,h(\mathbf{t}_l)f(\mathbf{t}_k)h(\mathbf{t}_k)\,\mathbf{1}_{\{\mathbf{t}_k\le T\}}$. Note, that

$$\int_0^T(f,h)_t\,f(t)g(t)\rho(t)\mathrm{d}t = \frac{1}{2\overline{\tau}}(f,h)_T^2 + \int_0^T(f,h)_t\,f(t)g(t)\Upsilon(t)\mathrm{d}t\,,$$

i.e.

$$\left|\int_0^T(f,h)_t\,f(t)g(t)\rho(t)\mathrm{d}t\right| \le \frac{1}{2\overline{\tau}}(f,h)_T^2 + T\|\Upsilon\|_1\,.$$

Furthermore, the expectation of $\mathbf{D}_{2,T}^{'}(f,h)$ can be represented as

$$\begin{aligned}\mathbf{E}_Q\,\mathbf{D}_{2,T}^{'}(f,h) &= \mathbf{E}_Q\sum_{l\ge1}f(\mathbf{t}_l)\,h(\mathbf{t}_l)\sum_{k\ge l+1}f(\mathbf{t}_k)h(\mathbf{t}_k)\,\mathbf{1}_{\{\mathbf{t}_k\le T\}}\\ &= \int_0^T f(x)g(x)\left(\int_0^{T-x}f(x+t)h(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x\\ &= \frac{1}{2\overline{\tau}^2}(f,h)_T^2 + \mathbf{D}_{2,T}^{''}(f,h)\,,\end{aligned}$$

where

$$\mathbf{D}^{''}_{2,T}(f,h) = \int_0^T f(x)h(x)\left(\int_0^{T-x} f(x+t)h(x+t)\Upsilon(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x + \frac{1}{\overline{\tau}}\int_0^T f(x)h(x)\left(\int_0^{T-x} f(x+t)h(x+t)\Upsilon(t)\mathrm{d}t\right)\Upsilon(x)\mathrm{d}x\,.$$

This implies $|\mathbf{D}^{''}_{2,T}(f,h)| \le T(|\rho|_* + 1/\overline{\tau})(1+\|\Upsilon\|_1^2)$ and, therefore,

$$|\mathbf{E}_Q\,\mathbf{D}_{2,T}(f,h)| \le \sigma_Q\left(\frac{(f,h)_T^2}{\overline{\tau}} + T(2+\overline{\tau}|\rho|_*)(1+\|\Upsilon\|_1^2)\right).$$

Moreover, we can calculate directly

$$\mathbf{E}_Q\,\mathbf{D}_{3,T}(f,h) = \varrho_2^4 \Pi(x^4)\int_0^T f^2(t)\,h^2(t)\mathrm{d}t + \varrho_3^4\int_0^T f^2(t)\,h^2(t)\rho(t)\mathrm{d}t$$

and $\mathbf{E}_Q\,\mathbf{D}_{3,T}(f,h) \le T\sigma_Q^2\left(\Pi(x^4)+|\rho|_*\right)(1+\overline{\tau})^2$. From here we obtain that $|\mathbf{E}_Q\,\mathbf{D}_T(f,h)| \le \sigma_Q^2(1+\overline{\tau})^2\left(4(f,h)_T^2 + T\widetilde{\mathbf{c}}\right)$, where $\widetilde{\mathbf{c}}$ is given in (57). From this and (60) we find $\mathbf{E}_Q\left[\widetilde{I}(f),\widetilde{I}(h)\right]_T \le 8\sigma_Q^2(1+\overline{\tau})^2\left((f,h)_T^2 + T\widetilde{\mathbf{c}}\right)$. This bound and (59) implies (57). Hence the proof is achieved. □

In order to prove the oracle inequalities we need to study the conditions introduced in Konev and Pergamenshchikov (2012) for the general semimartingale model (3). To this end, we set for any $x \in \mathbb{R}^p$ the functions

$$B_{1,Q}(x) = \sum_{j=1}^p x_j\left(\mathbf{E}_Q\xi_{j,p}^2 - \sigma_Q\right) \quad \text{and} \quad B_{2,Q}(x) = \sum_{j=1}^p x_j\,\widetilde{\xi}_{j,p}\,, \tag{61}$$

where σ_Q is defined in (15) and $\widetilde{\xi}_{j,p} = \xi_{j,p}^2 - \mathbf{E}_Q\xi_{j,p}^2$.

Proposition 4 *Assume that the conditions* $\mathbf{C}_1$)–$\mathbf{C}_2$), $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold true. Then there exists some constant* $\mathbf{c}^* > 0$ *such that for any* $Q \in \cup_{k\ge1}\mathcal{Q}_k$

$$\mathbf{L}_{1,Q} = \sup_{T\ge3}\sup_{x\in[-1,1]^p}\left|B_{1,Q}(x)\right| < \mathbf{c}^*\,\sigma_Q \tag{62}$$

and

$$\mathbf{L}_{2,Q} = \sup_{T\ge3}\sup_{|x|\le1}\mathbf{E}_Q\,B_{2,Q}^2(x) \le \mathbf{c}^*\,\sigma_Q^2\,, \tag{63}$$

where $|x|^2 = \sum_{j=1}^p x_j^2$.

Proof. Firstly, using here Lemma 1, we obtain that

$$\mathbf{E}\xi_{j,p}^2 = \varrho_1^2 + \varrho_2^2 + \frac{\varrho_3^2}{T}\int_0^T \psi_{j,p}^2(x)\,\rho(x)\mathrm{d}\,x = \sigma_Q + \frac{\varrho_3^2}{T}\int_0^T \psi_{j,p}^2(x)\,\Upsilon(x)\mathrm{d}\,x\,.$$

From (13) it follows that $\left|\mathbf{E}_Q\xi_{j,p}^2 - \sigma_Q\right| \leq 2\varrho_3^2\|\Upsilon\|_1/T$ and, therefore, the condition $\mathbf{H}_5$) implies the inequality (62). Next, note that

$$\mathbf{E}_Q\left(\sum_{j=2}^{p} x_j\widetilde{\xi}_{j,p}\right)^2 \leq \frac{1}{T^2}\sum_{j=1}^{p}\sum_{l=1}^{p}|x_j|\,|x_l|\left|\mathbf{E}_Q\,\widetilde{I}_T(\psi_{j,p})\widetilde{I}_T(\psi_{l,p})\right|, \tag{64}$$

where $\widetilde{I}_T(f) = I_T^2(f) - \mathbf{E}_Q I_T^2(f)$. Now Proposition 3 and the property (28) imply, that for some constant $\mathbf{c}^* > 0$ and for $|x| \leq 1$

$$\mathbf{E}_Q\left(\sum_{j=2}^{p} x_j\widetilde{\xi}_{j,p}\right)^2 \leq \mathbf{c}^*\left(|x|^2 + \frac{1}{T}\left(\sum_{j=1}^{p}|x_j|\right)^2\right) \leq c^*\left(1+\frac{p}{T}\right).$$

The condition $\mathbf{H}_5$) implies directly Proposition 4. □

7 Proofs

7.1 Proof of Theorem 1

Using the cost function given in (38), we can rewrite the empirical squared error in (34) as follows

$$\mathrm{Err}(\lambda) = J_T(\lambda) + 2\sum_{j=1}^{p}\lambda(j)\vartheta_{j,p} + \|S\|^2 - \delta\widehat{P}_T(\lambda), \tag{65}$$

where

$$\vartheta_{j,p} = \widetilde{\theta}_{j,p} - \overline{\theta}_{j,p}\widehat{\theta}_{j,p} = \frac{1}{\sqrt{T}}\overline{\theta}_{j,p}\xi_{j,p} + \frac{1}{T}\widetilde{\xi}_{j,p} + \frac{1}{T}\varsigma_{j,T} + \frac{\sigma_Q - \widehat{\sigma}_T}{T},$$

with $\varsigma_{j,p} = \mathbf{E}_Q\xi_{j,p}^2 - \sigma_Q$ and $\widetilde{\xi}_{j,p} = \xi_{j,p}^2 - \mathbf{E}_Q\xi_{j,p}^2$. Setting

$$\mathbf{M}(\lambda) = \frac{1}{\sqrt{T}}\sum_{j=1}^{p}\lambda(j)\theta_j\xi_{j,p} \quad \text{and} \quad L(\lambda) = \sum_{j=1}^{p}\lambda(j) \tag{66}$$

and using the functions (61) through the penalty term (37), we rewrite (65) as

$$\begin{aligned}\mathrm{Err}(\lambda) = J_T(\lambda) + 2\frac{\sigma_Q - \widehat{\sigma}_T}{T}L(\lambda) + 2M(\lambda) + \frac{2}{T}B_{1,Q}(\lambda)\\ + 2\sqrt{P_T(\lambda)}\frac{B_{2,Q}(\nu(\lambda))}{\sqrt{\sigma_Q T}} + \|S\|^2 - \delta\widehat{P}_T(\lambda),\end{aligned} \tag{67}$$

where $\nu(\lambda) = \lambda/|\lambda|$. Let $\lambda_0 = (\lambda_0(j))_{1\le j\le p}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be defined as in (39). Substituting λ_0 and $\widehat{\lambda}$ in (67), we obtain

$$\mathrm{Err}(\widehat{\lambda}) - \mathrm{Err}(\lambda_0) = J_T(\widehat{\lambda}) - J_T(\lambda_0) + 2\frac{\sigma_Q - \widehat{\sigma}_T}{T} L(\varpi) + \frac{2}{T} B_{1,Q}(\varpi) + 2M(\varpi)$$
$$+ 2\sqrt{P_T(\widehat{\lambda})}\frac{B_{2,Q}(\widehat{\nu})}{\sqrt{\sigma_Q T}} - 2\sqrt{P_T(\lambda_0)}\frac{B_{2,Q}(\nu_0)}{\sqrt{\sigma_Q T}} - \delta\widehat{P}_T(\widehat{\lambda}) + \delta\widehat{P}_T(\lambda_0),$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{\nu} = \nu(\widehat{\lambda})$ and $\nu_0 = \nu(\lambda_0)$. Now, in view of the inequality $2|ab| \le \delta a^2 + \delta^{-1}b^2$ we get that

$$2\sqrt{P_T(\lambda)}\frac{|B_{2,Q}(\nu(\lambda))|}{\sqrt{\sigma_Q T}} \le \delta P_T(\lambda) + \frac{B^2_{2,Q}(\nu(\lambda))}{\delta\sigma_Q T}.$$

Then, taking into account that $|L(\varpi)| \le L(\widehat{\lambda}) + L(\lambda) \le 2\Lambda_*$ and using the definition (62) we get

$$\mathrm{Err}(\widehat{\lambda}) \le \mathrm{Err}(\lambda_0) + 2\mathbf{M}(\varpi) + \frac{2\mathbf{L}_{1,Q}}{T} + \frac{2B^*_{2,Q}}{\delta\sigma_Q T} + \frac{4\Lambda_*|\widehat{\sigma}_T - \sigma_Q|}{T} + 2\delta\widehat{P}_T(\lambda_0)\,,$$

where $B^*_{2,Q} = \sup_{\lambda\in\Lambda} B^2_{2,Q}((\nu(\lambda))$. To estimate the second term in the right side of this inequality we set $S_x = \sum_{j=1}^p x(j)\overline{\theta}_{j,p}\psi_{j,p}$ for $x = (x(j))_{1\le j\le p} \in \mathbb{R}^p$. Thanks to (23) we estimate the term $\mathbf{M}(x)$ for any $x \in \mathbb{R}^p$ as

$$\mathbf{E}_Q\mathbf{M}^2(x) \le \varkappa_Q \frac{1}{T}\sum_{j=1}^p x^2(j)\overline{\theta}^2_{j,p} = \varkappa_Q \frac{\|S_x\|^2}{T}\,. \tag{68}$$

Setting here $Z^* = \sup_{x\in\Lambda_1} TM^2(x)/\|S_x\|^2$ and $\Lambda_1 = \Lambda - \lambda_0$, we get

$$2|\mathbf{M}(x)| \le \delta\|S_x\|^2 + \frac{Z^*}{T\delta}. \tag{69}$$

The last term here can be estimated from above as

$$\mathbf{E}_Q Z^* \le \sum_{x\in\Lambda_1} \frac{T\mathbf{E}_Q\mathbf{M}^2(x)}{\|S_x\|^2} \le \sum_{x\in\Lambda_1} \varkappa_Q = \varkappa_Q \mathbf{m}_*\,,$$

where $\mathbf{m}_* = \mathrm{card}(\Lambda)$. Moreover, note that, for any $x \in \Lambda_1$,

$$\|S_x\|^2 - \|\widehat{S}_x\|^2 = \sum_{j=1}^p x^2(j)(\overline{\theta}^2_{j,p} - \widehat{\theta}^2_{j,p}) \le -2\mathbf{M}_1(x), \tag{70}$$

where $\mathbf{M}_1(x) = T^{-1/2}\sum_{j=1}^p x^2(j)\overline{\theta}^2_{j,p}\xi_{j,T}$. Taking into account now that, for any $x \in \Lambda_1$, the components $|x(j)| \le 1$, we can estimate this term as in (68), i.e. $\mathbf{E}_Q\,\mathbf{M}^2_1(x) \le \varkappa_Q\,\|S_x\|^2/T$. Similarly to the previous reasoning setting

$Z_1^* = \sup_{x\in\Lambda_1} T\mathbf{M}_1^2(x)/\|S_x\|^2$, we get $\mathbf{E}_Q\, Z_1^* \leq \varkappa_Q\, \mathbf{m}_*$. Using the same type of arguments as in (69), we can derive

$$2|\mathbf{M}_1(x)| \leq \delta\|S_x\|^2 + \frac{Z_1^*}{T\delta}. \tag{71}$$

From here and (70), we get

$$\|S_x\|^2 \leq \frac{\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z_1^*}{T\delta(1-\delta)} \tag{72}$$

for any $0 < \delta < 1$. Using this bound in (69) yields

$$2M(x) \leq \frac{\delta\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z^* + Z_1^*}{T\delta(1-\delta)}\,.$$

Taking into account that $|\widehat{S}_\varpi|^2 \leq 2\,(\mathrm{Err}(\widehat{\lambda}) + \mathrm{Err}(\lambda_0))$, we obtain

$$2M(\varpi) \leq \frac{2\delta(\mathrm{Err}(\widehat{\lambda}) + \mathrm{Err}(\lambda_0))}{1-\delta} + \frac{Z^* + Z_1^*}{T\delta(1-\delta)}$$

and, therefore,

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \leq& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{Z^* + Z_1^*}{T\delta(1-3\delta)} + \frac{2\mathbf{L}_{1,Q}}{T(1-3\delta)} + \frac{2B_{2,Q}^*}{\delta(1-3\delta)\sigma_Q T}\\ &+ \frac{(4\Lambda_* + 2)}{T(1-3\delta)}|\widehat{\sigma}_T - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_T(\lambda_0).\end{aligned}$$

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

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \leq& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{2(Z^* + Z_1^*)}{T\delta} + \frac{4\mathbf{L}_{1,Q}}{T} + \frac{4B_{2,Q}^*}{\delta\sigma_Q T}\\ &+ \frac{(8\Lambda_* + 2)|\widehat{\sigma}_T - \sigma_Q|}{T} + \frac{2\delta}{(1-3\delta)}\,P_T(\lambda_0).\end{aligned}$$

Now, using the upper bound (63), we get

$$\mathbf{E}_Q\, B_{2,Q}^* \leq \sum_{\lambda\in\Lambda} \mathbf{E}_Q B_{2,Q}^2(\nu(\lambda)) \leq \mathbf{m}_*\mathbf{L}_{2,Q}\,.$$

Now, taking into account that $\Lambda_* \geq 1$, we get

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \leq& \frac{1+\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\mathbf{m}_*}{T\delta} + \frac{4\mathbf{L}_{1,Q}}{T} + \frac{4\mathbf{m}_*\mathbf{L}_{2,Q}}{\delta\sigma_Q T}\\ &+ \frac{10\Lambda_*\mathbf{E}_Q\,|\widehat{\sigma}_T - \sigma_Q|}{T} + \frac{2\delta}{(1-3\delta)}P_T(\lambda_0).\end{aligned}$$

By using the upper bound for $P_T(\lambda_0)$ in Lemma 4, we obtain that

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q \mathbf{m}_*}{T\delta} + \frac{4\mathbf{L}_{1,Q}}{T} + \frac{4\mathbf{m}_*\mathbf{L}_{2,Q}}{\delta\sigma_Q T} + \frac{10\Lambda_*}{T}\,\mathbf{E}_Q\,|\widehat{\sigma}_T - \sigma_Q| + \frac{2\delta\mathbf{L}_{1,Q}}{(1-3\delta)T}\,.$$

Taking into account here that $1-3\delta \geq 1/2$ for $0 < \delta < 1/6$ and, moreover, that $\varkappa_Q \leq (1+\overline{\tau}|\rho|_*)\sigma_Q$ and using the bounds (62) and (63) we obtain the inequality (41). Hence we obtain the desired result. □

7.2 Proof of Proposition 1

Let $x^{'} = (x_j^{'})_{1\leq j\leq p}$ with $x_j^{'} = \mathbf{1}_{\{[\sqrt{T}]\leqslant j\leqslant p\}}$. Then (30) and (35) yield

$$\widehat{\sigma}_T = \frac{T}{p}\sum_{j=l}^{p}(\overline{\theta}_{j,p})^2 + \frac{2T}{p}\mathbf{M}(x^{'}) + \frac{1}{p}\sum_{j=l}^{p}\xi_{j,p}^2\,, \tag{73}$$

where $\mathbf{M}$ is given in (66). Setting $x^{''} = (x_j^{''})_{1\leq j\leq p}$ and $x_j^{''} = p^{-1/2}\mathbf{1}_{\{[\sqrt{T}]\leqslant j\leqslant p\}}$, one can write the last term on the right hand side of (73) as

$$\frac{1}{p}\sum_{j=[\sqrt{T}]}^{p}\xi_{j,p}^2 = \frac{1}{\sqrt{p}}\,B_{2,Q}(x^{''}) + \frac{1}{p}B_{1,Q}(x^{'}) + \frac{(p-[\sqrt{T}]+1)\sigma_Q}{p}\,,$$

where the functions $B_{1,Q}$ and $B_{2,Q}$ are defined in (61). Using Proposition 4 and Lemma 6 , we come to the following upper bound

$$\mathbf{E}_Q|\widehat{\sigma}_T - \sigma_Q| \leq \frac{16|\dot{S}|^2T}{[\sqrt{T}]p} + \frac{2T}{p}\mathbf{E}_Q\,|\mathbf{M}(x^{'})| + \frac{\mathbf{L}_{1,Q}}{p} + \frac{\sqrt{\mathbf{L}_{2,Q}}}{\sqrt{p}} + \frac{\sigma_Q([\sqrt{T}]-1)}{p}\,.$$

In the same way as in (68) through Lemma 6, we obtain

$$\mathbf{E}_Q\,|M(x^{'})| \leq \left(\frac{\varkappa_Q}{T}\sum_{j=[\sqrt{T}]}^{p}\overline{\theta}_{j,p}^2\right)^{1/2} \leq \frac{4(\varkappa_Q|\dot{S}|^2)^{1/2}}{[\sqrt{T}]}\,.$$

Taking into account that $\kappa_Q \leq (1+\overline{\tau}|\rho|_*)\sigma_Q$ and using the bounds (62) and (63) we obtain the inequality (42). Hence Proposition 1 holds true. □

7.3 Proof of Theorem 2

This proof directly follows from Theorem 1 and Proposition 1. □

7.4 Proof of Theorem 4

First, we denote by Q_0 the distribution in $\mathcal{D}[0,n]$ of the noise (6) with the parameter $\varrho_1 = \varsigma^*$, $\varrho_2 = 0$ and $\varrho_3 = 0$, i.e., the distribution for the "signal + white noise" model. So, we can estimate from below the robust risk $\mathcal{R}^*_T(\widetilde{S}_T, S) \geq \mathcal{R}_{Q_0}(\widetilde{S}_T, S)$. Now, Theorem 6.1 from Konev Pergamenshchikov (2009b) yields the bound (50). Hence we obtain the desired result. □

7.5 Proof of Proposition 2

First, we note that in view of (31) one can represent the quadratic risk for the empiric norm $\|\cdot\|_p$ defined in (25) as

$$\mathbf{E}_Q \, \|\widehat{S}_{\lambda_0} - S\|_p^2 = \frac{1}{T} \sum_{j=1}^{p} \lambda_0^2(j) \, \mathbf{E}_Q \, \xi_{j,p}^2 + \overline{\Theta}_p \, ,$$

where $\overline{\Theta}_p = \sum_{j=1}^{p} \left(\theta_{j,p} - \lambda_0(j) \, \overline{\theta}_{j,p}\right)^2$. First, note that

$$\sup_{Q \in \mathcal{Q}_T} \mathbf{E}_Q \sum_{j=1}^{p} \lambda_0^2(j) \, \xi_{j,p}^2 \leq \varsigma_* \sum_{j=1}^{p} \lambda_0^2(j) + \mathbf{L}_{1,Q} \, .$$

where $\mathbf{L}^*_{1,T} = \sup_{Q \in \mathcal{Q}_T} \mathbf{L}_{1,Q}$. Taking into account that $\upsilon_T = T/\sigma^*$, we get

$$\sup_{Q \in \mathcal{Q}_T} \mathbf{E}_Q \, \|\widehat{S}_{\lambda_0} - S\|_p^2 \leq \frac{1}{\upsilon_T} \sum_{j=1}^{p} \lambda_0^2(j) + \frac{\mathbf{L}^*_{1,T}}{T} + \overline{\Theta}_p \, .$$

Note here that

$$\lim_{T \to \infty} \frac{\sum_{j=1}^{T} \lambda_0^2(j)}{\upsilon_T^{1/(2k+1)}} = \frac{2(\tau_k \, \mathbf{r})^{1/(2k+1)} \, k^2}{(k+1)(2k+1)} \, , \quad \tau_k = \frac{(k+1)(2k+1)}{\pi^{2k} k} \, . \tag{74}$$

Now, from (29) we obtain that for any $0 < \widetilde{\varepsilon} < 1$

$$\overline{\Theta}_p \leq (1 + \widetilde{\varepsilon}) \, \Theta_p + (1 + \widetilde{\varepsilon}^{-1}) \sum_{j=1}^{p} h_{j,p}^2 \, , \tag{75}$$

where $\Theta_p = \sum_{j=1}^{p} (1 - \lambda_0(j))^2 \, \theta_{j,p}^2$. Moreover, in view of the definition (51)

$$\Theta_p = \sum_{j=\iota_0}^{[\omega_0]} (1 - \lambda_0(j))^2 \, \theta_{j,p}^2 + \sum_{j=[\omega_0]+1}^{p} \theta_{j,p}^2 := \Theta_{1,p} + \Theta_{2,p} \, ,$$

where $\omega_0 = \omega_{\alpha_0} = (\tau_k r_0 \upsilon_T)^{1/(2k+1)}$ and $r_0 = [\mathbf{r}/\varepsilon] \, \varepsilon$. Lemma 8 yields

$$\Theta_{1,p} \leq (1 + \widetilde{\varepsilon}) \sum_{j=[\ln T]}^{[\omega_0]} (1 - \lambda_0(j))^2 \, \theta_j^2 + 4\pi^2 r (1 + \widetilde{\varepsilon}^{-1}) \, \omega_0^3 \, p^{-2} \, .$$

Through Lemma 7 we have $\Theta_{2,p} \le (1+\widetilde{\varepsilon})\sum_{j\ge[\omega_0]+1}\theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,\mathbf{r}\,p^{-2}$. Hence, $\Theta_p \le (1+\widetilde{\varepsilon})\,\Theta^* + (1+\widetilde{\varepsilon}^{-1})\left(4\pi^2 r\omega_0^3 + r\right)p^{-2}$, where the first term $\Theta^* = \sum_{j\ge\ln T}(1-\lambda_0(j))^2\,\theta_j^2$. Moreover, note that

$$\sup_{S\in\mathcal{W}_{\mathbf{r},1}}\max_{1\le j\le p} h_{j,p}^2 \le \|\dot{S}\|^2\,p^{-2} \le r\,p^{-2}\,.$$

Moreover, $\mathcal{W}_{\mathbf{r},\mathbf{k}} \subseteq \mathcal{W}_{\mathbf{r},2}$ for any $\mathbf{k}\ge 2$. From here and Lemma 9 we get

$$\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}}\sum_{j=1}^{p} h_{j,p}^2 \le r\left(p^{-1}\,\mathbf{1}_{\{\mathbf{k}=1\}} + 3p^{-2}\mathbf{1}_{\{\mathbf{k}\ge 2\}}\right)$$

and, therefore, in view of the condition $\mathbf{H}_5$)

$$\lim_{T\to\infty} \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\left(p^{-1}\mathbf{1}_{\{\mathbf{k}=1\}} + \omega_0^3 p^{-2}\mathbf{1}_{\{\mathbf{k}\ge 2\}}\right) = 0\,.$$

This implies, that

$$\limsup_{T\to\infty}\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}}\overline{\Theta}_p \le \limsup_{T\to\infty}\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}}\Theta^*\,.$$

To estimate the term Θ^* we set

$$\mathbf{U}_T = \upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{j\ge\ln T}(1-\lambda_0(j))^2/a_j\,,$$

where the sequence $(a_j)_{j\ge1}$ is defined in (48). This leads to the inequality

$$\sup_{S\in\mathcal{W}_{\mathbf{r},1}}\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\,\Theta^* \le \mathbf{U}_T\sum_{j\ge1}a_j\,\theta_j^2 \le \mathbf{U}_T\mathbf{r}\,.$$

Using $\lim_{T\to\infty} r_0 = \mathbf{r}$, we get $\limsup_{T\to\infty}\mathbf{U}_T \le \pi^{-2k}\,(\tau_k\,r)^{-2\mathbf{k}/(2\mathbf{k}+1)}$, where the coefficient τ_k is given in (74). This implies immediately that

$$\limsup_{T\to\infty}\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}}\overline{\Theta}_p \le \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2\mathbf{k}/(2\mathbf{k}+1)}}\,. \tag{76}$$

Therefore, from (74) and (76) it follows that

$$\lim_{T\to\infty}\upsilon_T^{2\mathbf{k}/(2\mathbf{k}+1)}\sup_{S\in\mathcal{W}_{\mathbf{r},\mathbf{k}}}\sup_{Q\in\mathcal{Q}_T}\mathbf{E}_Q\,\|\widehat{S}_{\lambda_0} - S\|_p^2 \le \mathbf{l}_*\,. \tag{77}$$

Using now Lemma 5 and the condition $\mathbf{H}_5$), we get the upper bound (52). Hence we obtain the desired result. □

8 Appendix

A.1 Property of the penalty term

Lemma 4 *For any $n \geq 1$ and $\lambda \in \Lambda$,*

$$P_T(\lambda) \leq \mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{\mathbf{L}_{1,Q}}{T},$$

where the coefficient $P_T(\lambda)$ is defined in (66) *and $\mathbf{L}_{1,Q}$ is defined in* (62).

Proof. From (30) and (33) we obtain

$$\mathrm{Err}(\lambda) \geq \sum_{j=1}^{p} \left(\lambda(j)\widehat{\theta}_{j,p} - \overline{\theta}_{j,p}\right)^2 = \sum_{j=1}^{p} \left((\lambda(j)-1)\overline{\theta}_{j,p} + \frac{\lambda(j)}{T}\xi_{j,p}\right)^2 .$$

Now Proposition 4 implies

$$\mathcal{R}_Q(\widehat{S}_\lambda, S) = \mathbf{E}_Q \,\mathrm{Err}(\lambda) \geq \frac{1}{T}\sum_{j=1}^{p} \lambda^2(j)\mathbf{E}_Q\,\xi_{j,p}^2 \geq P_T(\lambda) - \frac{\mathbf{L}_{1,Q}}{T}\,.$$

Hence we obtain the result. □

A.2 Properties of the Fourier coefficients

Lemma 5 *Let f be an absolutely continuous function, $f : [0,1] \to \mathbb{R}$, with $\|\dot{f}\| < \infty$ and g be a simple function, $g : [0,1] \to \mathbb{R}$ of the form $g(t) = \sum_{j=1}^{p} c_j\,\chi_{(t_{j-1},\mathbf{t}_j]}(t)$, where c_j are some constants. Then, for any $\varepsilon > 0$, the function $\Delta = f - g$ satisfies the following inequalities*

$$\|\Delta\|^2 \leq (1+\widetilde{\varepsilon})\|\Delta\|_p^2 + (1+\widetilde{\varepsilon}^{-1})\frac{\|\dot{f}\|^2}{p^2}\,, \quad \|\Delta\|_p^2 \leq (1+\widetilde{\varepsilon})\|\Delta\|^2 + (1+\widetilde{\varepsilon}^{-1})\frac{\|\dot{f}\|^2}{p^2}\,.$$

Lemma 6 *Let the function $S(t)$ in* (3) *be absolutly continuous and have an absolutely integrable derivative. Then the coefficients $(\overline{\theta}_{j,p})_{1\leqslant j\leqslant p}$ defined in* (29) *satisfy the inequalities $|\overline{\theta}_{1,p}| \leqslant \|S\|_1$ and $\max_{2\leqslant j\leqslant p} j|\overline{\theta}_{j,p}| \leqslant 2\sqrt{2}|\dot{S}|_1$.*

Lemma 7 *For any $p \geq 2$, $1 \leq N \leq p$ and $r > 0$, the coefficients $(\theta_{j,p})_{1\leq j\leq p}$ of functions S from the class $\mathcal{W}_{\mathbf{r},1}$ satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality $\sum_{j=N}^{p}\theta_{j,p}^2 \leq (1+\widetilde{\varepsilon})\sum_{j\geq N}\theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,r\,p^{-2}$.*

Lemma 8 *For any $p \geq 2$ and $r > 0$, the coefficients $(\theta_{j,p})_{1\leq j\leq p}$ of functions S satisfy the inequality $\max_{1\leq j\leq p}\sup_{S\in\mathcal{W}_{\mathbf{r},1}}\left(|\theta_{j,p}-\theta_j| - 2\pi\sqrt{r}\,j\,p^{-1}\right) \leq 0$.*

Lemma 9 *For any $p \geq 2$ and $r > 0$ the correction coefficients from* (29) *satisfy the inequality $\sup_{S\in\mathcal{W}_{\mathbf{r},2}}\sum_{j=1}^{p} h_{j,p}^2 \leq 3r\,p^{-2}$.*

Lemmas 5 – 9 are proven in Konev and Pergamenshchikov (2015).

References

Barbu, V., Beltaief, S. and Pergamenshchikov, S. M. (2019) Robust adaptive efficient estimation for semi - Markov nonparametric regression models. - *Statistical inference for stochastic processes*, **22**, 2, 187– 231

Barbu, V., Beltaief, S. and Pergamenshchikov, S. M. (2019) Robust statistical signal processing in semi-Markov nonparametric regression models. - *Les Annales de l'I.S.U.P.*, special issue, **63**, Fascicule 2-3, 45 – 56.

Beltaief, S., Chernoyarov O. and Pergamenshchikov, S. M. (2020) Model selection for the robust efficient signal processing observed with small Levy noise. *Ann. Inst. Statist. Math.*, **72**, 1205 - 1235.

Barbu, V. S. and N. Limnios, N. (2008) *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their use in Reliability and DNA Analysis.* Lecture Notes in Statistics, **191**, New York: Springer.

Barndorff-Nielsen, O.E. and Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics. *J. Royal Stat. Soc.*, **B 63**, 167–241.

Biard, R. and B. Saussereau, B. (2014) Fractional Poisson processes: long-range dependence and applications in ruin theory. - *J. Appl. Probab* **51**, 727 - 740

Fourdrinier, D. and Pergamenshchikov, S. M. (2007) Improved selection model method for the regression with dependent noise. *Ann. Inst. Statist. Math.*, **59**, 3, 435–464.

Fujimori, K. (2019) The Danzing selector for a linear model of diffusion processes. *Statistical Inference for Stochastic Processes*, **22**, p. 475 – 498.

Liptser, R.S. and Shiryaev A.N. (1989) *Theory of Martingales.* New-York: Springer.

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

Hasttie,T., Friedman, J. and Tibshirani, R. (2008) *The Elements of Statistical Leaning. Data Mining, Inference and Prediction.* Second Edition, Springer, Springer series in Statistics.

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

Konev, V. V. and Pergamenshchikov, S. M. (2009a) Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities. *Vestn. Tomsk. Gos. Univ. Mat. Mech.*, **3**, 7 , 23–41.

Konev, V. V. and Pergamenshchikov, S. M. (2009b) Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Vestn. Tomsk. Gos. Univ. Mat. Mech.*, **4**, 8, 31–45.

Konev, V. V. and Pergamenshchikov, S. M. (2012) Efficient robust nonparametric in a semimartingale regression model. *Ann. Inst. Henri Poincaré. Probab. Stat.*, **48**, 4, 1217–1244.

Konev, V. V. and Pergamenshchikov, S. M. (2015) Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic Process. Appl.*, **125**, 294–326.

Kutoyants, Yu. A. (1994) *Identification of dynamical systems with small noise.* Dordrecht: Kluwer Academic Publishers Group.

Laskin, N. (2003) Fractional Poisson processes. *Communications in Nonlinear Science and Numerical Simulation*, **8**, 201 – 213.

Maheshwari, A. and Vellaisamy, P. (2016) On the long - range dependence of fractional Poission processes. *J. Appl. Probab.*, **53** (4), 989 - 1000.

Middleton, D. (1979) Canonical non-Gaussian noise models: their implications for measurement and for prediction of receiver performance. *IEEE Trans. Electromagnetic Compatibility*, **21**, 209 – 220.

Novikov, A.A. (1975) On discontinuous martingales. - *Theory. Probab. Appl.*, **20** (1), 11 – 26.

Pinsker, M. S. (1981) Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission information*, **17**, 120–133.

Repin, O. N. and Saichev, A. I. (2000). Fractional Poisson law. - *Radiophys. and Quantum Electronics*, **43** (9), 738 – 741.