Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2012, Vol. 48, No. 4, 1217–1244
DOI: 10.1214/12-AIHP488

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Efficient robust nonparametric estimation in a semimartingale regression model[1]

Victor Konev[a] and Serguei Pergamenshchikov[b]

[a]*Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia. E-mail:* vvkonev@mail.tsu.ru
[b]*Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France and Department of Mathematics and Mechanics, Tomsk State University, Lenin str. 36, 634041 Tomsk, Russia. E-mail:* Serge.Pergamenchtchikov@univ-rouen.fr

Received 17 November 2010; revised 2 February 2012; accepted 12 March 2012

Abstract. The paper considers the problem of robust estimating a periodic function in a continuous time regression model with the dependent disturbances given by a general square integrable semimartingale with an unknown distribution. An example of such a noise is a non-Gaussian Ornstein–Uhlenbeck process with jumps (see (*J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2001) 167–241), (*Ann. Appl. Probab.* **18** (2008) 879–908)). An adaptive model selection procedure, based on the weighted least square estimates, is proposed. Under general moment conditions on the noise distribution, sharp non-asymptotic oracle inequalities for the robust risks have been derived and the robust efficiency of the model selection procedure has been shown. It is established that, in the case of the non-Gaussian Ornstein–Uhlenbeck noise, the sharp lower bound for the robust quadratic risk is determined by the limit value of the noise intensity at high frequencies. An example with a martinagale noise exhibits that the risk convergence rate becomes worse if the noise intensity is unbounded.

Résumé. Dans cette article nous considérons le problème d'estimation robuste d'une fonction périodique dans un modèle de régression en temps continu avec un bruit dépendant décrit par une semi martingale carrée intégrable de distribution inconnue. Un exemple de ce bruit est un processus d'Ornstein–Uhlenbeck non gaussien avec sauts (voir (*J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2001) 167–241), (*Ann. Appl. Probab.* **18** (2008) 879–908)). Nous proposons une procédure adaptative de sélection de modèle basée sur les estimateurs des moindres carrés pondérés. Sous des conditions générales sur les deux premiers moments de la distribution du bruit, des inégalités d'Oracle non asymptotiques pointues pour des risques quadratiques robustes sont obtenues et l'efficacité robuste est établie. Nous avons établi aussi que dans le cas du processus d'Ornstein–Uhlenbeck non Gaussian, la borne inférieure pour le risque quadratique robuste est donnée par la limite de l'intensité du bruit quand la fréquence tend vers l'infini. Nous donnons un exemple d'un modèle de régression avec un bruit martingale où la vitesse de convergence du risque quadratique devient plus lente si l'intensité du bruit tend vers l'infini.

MSC: Primary 62G08; secondary 62G05

Keywords: Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequality; Asymptotic efficiency

1. Introduction

Consider a regression model in continuous time

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

[1]Supported by the RFBR-Grant 09-01-00172-a.

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

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

is well defined and has the following properties

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

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

The problem is to estimate the unknown function S in the model (1.1) on the basis of observations $(y_t)_{0\le t\le n}$.

The class of the disturbances ξ satisfying conditions (1.3) is rather wide and comprises, in particular, the Lévy processes which are used in different applied problems (see [4,16], for details). The models (1.1) with the Lévy's type noise naturally arise (see [18]) in the nonparametric functional statistics problems (see, for example, [8]). Moreover, as is shown in Section 2, non-Gaussian Ornstein–Uhlenbeck-based models, introduced in [2], enter this class.

We define the error of an estimate $\widehat{S}$ (any real-valued function measurable with respect to $\sigma\{y_t, 0 \le t \le n\}$) for S by its integral quadratic risk

$$\mathcal{R}_Q(\widehat{S}, S) := \mathbf{E}_{Q,S} \|\widehat{S} - S\|^2, \tag{1.4}$$

where $\mathbf{E}_{Q,S}$ stands for the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1.1) with a fixed distribution Q of the noise $(\xi_t)_{0\le t\le n}$ and a given function S; $\|\cdot\|$ is the norm in $\mathbf{L}_2[0,1]$, i.e.

$$\|f\|^2 := \int_0^1 f^2(t)\,\mathrm{d}t. \tag{1.5}$$

Since in our case the noise distribution Q is unknown, it seems natural similar to [10] to measure the quality of an estimate $\widehat{S}$ by the robust risk defined as

$$\mathcal{R}_n^*(\widehat{S}, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widehat{S}, S) \tag{1.6}$$

which assumes taking supremum of the error (1.4) over the whole family of admissible distributions $\mathcal{Q}_n$.

We will treat the stated problem from the standpoint of the model selection approach. It will be noted that the origin of this method goes back to early seventies with the pioneering papers by Akaike [1] and Mallows [22] who proposed to introduce penalizing in a log-likelihood type criterion. The further progress has been made by Barron, Birgé and Massart [3,23], who developed a non-asymptotic model selection method which enables one to derive non-asymptotic oracle inequalities for nonparametric regression models with the i.i.d. Gaussian disturbances. An oracle inequality yields the upper bound for the estimate risk via the minimal risk corresponding to a chosen family of estimates. Galtchouk and Pergamenshchikov [9] applied the Barron–Birgé–Massart technic to the problem of estimating nonparametric drift functions in ergodic diffusion processes. Fourdrinier and Pergamenshchikov [7] extended the Barron–Birgé–Massart method to the models with the spherically symmetric dependent observations. They proposed a model selection procedure based on the improved least squares estimates. Lately, the authors [17] applied this method to the nonparametric problem of estimating a periodic function in a continuous time model with a Gaussian colored noise. In all cited papers, the non-asymptotic oracle inequalities have been derived, which enable one to establish the optimal convergence rate for the minimax risks. In addition to the optimal convergence rate, the other important problem is that of the efficiency of adaptive estimation procedures. In order to examine the efficiency property of a procedure one has to obtain the *sharp oracle inequalities*, i.e. such in which the factor at the principal term in the right-hand of the inequality is close to unity.

The first result on sharp inequalities is most likely due to Kneip [15] who studied a Gaussian regression model in discrete time. It will be observed that the derivation of oracle inequalities usually rests upon the fact that the initial model, by applying the Fourier transformation, can be reduced to a Gaussian model with independent observations. However, such a transform is possible only for Gaussian models with independent homogeneous observations or for the inhomogeneous ones with the known correlation characteristics. This restriction significantly narrows the area of application of the proposed model selection procedures and rules out a broad class of models including, in particular, heteroscedastic regression models widely used in econometrics (see, for example, [5,14]). For constructing adaptive procedures in the case of inhomogeneous observations one needs to modify the approach to the estimation problem. Galtchouk and Pergamenshchikov [11–13] have developed a new estimation method intended for the heteroscedastic regression models in discrete time. The heart of this method is to combine the Barron–Birgé–Massart non-asymptotic penalization method [3] and the Pinsker weighted least square method which minimizes the asymptotic risk (see, for example, [24,25]). This yields a significant improvement in the performance of the procedure (see numerical example in [11]).

The goal of this paper is to develop the robust efficient model selection method for the model (1.1) with dependent disturbances having unknown distribution. We follow the approach proposed by Galtchouk and Pergamenshchikov [11] in the construction of the procedure. Unfortunately, their method of obtaining the oracle inequalities is essentially based on the independence of observations and can not be applied here. This paper proposes the new analytical tools which allow one to obtain the sharp non-asymptotic oracle inequalities for robust risks under general conditions on the distribution of the noise in the model (1.1). This method enables us to treat both the cases of dependent and independent observations from the same standpoint, it does not assume the knowledge of the noise distribution and leads to the efficient estimation procedure with respect to the risk (1.6). The validity of the conditions, imposed on the noise in Eq. (1.1) is verified for a non-Gaussian Ornstein–Uhlenbeck process and for a martinagale with the increasing variance (see Section 2).

The rest of the paper is organized as follows. In Section 3 we construct the model selection procedure on the basis of weighted least squares estimates and state the main results in the form of oracle inequalities for the quadratic risk (1.4) and the robust risk (1.6). Here we also specify the set of admissible weight sequences in the model selection procedure. In Section 4 we establish some properties of the stochastic integrals with respect to the non-Gaussian Ornstein–Uhlenbeck process (2.1). Section 7 gives the proofs of the main results. In Sections 5, 6 it is shown that the proposed model selection procedure for estimating S in (1.1) is asymptotically efficient with respect to the robust risk (1.6). Section 7 gives the proofs of the oracle inequalities for the regression model (1.1) with the noises introduced in Section 2. In the Appendix some auxiliary propositions are given.

2. Semimartingale noises

In this section two examples of the disturbances $(\xi_t)_{t\geq 0}$ in (1.1) are given.

2.1. *Non-Gaussian Ornstein–Uhlenbeck process*

First we consider the disturbances $(\xi_t)_{t\geq 0}$ in (1.1) given by a non-Gaussian Ornstein–Uhlenbeck process with the Lévy subordinator. Such processes are used in the financial Black–Scholes type markets with jumps (see for example [6] and the references therein). Let the noise process in (1.1) obey the equation

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

where $a \leq 0$, $u_t = \varrho_1 w_t + \varrho_2 z_t$, ϱ_1 and ϱ_2 are unknown constants, $(w_t)_{t\geq 0}$ is a standard Brownian motion, $(z_t)_{t\geq 0}$ is a compound Poisson process defined as

$$z_t = \sum_{j=1}^{N_t} Y_j. \tag{2.2}$$

Here $(N_t)_{t\geq 0}$ is a standard homogeneous Poisson process with unknown intensity $\lambda > 0$ and $(Y_j)_{j\geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E}Y_j = 0, \qquad \mathbf{E}Y_j^2 = 1 \quad \text{and} \quad \mathbf{E}Y_j^4 < \infty. \tag{2.3}$$

Let $(T)_{k\geq 1}$ denote the arrival times of the process $(N_t)_{t\geq 0}$, that is,

$$T_k = \inf\{t \geq 0:\ N_t = k\}. \tag{2.4}$$

We assume that the parameters λ, a, ϱ_1 and ϱ_2 satisfy the conditions

$$-a_{\max} \leq a \leq 0, \qquad \lambda \geq \lambda_*, \qquad \varrho_{\min} \leq \varrho_1^2 + \lambda \varrho_2^2 \leq \varrho_{\max}. \tag{2.5}$$

Let $\mathcal{Q}_n$ denote the family of all distributions of process (2.1) on the space $\mathcal{D}[0,n]$ with the parameters a, λ, ϱ_1 and ϱ_2 satisfying the conditions (2.5) with fixed bounds $\lambda_* > 0$, $a_{\max} > 0$, $\varrho_{\min} > 0$ and $\varrho_{\max} > 0$.

It will be observed that the process (1.1)–(2.1) may be used for modelling of the stock prices in the financial markets of the Black–Scholes type with jumps (see, e.g., [20], p. 141). In this case the price process $(X_t)_{0\leq t\leq n}$ is governed by the stochastic differential equation:

$$\frac{\mathrm{d}X_t}{X_t} = S(t)\,\mathrm{d}t + \mathrm{d}\xi_t, \tag{2.6}$$

where $(\xi_t)_{t\geq 0}$ is an internal random factor specified by Eq. (2.1) and $S(t)$ is a periodic stock-apprecaition rate which has to be estimated from the observations

$$y_t = y_0 + \int_0^t X_u^{-1}\,\mathrm{d}X_u.$$

The solution to Eq. (2.6) is given by the Dolean exponent, i.e.

$$X_t = X_0 \exp\left\{ y_t - y_0 + \int_0^t \left(S(u) - \frac{\varrho_1^2}{2} \right) \mathrm{d}u + \Lambda_t \right\},$$

where $\Lambda_t = \sum_{T_j \leq t} (\ln(1 + \varrho_2 Y_j) - \varrho_2 Y_j)$.

To use the model (2.6) for describing the stock prices dynamics one needs to require that for all $j \geq 1$

$$1 + \varrho_2 Y_j > 0 \quad \text{a.s.}$$

2.2. *Martingale noise*

Next we consider a martinagale noise obeying the equation

$$\mathrm{d}\xi_t = \varrho_1(t)\,\mathrm{d}w_t + \varrho_2(t)\,\mathrm{d}z_t, \tag{2.7}$$

where ϱ_1 and ϱ_2 are continuously differentiable $\mathbb{R}_+ \to \mathbb{R}$ nonrandom functions; the process $(z_t)_{t\geq 0}$ is defined in (2.2)–(2.3). Assume that, there exist constants $\lambda_* > 0$, $\varrho_{\min} > 0$ and a $\mathbb{R}_+ \to \mathbb{R}_+$ continuous function $\varrho_{\max}(\cdot)$ such that for all $t \geq 0$

$$\lambda \geq \lambda_*, \quad \varrho_{\min} \leq \varrho_1^2(t) + \lambda \varrho_2^2(t) \leq \varrho_{\max}(t) \tag{2.8}$$

and, for any $\delta > 0$,

$$\lim_{t\to\infty} \frac{\varrho_{\max}(t)}{t^\delta} = 0. \tag{2.9}$$

Moreover, we assume that the derivatives of functions ϱ_i for some positive constants ϱ_*' and ϱ_*'' satisfy the following conditions

$$\sup_{t\geq 0}(t+1) \max_{1\leq i\leq 2} \left| \frac{\mathrm{d}}{\mathrm{d}t} \varrho_i^2(t) \right| \leq \varrho_*', \qquad \sup_{t\geq 0} \left| \frac{\mathrm{d}^2}{\mathrm{d}t^2} \varrho_i^2(t) \right| \leq \varrho_*''. \tag{2.10}$$

In this case we denote by $\mathcal{Q}_n$ the family of all distributions of the process (2.7) on $\mathcal{D}[0,n]$ satisfying the conditions (2.8) and (2.10) for some λ_*, $\varrho_{\min}$, $\varrho_{max}(\cdot)$, ϱ_*' and ϱ_*''.

3. Model selection

This section gives the construction of a model selection procedure for estimating a function S in (1.1) on the basis of weighted least square estimates and states the main results.

For estimating the unknown function S in the model (1.1), we apply its Fourier expansion in the trigonometric basis $(\phi_j)_{j\ge 1}$ in $\mathcal{L}_2[0,1]$ defined as

$$\phi_1 = 1, \qquad \phi_j(x) = \sqrt{2}\,\mathrm{Tr}_j\bigl(2\pi[j/2]x\bigr), \quad j \ge 2, \tag{3.1}$$

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. The corresponding Fourier coefficients

$$\theta_j = (S, \phi_j) = \int_0^1 S(t)\phi_j(t)\,\mathrm{d}t \tag{3.2}$$

can be estimated as

$$\widehat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t. \tag{3.3}$$

In view of (1.1), one obtains

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}, \qquad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j), \tag{3.4}$$

where $I_n(\phi_j)$ is given in (1.2).

For any sequence $x = (x_j)_{j\ge 1}$, we set

$$|x|^2 = \sum_{j=1}^{\infty} x_j^2 \quad \text{and} \quad \#(x) = \sum_{j=1}^{\infty} \mathbf{1}_{\{|x_j|>0\}}. \tag{3.5}$$

Now we impose some additional conditions on the family $\mathcal{Q}_n$ of distributions of the noise $(\xi_t)_{t\ge 0}$ in (1.1).

(C_1) *There exists a variance proxy $\varsigma_Q > 0$ such that for any $n \ge 1$*

$$\mathbf{L}_{1,n}(Q) = \sup_{x\in\mathcal{H},\#(x)\le n} \left| \sum_{j=1}^{\infty} x_j \bigl(\mathbf{E}_Q \xi_{j,n}^2 - \varsigma_Q\bigr) \right| < \infty,$$

where $\mathcal{H} = [-1,1]^{\infty}$.

(C_2) *Assume that for each $n \ge 1$*

$$\mathbf{L}_{2,n}(Q) = \sup_{|x|\le 1,\#(x)\le n} \mathbf{E}_Q \left(\sum_{j=1}^{\infty} x_j \widetilde{\xi}_{j,n} \right)^2 < \infty,$$

where $\widetilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2$.

As is shown in the proof of Theorem 3.5 in Section 7, both conditions (C_1) and (C_2) hold for the process (2.1). Further we introduce a class of weighted least squares estimates for $S(t)$ as

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

where $\gamma = (\gamma(j))_{j\ge 1}$ is a sequence of weight coefficients such that

$$0 \le \gamma(j) \le 1 \quad \text{and} \quad 0 < \#(\gamma) \le n. \tag{3.7}$$

Let Γ denote a finite set of such weight sequences $\gamma = (\gamma(j))_{j\ge 1}$, $\nu = \mathrm{card}(\Gamma)$ be its cardinal number and

$$\mu = \max_{\gamma\in\Gamma} \#(\gamma). \tag{3.8}$$

The model selection procedure for the unknown function S in (1.1) will be constructed on the basis of a family of estimates $(\widehat{S}_\gamma)_{\gamma\in\Gamma}$. The choice of a specific set of weight sequences Γ is discussed at the end of this section. To find a proper weight sequence γ in the set Γ, one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. The empirical squared error

$$\mathrm{Err}_n(\gamma) = \|\widehat{S}_\gamma - S\|^2$$

can be written as

$$\mathrm{Err}_n(\gamma) = \sum_{j=1}^\infty \gamma^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^\infty \gamma(j)\widehat{\theta}_{j,n}\theta_j + \sum_{j=1}^\infty \theta_j^2. \tag{3.9}$$

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

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{\widehat{\sigma}_n}{n}, \tag{3.10}$$

where $\widehat{\sigma}_n$ is some estimator for the variance proxy ς_Q in the condition (C_1).

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

$$J_n(\gamma) = \sum_{j=1}^\infty \gamma^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^\infty \gamma(j)\widetilde{\theta}_{j,n} + \rho\widehat{P}_n(\gamma), \tag{3.11}$$

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

$$\widehat{P}_n(\gamma) = \frac{\widehat{\sigma}_n|\gamma|^2}{n}. \tag{3.12}$$

In the case, when the value of ς_Q in (C_1) is known, one can take $\widehat{\sigma}_n = \varsigma_Q$ and

$$P_n(\gamma) = \frac{\varsigma_Q|\gamma|^2}{n}. \tag{3.13}$$

Substituting the weight coefficients, minimizing the cost function

$$\widehat{\gamma} = \operatorname*{argmin}_{\gamma\in\Gamma} J_n(\gamma), \tag{3.14}$$

in (3.6) leads to the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\widehat{\gamma}}. \tag{3.15}$$

It will be noted that $\widehat{\gamma}$ exists because Γ is a finite set. If the minimizing sequence in (3.14) $\widehat{\gamma}$ is not unique, one can take any minimizer.

First we consider the case when the proxy variance ς_Q in (C_1) known.

Proposition 3.1. *If the conditions* (C_1) *and* (C_2) *hold for the distribution Q of the process ξ in* (1.1), *then the risk* (1.4) *of estimate* (3.15) *for S satisfies the oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\rho-2\rho^2}{1-3\rho}\min_{\gamma\in\Gamma}\mathcal{R}_Q(\widehat{S}_\gamma, S) + \frac{1}{n}\mathcal{B}_Q(n,\rho), \tag{3.16}$$

where

$$\mathcal{B}_Q(n,\rho) = 2\frac{\varsigma_Q\sigma_Q\nu + 2\varsigma_Q\mathbf{L}_{1,n}(Q) + \nu\mathbf{L}_{2,n}(Q)}{\varsigma_Q\rho(1-3\rho)} + \frac{6\mu\mathcal{R}_Q(\widehat{\sigma}_n,\varsigma_Q)}{1-3\rho}.$$

This result can be proved along the lines of Theorem 1 in [18].

Now we specify the class $\mathcal{Q}_n$ of admissible distributions Q in the robust risk in (1.6).

Let $\mathcal{Q}_n$ be a set of noise distributions Q on the space $\mathcal{D}[0,n]$ satisfying (1.3), (C_1), (C_2) and the following conditions.

(H_0) *The factor σ_Q in* (1.3) *and the proxy variance ς_Q in* (C_1) *are such that for each $n \ge 1$*

$$\begin{aligned} &\varsigma_n^* := \sup_{Q\in\mathcal{Q}_n} \varsigma_Q < \infty \quad \textit{and} \quad \sigma_n^* := \sup_{Q\in\mathcal{Q}_n} \sigma_Q < \infty, \\ &\varsigma_* := \inf_{Q\in\mathcal{Q}_n} \varsigma_Q > 0 \end{aligned} \tag{3.17}$$

and, moreover, for any $\delta > 0$

$$\lim_{n\to\infty} \frac{\varsigma_n^* + \sigma_n^*}{n^\delta} = 0. \tag{3.18}$$

(H_1) *The functionals $\mathbf{L}_{1,n}(Q)$ and $\mathbf{L}_{2,n}(Q)$ in* (C_1), (C_2) *are uniformly bounded on the set $\mathcal{Q}_n$, i.e. for each $Q \in \mathcal{Q}_n$*

$$\mathbf{L}_{1,n}(Q) \le \mathbf{L}_{1,n}^*, \qquad \mathbf{L}_{2,n}(Q) \le \mathbf{L}_{2,n}^*$$

and the numerical sequences $(\mathbf{L}_{i,n}^*)_{n\ge1}$, $i = 1, 2$, are such that for any $\delta > 0$

$$\lim_{n\to\infty} \frac{\mathbf{L}_{1,n}^* + \mathbf{L}_{2,n}^*}{n^\delta} = 0.$$

Theorem 3.2. *Suppose that the family of admissible noise distributions $\mathcal{Q}_n$ for the model* (1.1) *is defined by the conditions* (C_1), (C_2) *and* (H_0), (H_1). *Then the robust risk* (1.6) *of the estimate* (3.15) *for $S(t)$ satisfies for any $n \ge 1$ and $0 < \rho < 1/3$ the oracle inequality*

$$\mathcal{R}_n^*(\widehat{S}_*, S) \le \frac{1+3\rho-2\rho^2}{1-3\rho} \min_{\gamma\in\Gamma} \mathcal{R}_n^*(\widehat{S}_\gamma, S) + \frac{1}{n}\mathcal{B}^*(n,\rho), \tag{3.19}$$

where

$$\mathcal{B}^*(n,\rho) = 2\frac{\varsigma_n^*\sigma_n^*\nu + 2\varsigma_n^*\mathbf{L}_{1,n}^* + \nu\mathbf{L}_{2,n}^*}{\varsigma_*\rho(1-3\rho)} + \frac{6\mu\mathcal{R}_n^*(\widehat{\sigma}_n)}{1-3\rho}$$

and $\mathcal{R}_n^*(\widehat{\sigma}_n) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widehat{\sigma}_n, \varsigma_Q)$.

3.1. *The case of unknown ς_Q*

If the variance proxy ς_Q in the condition (C_1) is unknown it can be estimated as

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \widehat{\theta}_{j,n}^2, \quad n \ge 2. \tag{3.20}$$

Proposition 3.3. *Suppose that the conditions* (C_1) *and* (C_2) *hold for the model* (1.1) *and $S(\cdot)$ is a continuously differentiable function such that*

$$|\dot{S}|_1 = \int_0^1 \left|\dot{S}(t)\right| \mathrm{d}t < +\infty. \tag{3.21}$$

Then, for any $n \geq 2$,

$$\mathcal{R}_Q(\widehat{\sigma}_n, \varsigma_Q) \leq \frac{\kappa_n(Q,S)}{\sqrt{n}}, \tag{3.22}$$

where

$$\kappa_n(Q,S) = 4|\dot{S}|_1^2\left(1 + \frac{\sqrt{\sigma_Q}}{n^{1/4}}\right) + \varsigma_Q + \sqrt{\mathbf{L}_{2,n}(Q)} + \frac{\mathbf{L}_{1,n}(Q)}{n^{1/2}}.$$

This assertion is a direct consequence of Proposition 4 in [18].

Propositions 3.1 and 3.3 allow one to obtain the following non-asymptotic oracle inequality.

Theorem 3.4. *Let the distribution family $\mathcal{Q}_n$ be as in Theorem 3.2 with unknown ς_Q and S in (1.1) satisfy (3.21). Then, for any $n \geq 1$ and $0 < \rho < 1/3$, the model selection procedure (3.15), (3.20) satisfies the oracle inequality*

$$\mathcal{R}_n^*(\widehat{S}_*, S) \leq \frac{1 + 3\rho - 2\rho^2}{1 - 3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_n^*(\widehat{S}_\gamma, S) + \frac{1}{n}\mathcal{B}_1^*(n,\rho), \tag{3.23}$$

where

$$\mathcal{B}_1^*(n,\rho) = 2\frac{\varsigma_n^*\sigma_n^*\nu + 2\varsigma_n^*\mathbf{L}_{1,n}^* + \nu\mathbf{L}_{2,n}^*}{\varsigma_*\rho(1-3\rho)} + \frac{6\mu\kappa_n^*(S)}{(1-3\rho)\sqrt{n}}$$

and

$$\kappa_n^*(S) = 4|\dot{S}|_1^2\left(1 + \frac{\sqrt{\sigma_n^*}}{n^{1/4}}\right) + \varsigma_n^* + \sqrt{\mathbf{L}_{2,n}^*} + \frac{\mathbf{L}_{1,n}^*}{n^{1/2}}.$$

Moreover, for any $\delta > 0$,

$$\lim_{n\to\infty} \frac{\mathcal{B}_1^*(n,\rho)}{n^\delta} = 0.$$

Now we will obtain the oracle inequalities for the model (1.1) with the noises introduced in Section 2. We will need the following parameter

$$\mathbf{M}^* = 116\varrho_{\max}^2 + 33\varrho_{\max}^2 \frac{\mathbf{E}Y_1^4}{\lambda_*}. \tag{3.24}$$

Theorem 3.5. *Let $\mathcal{Q}_n$ be the distribution family for the Ornstein–Uhlenbeck process (2.1) with the parameters meeting (2.5). Then, for any $n \geq 1$ and $0 < \rho < 1/3$, the estimator (3.15) satisfies the oracle inequality (3.23) with the parameters $\sigma_n^* = 3\varrho_{\max}$, $\varsigma_n^* = \varrho_{\max}$, $\varsigma_* = \varrho_{\min}$, $\mathbf{L}_{1,n}^* = 2(4a_{\max}^2 + 15a_{\max} + 2)\varrho_{\max}$ and $\mathbf{L}_{2,n}^* = 82\mathbf{M}^*$.*

Proof of this theorem is given in Section 7.

Remark 3.1. *It will be noted that the oracle inequality (3.23) for the model (1.1)–(2.1) holds uniformly in the stability region of the process (2.1) including its boundary, i.e. the case when $a = 0$.*

When considering the estimation problem for the model (1.1) with the martingale noise (2.7) we will use two sequences

$$l_{1,n} = 2(1+\lambda_*)\varrho_*' \ln(n+1)$$

and

$$l_{2,n} = (1+\lambda_*)\big(\varrho_{\max}(0) + 2\varrho_*' \ln(n+1)\big) \tag{3.25}$$

with the constants defined in (2.8), (2.10).

Theorem 3.6. *Let $\mathcal{Q}_n$ be the family of distributions of the process* (2.7) *with the parameters meeting* (2.8)–(2.10). *Then, for any $n \geq 1$ and $0 < \rho < 1/3$, the estimator* (3.15) *satisfies the oracle inequality* (3.23) *with $\sigma_n^* = \|\varrho_{\max}\|_{*,n}$, $\varsigma_* = \varrho_{\min}$,*

$$\varsigma_n^* = n^{-1} \int_0^n \varrho_{\max}(u)\,\mathrm{d}u, \qquad \mathbf{L}_{1,n}^* = \varrho_*' + \varrho_*''/2 \quad \textit{and}$$

$$\mathbf{L}_{2,n}^* = \frac{l_{1,n}^2}{n} + 2l_{2,n}^2 + 4\lambda_* \mathbf{E} Y_1^4 \|\varrho_{\max}\|_{*,n}^4.$$

Proof of this theorem is given in Section 7.

Remark 3.2. *If in the model* (2.7) $\lim_{t\to\infty} \varrho_{\max}(t) = \infty$, *then $\varsigma_n^* \approx \varrho_{\max}(n) \to \infty$ as $n \to \infty$ and, by virtue of the condition* (2.9), $\lim_{n\to\infty} \varsigma_n^*/n^\delta = 0$ *for each $\delta > 0$.*

3.2. *Specification of weights in the model selection procedure* (3.15)

We will specify the weight coefficients $(\gamma(j))_{j\geq 1}$ in the way proposed in [11] for a heteroscedastic regression model in discrete time. Consider a numerical grid of the form

$$\mathcal{A}_n = \{1, \dots, k^*\} \times \{t_1, \dots, t_m\}, \tag{3.26}$$

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

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty, & \lim_{n\to\infty} \frac{k^*(n)}{\ln n} = 0, \\ \lim_{n\to\infty} \varsigma_n^* \varepsilon(n) = 0 \quad \text{and} & \lim_{n\to\infty} n^\delta \varepsilon(n) = +\infty, \end{cases} \tag{3.27}$$

where ς_n^* is the least upper bound of the noise variance proxy defined in (3.17). One can take, for example,

$$\varepsilon(n) = \frac{1}{\ln(n+1)} \quad \text{and} \quad k^*(n) = \sqrt{\ln(n+1)}.$$

For each $\alpha = (\beta, t) \in \mathcal{A}_n$, we introduce the weight sequence $\gamma_\alpha = (\gamma_\alpha(j))_{j\geq 1}$ as

$$\gamma_\alpha(j) = \mathbf{1}_{\{1\leq j\leq j_0\}} + \left(1 - (j/\omega_\alpha)^\beta\right)\mathbf{1}_{\{j_0 < j \leq \omega_\alpha\}}, \tag{3.28}$$

where $j_0 = j_0(\alpha) = [\omega_\alpha/\ln(n+1)]$,

$$\omega_\alpha = (\tau_\beta t n)^{1/(2\beta+1)} \quad \text{and} \quad \tau_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}.$$

We set

$$\Gamma = \{\gamma_\alpha, \alpha \in \mathcal{A}_n\}. \tag{3.29}$$

It will be noted that in this case $\nu = k^* m$.

Remark 3.3. *It will be observed that the specific form of weights* (3.28) *was proposed by Pinsker* [25] *for the filtration problem with known smoothness of the regression function observed with an additive Gaussian white noise in continuous time. Nussbaum* [24] *used these weights for the Gaussian regression estimation problem in discrete time.*

The minimal mean square risk, called the Pinsker constant, is provided by the weight least squares estimate with the weights where the index α depends on the smoothness order of the function S. If the smoothness order is unknown one has to use, instead of one estimate, a whole family of estimates containing, in particular, the optimal one.

In this case the problem is to study the properties of the whole class of estimates. Below we derive an oracle inequality for this class which yields the best mean square risk up to a multiplicative and additive constants provided that the smoothness of the unknown function S is not available. Moreover, it will be shown that the multiplicative constant tends to unity and the additive one vanishes as $n \to \infty$ with the rate higher than any minimax rate.

In view of the assumptions (3.27), for any $\delta > 0$, one has

$$\lim_{n\to\infty} \frac{\nu}{n^{\delta}} = 0.$$

Moreover, by (3.28) for any $\alpha \in \widetilde{U}_n$

$$\sum_{j=1}^{\infty} \mathbf{1}_{\{\gamma_\alpha(j)>0\}} \le \omega_\alpha .$$

Therefore, taking into account that $A_\beta \le A_1 < 1$ for $\beta \ge 1$, we get

$$\mu = \mu_n \le (n/\varepsilon)^{1/3}$$

and for any $\delta > 0$

$$\lim_{n\to\infty} \frac{\mu_n}{n^{1/3+\delta}} = 0.$$

To study the asymptotic behaviour of the term $\mathcal{B}_1^*(n,\rho)$ we assume that the parameter ρ in the cost function (3.11) depends on n, i.e. $\rho = \rho_n$ such that $\rho_n \to 0$ as $n \to \infty$ and for any $\delta > 0$

$$\lim_{n\to\infty} n^{\delta} \rho_n = 0. \tag{3.30}$$

Applying this limiting relation in the analysis of the additive term $\mathcal{B}_1^*(n,\rho)$ in (3.23) yields the following result.

Theorem 3.7. *Assume that the family distribution $\mathcal{Q}_n$ satisfies the condition* (H_0) *and the unknown function S is continuously differentiable satisfying the condition* (3.21). *Then, for any $n \ge 1$, the model selection procedure* (3.15), (3.30), (3.20), (3.29) *satisfies the oracle inequality* (3.23) *with the additive term $\mathcal{B}_1^*(n,\rho)$ obeying, for any $\delta > 0$, the following limiting relation*

$$\lim_{n\to\infty} \frac{\mathcal{B}_1^*(n,\rho_n)}{n^{\delta}} = 0.$$

4. Stochastic integrals with respect to the process (2.1)

In this section we establish some properties of the stochastic integral

$$I_t(f) = \int_0^t f(s)\,\mathrm{d}\xi_s, \quad 0 \le t \le n, \tag{4.1}$$

with respect to the process (2.1). We need some notations. Let us denote

$$\varepsilon_f(t) = a\int_0^t \mathrm{e}^{a(t-v)} f(v)\big(1+\mathrm{e}^{2av}\big)\,\mathrm{d}v, \tag{4.2}$$

where f is a $[0,+\infty) \to \mathbb{R}$ function integrated on any finite interval. We introduce also the following transformation

$$\tau_{f,g}(t) = \frac{1}{2}\int_0^t \big(2f(s)g(s) + \varepsilon^*_{f,g}(s)\big)\,\mathrm{d}s \tag{4.3}$$

of square integrable $[0, +\infty) \to \mathbb{R}$ functions f and g. Here

$$\varepsilon^*_{f,g}(t) = f(t)\varepsilon_g(t) + \varepsilon_f(t)g(t).$$

It will be noted that

$$a\tau_{f,1}(t) = \frac{1}{2}\varepsilon_f(t) \quad \text{and} \quad a\tau_{1,1}(t) = \frac{1}{2}\big(\mathrm{e}^{2at} - 1\big). \tag{4.4}$$

Moreover, we set

$$\tau^*_{f,g}(t) = \tau_{f,g}(t) + f(t)\tau_{1,g}(t) + g(t)\tau_{f,1}(t) + f(t)g(t)\tau_{1,1}(t). \tag{4.5}$$

We can rewrite this function as

$$\tau^*_{f,g}(t) = \tau_{f,g}(t) + \frac{\varepsilon^*_{f,g}(t) + f(t)g(t)(\mathrm{e}^{2at} - 1)}{2a}.$$

Proposition 4.1. *If f and g are from $\mathcal{L}_2[0, n]$ then*

$$\mathbf{E} I_t(f) I_t(g) = \widetilde{\varrho}\tau_{f,g}(t), \tag{4.6}$$

where $\widetilde{\varrho} = \varrho_1^2 + \lambda\varrho_2^2$.

Proof. Noting that the process $I_t(f)$ satisfies the stochastic equation

$$\mathrm{d}I_t(f) = af(t)\xi_t \,\mathrm{d}t + f(t)\,\mathrm{d}u_t, \quad I_0(f) = 0,$$

and applying the Ito formula (see, for example, [21]) one obtains

$$\begin{aligned} I_t(f)I_t(g) = &\int_0^t \big(\varrho_1^2 f(s)g(s) + a\big(f(s)\zeta_s(g) + g(s)\zeta_s(f)\big)\big)\,\mathrm{d}s \\ &+ \varrho_2^2 \sum_{l\ge 1} f(T_l)g(T_l)Y_l^2 \mathbf{1}_{\{T_l\le t\}} + \int_0^t \Upsilon_{s-}(f,g)\,\mathrm{d}u_s, \end{aligned} \tag{4.7}$$

where $\zeta_s(f) = I_s(f)\xi_s$ and $\Upsilon_s(f,g) = f(s)I_s(g) + g(s)I_s(f)$. This yields

$$\begin{aligned} \mathbf{E} I_t(f)I_t(g) = a &\int_0^t \big(f(s)Z_s(g) + g(s)Z_s(f)\big)\,\mathrm{d}s \\ &+ \widetilde{\varrho}\int_0^t f(s)g(s)\,\mathrm{d}s, \end{aligned} \tag{4.8}$$

where $Z_s(f) = \mathbf{E}\zeta_s(f)$. Putting here $g = 1$ and taking into account that $\mathbf{E}\xi_s^2 = \widetilde{\varrho}(\mathrm{e}^{2as} - 1)/2a$, we obtain $Z_s(f) = \widetilde{\varrho}\varepsilon_f(s)/2a$. This implies immediately (4.6). Hence Proposition 4.1. □

Further, for integrated $[0, +\infty) \to \mathbb{R}$ functions f and g, we define the $[0, +\infty) \times [0, +\infty) \to \mathbb{R}$ function

$$D_{f,g}(x, z) = \int_0^x L^*_{f,g}(y, z)\,\mathrm{d}y + f(z)g(z), \tag{4.9}$$

where $L^*_{f,g}(y, z) = g(y + z)L_f(y, z) + f(y + z)L_g(y, z)$;

$$L_f(x, z) = a\mathrm{e}^{ax}\bigg(f(z) + a\int_0^x \mathrm{e}^{av} f(v + z)\,\mathrm{d}v\bigg).$$

Proposition 4.2. *Let $\mathcal{G} = \sigma\{T_k, k \geq 1\}$, be σ-algebra generated by the stopping times* (2.4), *f and g be bounded left-continuous* $[0,\infty) \times \Omega \to \mathbb{R}$ *functions measurable with respect to* $\mathcal{B}[0,+\infty) \otimes \mathcal{G}$ (*the product σ-algebra created by* $\mathcal{B}[0,+\infty)$ *and* $\mathcal{G}$). *Then, for any $k \geq 1$,*

$$\mathbf{E}\big(I_{T_k-}(f)|\mathcal{G}\big) = 0$$

and $\mathbf{E}(I_{T_k-}(f)I_{T_k-}(g)|\mathcal{G}) = \varrho_1^2\tau_{f,g}(T_k) + \varrho_2^2\sum_{l=1}^{k-1} D_{f,g}(T_k - T_l, T_l)$.

Proof. Taking the conditional expectation $\mathbf{E}(\cdot|\mathcal{G})$ in (4.7) yields

$$\mathbf{E}\big(I_t(f)I_t(g)|\mathcal{G}\big) = \int_0^t \varrho_1^2 f(s)g(s)\,\mathrm{d}s + \varrho_2^2\sum_{l\geq 1} f(T_l)g(T_l)\mathbf{1}_{\{T_l\leq t\}} + a\int_0^t \big(f(s)\widetilde{Z}_g(s) + g(s)\widetilde{Z}_f(s)\big)\,\mathrm{d}s,$$

where $\widetilde{Z}_f(s) = \mathbf{E}(I_s(f)\xi_s|\mathcal{G})$. By direct calculation we find

$$\widetilde{Z}_f(t) = a\int_0^t f(s)\mathbf{E}(\xi_t\xi_s|\mathcal{G})\,\mathrm{d}s + \varrho_1^2\int_0^t \mathrm{e}^{a(t-s)}f(s)\,\mathrm{d}s + \varrho_2^2\sum_{l\geq 1} f(T_l)\mathrm{e}^{a(t-T_l)}\mathbf{1}_{\{T_l\leq t\}}.$$

Taking into account here that for any $0 \leq s \leq t$

$$\mathbf{E}(\xi_t\xi_s|\mathcal{G}) = \mathrm{e}^{a(t-s)}\left(\frac{\varrho_1^2}{2a}\big(\mathrm{e}^{2as} - 1\big) + \varrho_2^2\sum_{l\geq 1}\mathrm{e}^{2a(s-T_l)}\mathbf{1}_{\{T_l\leq s\}}\right),$$

one obtains,

$$a\widetilde{Z}_f(t) = \frac{\varrho_1^2}{2}\varepsilon_f(t) + \varrho_2^2\sum_{j\geq 1} L_f(t - T_j, T_j)\mathbf{1}_{\{T_j\leq t\}}.$$

From here one comes to the desired equality. □

Proposition 4.3. *Let F, f and g be nonrandom bounded left-continuous* $[0,\infty) \to \mathbb{R}$ *functions. Then*

$$\mathbf{E}\sum_{k\geq 1} F(T_k)I_{T_k-}(f)I_{T_k-}(g)\mathbf{1}_{\{T_k\leq t\}} = \lambda\widetilde{\varrho}\int_0^t F(v)\tau_{f,g}(v)\,\mathrm{d}v.$$

Proof. We set $\iota(t) = \mathbf{E}\sum_{k\geq 1} F(T_k)I_{T_k-}(f)I_{T_k-}(g)\mathbf{1}_{\{T_k\leq t\}}$. By applying Proposition 4.2 one gets

$$\iota(t) = \varrho_1^2\mathbf{E}\sum_{k\geq 1} F(T_k)\tau_{f,g}(T_k)\mathbf{1}_{\{T_k\leq t\}} + \varrho_2^2\mathbf{E}\sum_{k\geq 1} F(T_k)\sum_{l=1}^{k-1} D_{f,g}(T_k - T_l, T_l)\mathbf{1}_{\{T_k\leq t\}}$$
$$:= \varrho_1^2\iota_1(t) + \varrho_2^2\iota_2(t),$$

where

$$\iota_1(t) = \lambda\int_0^t \sum_{l\geq 1} F(z)\tau_{f,g}(z)\frac{(\lambda z)^{l-1}}{(l-1)!}\mathrm{e}^{-\lambda z}\,\mathrm{d}z = \lambda\int_0^t F(z)\tau_{f,g}(z)\,\mathrm{d}z.$$

To calculate $\iota_2(t)$ we note that

$$\iota_2(t) = \mathbf{E} \sum_{l \geq 1} \mathbf{1}_{\{T_l \leq t\}} \sum_{k \geq l+1} F(T_k) D_{f,g}(T_k - T_l, T_l) \mathbf{1}_{\{T_k \leq t\}}.$$

Taking into account that $T_k - T_l$ is independent of T_l for any $k > l$ we obtain

$$\begin{aligned}
\iota_2(t) &= \lambda \mathbf{E} \sum_{l \geq 1} \mathbf{1}_{\{T_l \leq t\}} \int_0^{t-T_l} \sum_{k \geq l+1} F(z + T_l) D_{f,g}(z, T_l) \frac{(\lambda z)^{k-l-1}}{(k-l-1)!} \mathrm{e}^{-\lambda z} \, \mathrm{d}z \\
&= \lambda \mathbf{E} \sum_{l \geq 1} \mathbf{1}_{\{T_l \leq t\}} \int_0^{t-T_l} F(z + T_l) D_{f,g}(z, T_l) \, \mathrm{d}z \\
&= \lambda^2 \int_0^t \int_0^{t-x} \big(F(z+x) D_{f,g}(z, x) \, \mathrm{d}z \big) \, \mathrm{d}x = \lambda^2 \int_0^t F(z) \tau_{f,g}(z) \, \mathrm{d}z.
\end{aligned}$$

Hence Proposition 4.3. □

Now we set

$$\widetilde{I}_t(f) = I_t^2(f) - \mathbf{E} I_t^2(f). \tag{4.10}$$

Further we need the following correlation measures for two integrated $[0, +\infty) \to \mathbb{R}$ functions f and g

$$\varpi_{f,g} = \max_{0 \leq v \leq n} \max_{0 \leq t \leq n-v} \left| \int_0^t f(u+v) g(u) \, \mathrm{d}u \right| \tag{4.11}$$

and

$$\varpi^*_{f,g} = \max(\varpi_{f,g}, \varpi_{g,f}). \tag{4.12}$$

For any bounded $[0, \infty) \to \mathbb{R}$ function f we introduce the following uniform norm

$$\|f\|_{*,n} = \sup_{0 \leq t \leq n} \big| f(t) \big|.$$

To check the condition (C_2) we need the following non-asymptotic upper bound

Theorem 4.4. *For any left-continuous* $[0, \infty) \to \mathbb{R}$ *functions* f, g *with* $\|f\|_{*,n} \leq 1$ *and* $\|g\|_{*,n} \leq 1$

$$\big| \mathbf{E} \widetilde{I}_n(f) \widetilde{I}_n(g) \big| \leq n \mathbf{M}_Q \big(1 + \varpi^*_{f,g} + \varpi^*_{f,1} + \varpi^*_{1,g} \big), \tag{4.13}$$

where $\mathbf{M}_Q = 116 \widetilde{\varrho}^2 + 33 \lambda \varrho_2^4 \mathbf{E} Y_1^4$.

Proof. Taking in (4.7)–(4.8) $g = f$ and $V_t(f) = \zeta_t(f) - Z_t(f)$, one comes to the following stochastic equation

$$\mathrm{d}\widetilde{I}_t(f) = 2a V_t(f) f(t) \, \mathrm{d}t + \mathrm{d}M_t(f), \quad \widetilde{I}_0(f) = 0, \tag{4.14}$$

where $M_t(f) = 2 \int_0^t I_{s-}(f) f(s) \, \mathrm{d}u_s + \varrho_2^2 \int_0^t f^2(s) \, \mathrm{d}m_s$ and

$$m_t = \sum_{0 \leq s \leq t} \Delta z_s^2 - \lambda t.$$

Moreover, by the Ito formula one finds for $t \ge 0$

$$\mathbf{E}\widetilde{I}_t(f)\widetilde{I}_t(g) = \mathbf{E}\big[\widetilde{I}(f), \widetilde{I}(g)\big]_t + 2\int_0^t \big(f(s)\mathcal{T}_{f,g}(s) + g(s)\mathcal{T}_{g,f}(s)\big)\,\mathrm{d}s, \tag{4.15}$$

where $\mathcal{T}_{f,g}(t) = a\mathbf{E}V_t(f)\widetilde{I}_t(g)$. To calculate the first expectation in the right-hand side of this equality we note that continuous martingale component of the semimartingale (4.14) is $\widetilde{I}_t^c(f) = 2\varrho_1\int_0^t I_s(f)f(s)\,\mathrm{d}w_s$. Therefore, by Proposition 4.1

$$\mathbf{E}\big\langle\widetilde{I}^c(f), \widetilde{I}^c(g)\big\rangle_t = 4\varrho_1^2\widetilde{\varrho}\int_0^t \tau_{f,g}(s)f(s)g(s)\,\mathrm{d}s.$$

Moreover, in view of Proposition 4.3, one finds

$$\mathbf{E}\Delta\widetilde{I}_t(f)\Delta\widetilde{I}_t(g) = 4\lambda\varrho_2^2\widetilde{\varrho}\int_0^t \tau_{f,g}(s)f(s)g(s)\,\mathrm{d}s + \varrho_3\int_0^t f^2(s)g^2(s)\,\mathrm{d}s,$$

where $\varrho_3 = \lambda\varrho_2^4\mathbf{E}Y_1^4$. This yields

$$\mathbf{E}\big[\widetilde{I}(f), \widetilde{I}(g)\big]_t = \int_0^t G_{f,g}(s)\,\mathrm{d}s, \tag{4.16}$$

where $G_{f,g}(t) = 4\widetilde{\varrho}^2 f(t)g(t)\tau_{f,g}(t) + \varrho_3 f^2(t)g^2(t)$. Lemma A.1 implies

$$\|G_{f,g}\|_{*,n} \le 16\widetilde{\varrho}^2\varpi^*_{f,g} + \varrho_3. \tag{4.17}$$

Further from (4.4) we obtain

$$G_{1,1}(t) = \frac{2\widetilde{\varrho}^2}{a}\big(\mathrm{e}^{2at} - 1\big) + \varrho_3. \tag{4.18}$$

Putting in (4.7)–(4.8) $g = 1$, we get

$$\mathrm{d}V_t(f) = aV_t(f)\,\mathrm{d}t + af(t)\widetilde{\xi}_t\,\mathrm{d}t + \mathrm{d}K_t(f), \tag{4.19}$$

where $\widetilde{\xi}_t = \xi_t^2 - \mathbf{E}\xi_t^2$, $K_t(f) = \int_0^t I^*_{s-}(f)\,\mathrm{d}u_s + \varrho_2^2\int_0^t f(s)\,\mathrm{d}m_s$ and $I_t^*(f) = \Upsilon_s(f,1) = I_t(f) + f(t)\xi_t$. To calculate the function $\mathcal{T}_{f,g}(t)$, we note that, from the Ito formula

$$\mathbf{E}V_t(f)\widetilde{I}_t(g) = a\int_0^t \big(\mathbf{E}V_s(f)\widetilde{I}_s(g) + 2g(s)\mathbf{E}V_s(f)V_s(g)\big)\,\mathrm{d}s + a\int_0^t f(s)\mathbf{E}\widetilde{\xi}_s\widetilde{I}_s(g)\,\mathrm{d}s + \mathbf{E}\big[K(f), M(g)\big]_t.$$

To calculate the last expectation we note that Proposition 4.1 yields

$$\mathbf{E}I_t(g)I_t^*(f) = \widetilde{\varrho}\big(\tau_{f,g} + f(t)\tau_{g,1}\big).$$

Now, by applying Propositions 4.1–4.3, we find

$$\mathbf{E}\big[K(f), M(g)\big]_t = \int_0^t U_{f,g}(s)\,\mathrm{d}s, \tag{4.20}$$

where $U_{f,g}(s) = 2\widetilde{\varrho}^2 g(s)\tau_{f,g}(s) + 2\widetilde{\varrho}^2 g(s)f(s)\tau_{1,g}(s) + \varrho_3 f(s)g^2(s)$. Therefore,

$$\mathbf{E}V_t(f)\widetilde{I}_t(g) = a\int_0^t \big(\mathbf{E}V_s(f)\widetilde{I}_s(g) + 2g(s)\mathbf{E}V_s(f)V_s(g)\big)\,\mathrm{d}s \\ + a\int_0^t f(s)\mathbf{E}\widetilde{\xi}_s\widetilde{I}_s(g)\,\mathrm{d}s + \int_0^t U_{f,g}(s)\,\mathrm{d}s. \tag{4.21}$$

Since $\widetilde{I}_t(1) = V_t(1) = \widetilde{\xi}_t$, the last equality for $f = g = 1$ implies

$$\mathbf{E}\widetilde{\xi}_t^2 = \int_0^t \mathrm{e}^{4a(t-s)}U_{1,1}(s)\,\mathrm{d}s = \mathrm{e}^{4at}\frac{2\widetilde{\varrho}^2 + a\varrho_3}{4a^2} - \mathrm{e}^{2at}\frac{\widetilde{\varrho}^2}{a^2} + \frac{2\widetilde{\varrho}^2 - a\varrho_3}{4a^2}. \tag{4.22}$$

We define the function

$$A_f(t) = \int_0^t \mathrm{e}^{3a(t-s)}\big(f(s)a^2\mathbf{E}\widetilde{\xi}_s^2 + \kappa_f(s)\big)\,\mathrm{d}s, \tag{4.23}$$

where $\kappa_f(t) = \widetilde{\varrho}^2(\varepsilon_f(t) + f(t)(\mathrm{e}^{2at} - 1)) + a\varrho_3 f(t)$.

In Lemma A.2 (see the Appendix) we prove that

$$\mathbf{E}V_t(f)V_t(g) = \int_0^t \mathrm{e}^{2a(t-s)}H_{f,g}(s)\,\mathrm{d}s, \tag{4.24}$$

where $H_{f,g}(s) = g(s)A_f(s) + f(s)A_g(s) + \widetilde{\varrho}^2\tau^*_{f,g}(s) + \varrho_3 f(s)g(s)$.

Moreover, substituting $f = 1$ in (4.21) yields

$$\mathbf{E}\widetilde{\xi}_t\widetilde{I}_t(g) = \int_0^t \mathrm{e}^{2a(t-s)}\big(g(s)\widetilde{H}_{1,g}(s) + U_{1,g}(s)\big)\,\mathrm{d}s,$$

where $\widetilde{H}_{f,g}(t) = 2a\int_0^t \mathrm{e}^{2a(t-s)}H_{f,g}(s)\,\mathrm{d}s$. Furthermore, (4.21) implies

$$\mathbf{E}V_t(f)\widetilde{I}_t(g) = \int_0^t \mathrm{e}^{a(t-s)}g(s)\widetilde{H}_{f,g}(s)\,\mathrm{d}s + \int_0^t \mathrm{e}^{a(t-s)}f(s)\mathcal{H}_g(s)\,\mathrm{d}s \\ + \int_0^t \mathrm{e}^{a(t-s)}f(s)\widetilde{U}_{1,g}(s)\,\mathrm{d}s + \int_0^t \mathrm{e}^{a(t-s)}U_{f,g}(s)\,\mathrm{d}s,$$

where $\mathcal{H}_g(t) = a\int_0^t \mathrm{e}^{2a(t-s)}g(s)\widetilde{H}_{1,g}(s)\,\mathrm{d}s$ and $\widetilde{U}_{f,g}(t) = a\int_0^t \mathrm{e}^{2a(t-s)}U_{f,g}(s)\,\mathrm{d}s$. Therefore,

$$\mathcal{T}_{f,g}(t) = a\int_0^t \mathrm{e}^{a(t-s)}\big(g(s)\widetilde{H}_{f,g}(s) + f(s)\mathcal{H}_g(s) + f(s)\widetilde{U}_{1,g}(s)\big)\,\mathrm{d}s + \widetilde{U}_{f,g}(t).$$

Lemmas A.5–A.8 imply

$$\|\mathcal{T}_{f,g}\|_{*,n} \le 25\widetilde{\varrho}^2\varpi^*_{f,g} + 9\widetilde{\varrho}^2\varpi^*_{1,g} + 8\varrho_3 + 7\widetilde{\varrho}^2.$$

Combining (4.15)–(4.17) yields

$$\frac{|\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g)|}{n} \le 16\widetilde{\varrho}^2\varpi^*_{f,g} + \varrho_3 + 2\big(\|\mathcal{T}_{f,g}\|_{*,n} + \|\mathcal{T}_{g,f}\|_{*,n}\big).$$

From here one comes to the upper bound (4.13). □

5. Robust asymptotic efficiency

In this section we show that the model selection procedure (3.15), (3.30), (3.20), (3.29) for estimating S in the model (1.1) is asymptotically efficient with respect to the robust risk (1.6). We assume that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W_r^k = \left\{ f \in \mathcal{C}_{\mathrm{per}}^k[0,1], \sum_{j=0}^{k} \| f^{(j)} \|^2 \le r \right\}, \tag{5.1}$$

where $r > 0$, $k \ge 1$ are some parameters, $\mathcal{C}_{\mathrm{per}}^k[0,1]$ is the set of k times continuously differentiable functions $f : [0,1] \to \mathbb{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \le i \le k$. The function class W_r^k can be written as an ellipsoid in l_2, i.e.

$$W_r^k = \left\{ f \in \mathcal{C}_{\mathrm{per}}^k[0,1]\colon \sum_{j=1}^{\infty} a_j \theta_j^2 \le r \right\}, \tag{5.2}$$

where $a_j = \sum_{i=0}^{k} (2\pi[j/2])^{2i}$. We denote by Q_0 the distribution of the Winer process with the scale parameter ς_n^* defined in (3.17).

(H_3) *Assume that the distribution Q_0 belongs to the family $\mathcal{Q}_n$.*

In this section we will show that the Pinsker constant for the robust risk (1.6) is given by the equation

$$R_{k,n}^* = \left(\varsigma_n^*\right)^{2k/(2k+1)} R_k^0, \tag{5.3}$$

where

$$R_k^0 = \left((2k+1)r\right)^{1/(2k+1)} \left(\frac{k}{(k+1)\pi} \right)^{2k/(2k+1)}.$$

Note that R_k^0 is the well-known Pinsker constant obtained for the nonadaptive filtration problem in "signal + small white noise" model (see, for example, [25]).

It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball W_r^k is $n^{2k/(2k+1)}$ (see, for example, [24,25]).

We will see that, asymptotically, the robust risk (1.6) normalized by this rate is bounded from below by $R_{k,n}^*$, i.e. this bound can not be diminished if one considers the class of all admissible estimates for S. Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the sigma-algebra $\sigma\{y_t, 0 \le t \le n\}$ generated by the process (1.1).

Theorem 5.1. *Under the condition* (H_3)

$$\liminf_{n\to\infty} \frac{n^{2k/(2k+1)}}{R_{k,n}^*} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_n, S) \ge 1. \tag{5.4}$$

Proof of this theorem is similar to that of Theorem 3.2 in [19].

Now we show that, under some conditions, the normalized robust risk for the model selection procedure is bounded from above by the same constant $R_{k,n}^*$.

Theorem 5.2. *Assume that, in model* (1.1), *for each $n \ge 1$ the distribution of $(\xi_t)_{0\le t\le n}$ belongs to the family $\mathcal{Q}_n$ satisfying the conditions* (H_0), (H_1). *Then the robust risk* (1.6) *of the model selection procedure $\widehat{S}_*$, defined in* (3.30), (3.20), (3.29), *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \frac{n^{2k/(2k+1)}}{R_{k,n}^*} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le 1. \tag{5.5}$$

Theorem 5.1 and Theorem 5.2 imply the following result

Corollary 5.3. *Under the conditions* (H_0)–(H_2)

$$\lim_{n\to\infty} \frac{n^{2k/(2k+1)}}{R^*_{k,n}} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S\in W^k_r} \mathcal{R}^*_n(\widehat{S}_n, S) = 1. \tag{5.6}$$

Remark 5.1. *Equation* (5.6) *means that the parameter $R^*_{k,n}$ defined in* (5.3) *is the Pinsker constant* (*see*, *for example*, [25]) *for the model* (1.1) *and that the model selection procedure* (3.30), (3.20), (3.29) *is asymptotically robust efficient*.

Remark 5.2. *Note that $R^*_{k,n}$ coincides in the form with the well-known Pinsker constant in a nonparametric fitration problem of signal observed with white Gaussian noise* [25]. *In the case of white noise model* (*i.e. $\xi_t = w_t$ in* (1.1)) *the intensity of noise in* (3.4) *remains constant in the whole range of frequencies*, *i.e.* $\mathbf{E}\xi^2_{j,n} = 1$ *for all* $j \geq 1$. *For the model* (1.1), (2.1), *the noise intensity of $\xi_{j,n}$ stabilizes*, *in view of* (2.1), *only in the range of high frequencies*, *i.e. for each* $n \geq 1$

$$\mathbf{E}\xi^2_{j,n} = \rho^* + \mathrm{O}\left(\frac{1}{j^2}\right) \quad \text{as } j \to \infty.$$

When the distribution of the noise $(\xi_t)_{t\geq 0}$ *is known* (*the class $\mathcal{Q}_n$ in* (2.1) *consists of a single distribution*), *the quantity $R^*_{k,n}$ is given by* (5.3) *with* $\varsigma^*_n = \rho^*$. *It means that the lower bound for the quadratic risk*, *in the case of the colored noise*, *is determined by the noise intensity only at high frequencies*. *When the noise distribution is unknown*, *one should use ς^*_n defined in* (3.17) *which equals the supremum of ρ^* over the set of all admissible distributions*. *In other words*, *the quadratic risk lower bound is determined by the maximum of the noise intensity at high frequencies taken over the whole class of admissible distributions*.

Remark 5.3. *It will be observed that the standard optimal convergence rate of the robust risk of the model selection procedure* (3.15) *for the model* (1.1) *essentially rests an the assumptions providing stabilization of the maximal noise intensity ς^*_n as $n \to \infty$*. *Less stringent assumptions on the noise process may result in worsening the convergence rate*. *As an example of this phenomenon*, *we consider the model* (1.1) *with the noise* (2.7) *and assume that $\varrho_{\max} \to \infty$ in such a way that ς^*_n tends to infinity more slowly than any power function n^δ*, $\delta > 0$ *as* $n \to \infty$. *Then*, *in virtue of* (5.3), (5.4), *the risk convergence rate becomes worse*, *namely*, $(n/\varsigma^*_n)^{2k/(2k+1)}$.

6. Upper bound

6.1. *Known smoothness*

First we suppose that the parameters $k \geq 1$, $r > 0$ in (5.1) and ς^*_n in (3.17) are known. Let the family of admissible weighted least squares estimates $(\widehat{S}_\gamma)_{\gamma\in\Gamma}$ for the unknown function $S \in W^k_r$ be given by (3.29). Consider the pair

$$\alpha_0 = (k, t_0),$$

where $t_0 = [\overline{r}/\varepsilon]\varepsilon$, $\overline{r} = r/\varsigma^*_n$ and ε satisfies the conditions (3.27). Denote the corresponding weight sequence in Γ as

$$\gamma_0 = \gamma_{\alpha_0}. \tag{6.1}$$

Note that for sufficiently large n the pair α_0 belongs to the set (3.26).

Theorem 6.1. *The estimator $\widehat{S}_{\gamma_0}$ has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \frac{n^{2k/(2k+1)}}{R^*_{k,n}} \sup_{S\in W^k_r} \mathcal{R}^*_n(\widehat{S}_{\gamma_0}, S) \leq 1. \tag{6.2}$$

Proof. Substituting (3.4) in (3.6) and using the weight sequence (6.1) one gets

$$\|\widehat{S}_{\gamma_0}-S\|^2=\sum_{j=1}^{\infty}\bigl(1-\gamma_0(j)\bigr)^2\theta_j^2-2M_n+\frac{1}{n}\sum_{j=1}^{\infty}\gamma_0^2(j)\xi_{j,n}^2,$$

where

$$M_n=\frac{1}{\sqrt{n}}\sum_{j=1}^{\infty}\bigl(1-\gamma_0(j)\bigr)\gamma_0(j)\theta_j\xi_{j,n}.$$

It should be observed that $\mathbf{E}_{Q,S}M_n=0$ for any $Q\in\mathcal{Q}_n^*$. Moreover, by the condition (C_1)

$$\mathbf{E}_{Q,S}\sum_{j=1}^{\infty}\gamma_0^2(j)\xi_{j,n}^2\le\varsigma_Q\sum_{j=1}^{\infty}\gamma_0^2(j)+\mathbf{L}_{1,n}(Q)$$

and, in view of the condition (H_0), this implies

$$\sup_{Q\in\mathcal{Q}_n}\mathbf{E}_{Q,S}\sum_{j=1}^{\infty}\gamma_0^2(j)\xi_{j,n}^2\le\varsigma_n^*\sum_{j=1}^{\infty}\gamma_0^2(j)+l_n.$$

Thus,

$$\mathcal{R}_n^*(\widehat{S}_{\gamma_0},S)\le\sum_{j=\iota_0}^{\infty}\bigl(1-\gamma_0(j)\bigr)^2\theta_j^2+\frac{\varsigma_n^*}{n}\sum_{j=1}^{\infty}\gamma_0^2(j)+\frac{l_n}{n},\tag{6.3}$$

where $\iota_0=j_0(\alpha_0)$. Setting

$$\upsilon_n=n^{2k/(2k+1)}\sup_{j\ge\iota_0}\bigl(1-\gamma_0(j)\bigr)^2/a_j,$$

we estimate the first summand in the right-hand of (6.3) as

$$n^{2k/(2k+1)}\sum_{j=\iota_0}^{\infty}\bigl(1-\gamma_0(j)\bigr)^2\theta_j^2\le\upsilon_n\sum_{j\ge1}a_j\theta_j^2.$$

From here and (5.2), we obtain that for each $S\in W_r^k$

$$\Upsilon_{1,n}(S)=n^{2k/(2k+1)}\sum_{j=\iota_0}^{\infty}\bigl(1-\gamma_0(j)\bigr)^2\theta_j^2\le\upsilon_n r.$$

Further we note that

$$\limsup_{n\to\infty}(t_0)^{2k/(2k+1)}\upsilon_n\le\frac{1}{\pi^{2k}(\tau_k)^{2k/(2k+1)}},$$

where τ_k is given in (3.28). Moreover, by the condition (3.27), $\lim_{n\to\infty}t_0/\overline{r}=1$. Therefore,

$$\limsup_{n\to\infty}\frac{1}{(\varsigma_n^*)^{2k/(2k+1)}}\sup_{S\in W_r^k}\Upsilon_{1,n}(S)\le\frac{r^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2k/(2k+1)}}:=\Upsilon_1^*.$$

To examine the second summand in the right-hand of (6.3), we set

$$\Upsilon_{2,n} = \frac{(\varsigma_n^*)^{1/(2k+1)}}{n^{1/(2k+1)}} \sum_{j=1}^{n} \gamma_0^2(j).$$

It is easy to check that

$$\limsup_{n\to\infty} \Upsilon_{2,n} \le \frac{2(r\tau_k)^{1/(2k+1)} k^2}{(k+1)(2k+1)} := \Upsilon_2^*.$$

Therefore, taking into account that by the definition of the Pinsker constant in (5.3) $\Upsilon_1^* + \Upsilon_2^* = R_k^0$, we arrive at the inequality

$$\lim_{n\to\infty} \frac{n^{2k/(2k+1)}}{R_{k,n}^*} \sup_{S\in W_r^k} \mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \le 1.$$

Hence Theorem 6.1. □

6.2. *Unknown smoothness*

Combining Theorem 6.1 and Theorem 3.7 yields Theorem 5.2.

7. Proofs

7.1. *Proof of Theorem* 3.5

Applying the Proposition 4.1 we find that the inequality (1.3) for the process (2.1) holds with $\sigma_Q = 3\widetilde{\varrho}$. This, in view of (2.5), yields $\sigma_n^* = 3\varrho_{\max}$. Now we verify conditions (C_1) and (C_2) for the family of processes (2.1) satisfying (2.5). First we note that

$$\mathbf{E}_{Q,S} \xi_{j,n}^2 = \widetilde{\varrho}(1 + b_{j,n}), \tag{7.1}$$

where $b_{j,n} = n^{-1} a \int_0^n \mathrm{e}^{av} \Upsilon_j(v)\,\mathrm{d}v$ and

$$\Upsilon_j(v) = \int_0^{n-v} \phi_j(t+v)\phi_j(t)\bigl(1 + \mathrm{e}^{2at}\bigr)\,\mathrm{d}t.$$

If $j = 1$, one has

$$\bigl|\mathbf{E}_{Q,S} \xi_{1,n}^2 - \widetilde{\varrho}\bigr| \le 2\widetilde{\varrho}. \tag{7.2}$$

Since for the trigonometric basis (3.1) for $j \ge 2$

$$\phi_j(t+v)\phi_j(t) = \cos(\gamma_j v) + (-1)^j \cos\bigl(\gamma_j(2t+v)\bigr),$$

where $\gamma_j = 2\pi[j/2]$, therefore,

$$\Upsilon_j(v) = \cos(\gamma_j v) F(v) + (-1)^j \Upsilon_{0,j}(v), \qquad F(v) = \int_0^{n-v} \bigl(1 + \mathrm{e}^{2at}\bigr)\,\mathrm{d}t$$

and

$$\Upsilon_{0,j}(v) = \int_0^{n-v} \cos\bigl(\gamma_j(2t+v)\bigr)\bigl(1 + \mathrm{e}^{2at}\bigr)\,\mathrm{d}t.$$

Integrating by parts one finds

$$\Upsilon_{0,j}(v) = -\frac{2+\mathrm{e}^{2a(n-v)}}{2\gamma_j}\sin(v\gamma_j) + \frac{a}{2\gamma_j^2}\Upsilon_{1,j}(v),$$

where

$$\Upsilon_{1,j}(v) = \cos(v\gamma_j)\big(\mathrm{e}^{2a(n-v)} - 1\big) - 2a\int_0^{n-v}\mathrm{e}^{2at}\cos\big((2t+v)\gamma_j\big)\,\mathrm{d}t.$$

It is obvious that $|\Upsilon_{1,j}(v)| \le 2$. Further we calculate

$$\begin{aligned} b_{j,n} &= \frac{a}{n}\int_0^n \mathrm{e}^{av}F(v)\cos(v\gamma_j)\,\mathrm{d}v + \frac{a}{n}(-1)^j\int_0^n \mathrm{e}^{av}\Upsilon_{0,j}(v)\,\mathrm{d}v \\ &:= aD_{1,j} + a(-1)^j D_{2,j}. \end{aligned}$$

Integrating by parts two times yields

$$D_{1,j} = \frac{1}{n\gamma_j^2}\left(\mathrm{e}^{an}\dot{F}(n) - \dot{F}(0) - aF(0) - \int_0^n \mathrm{e}^{av}F_1(v)\,\mathrm{d}v\right),$$

where $F_1(v) = a^2F(v) + 2a\dot{F}(v) + \ddot{F}(v)$. Since $\gamma_j \ge j$ for $j \ge 2$, we obtain

$$|D_{1,j}| \le \frac{1}{j^2}\big(4|a| + 10\big).$$

Similarly, one gets $|D_{2,j}| \le 5/j^2$. Substituting these estimates in (7.1) and using the upper bound (7.2), we obtain for all $j \ge 1$

$$\left|\mathbf{E}_{Q,S}\xi_{j,n}^2 - \widetilde{\varrho}\right| \le \widetilde{\varrho}\,\frac{(4a^2 + 15|a| + 2)}{j^2}. \tag{7.3}$$

Thus we arrive at the inequality

$$\mathbf{L}_{1,n}(Q) \le 2\widetilde{\varrho}\big(4a^2 + 15|a| + 2\big),$$

which implies that

$$\mathbf{L}_{1,n}(Q) \le \mathbf{L}_1^*, \tag{7.4}$$

where $\mathbf{L}_1^*$ is defined in (3.24). Therefore the condition (C_1) holds with $\varsigma_Q = \widetilde{\varrho}$. Applying the conditions (2.5) we find $\varsigma_n^* = \varrho_{\max}$ and $\varsigma_* = \varrho_{\min}$. To check ($\mathrm{C}_2$) we represent the sum as

$$\sum_{j=1}^{\infty} x_j\widetilde{\xi}_{j,n} = \frac{1}{n}J_{1,n} + \frac{1}{n}J_{2,n},$$

where $J_{1,n} = x_1\widetilde{I}_n(\phi_1) + x_2\widetilde{I}_n(\phi_2)$ and $J_{2,n} = \sum_{j\ge 3}^{\infty} x_j\widetilde{I}_n(\phi_j)$. From here we have

$$\mathbf{E}_{Q,S}\left(\sum_{j=1}^{\infty} x_j\widetilde{\xi}_{j,n}\right)^2 \le \frac{2}{n^2}\big(\mathbf{E}_{Q,S}J_{1,n}^2 + \mathbf{E}_{Q,S}J_{2,n}^2\big). \tag{7.5}$$

By applying the Cauchy–Schwarz–Bounyakovskii inequality and noting that $x_1^2 + x_2^2 \le 1$, one gets

$$\mathbf{E}_{Q,S}J_{1,n}^2 \le \mathbf{E}_{Q,S}\widetilde{I}_n^2(\phi_1) + \mathbf{E}_{Q,S}\widetilde{I}_n^2(\phi_2).$$

By Theorem 4.4 this implies

$$\mathbf{E}_{Q,S} J_{1,n}^2 \le 2n\mathbf{M}_Q\big(1 + 3\varpi_{1,1}^* + 2\varpi_{1,2}^* + \varpi_{2,2}^*\big),$$

where $\varpi_{i,j}^* = \varpi_{\phi_i,\phi_j}^*$. We note that each $\varpi_{i,j}^*$ can be estimated as

$$\varpi_{i,j}^* \le \sqrt{\int_0^n \phi_i^2(u)\,\mathrm{d}u}\sqrt{\int_0^n \phi_j^2(u)\,\mathrm{d}u} = n.$$

Therefore

$$\mathbf{E}_{Q,S} J_{1,n}^2 \le 14\mathbf{M}_Q n^2. \tag{7.6}$$

Applying Theorem 4.4 and taking into account that $\|\phi_j\|_{*,n} \le \sqrt{2}$, one gets

$$\mathbf{E}_{Q,S} J_{2,n}^2 \le 4n\mathbf{M}_Q \sum_{i,j\ge 3} |x_i||x_j|\widetilde{\kappa}_{i,j}, \tag{7.7}$$

where $\widetilde{\kappa}_{i,j} = 1 + \varpi_{i,j}^* + \varpi_{1,j}^* + \varpi_{1,i}^*$. To estimate the coefficient $\varpi_{i,j}^*$ we note, that for any $i \ge 3$,

$$\phi_i(v+u) = \varkappa_{1,i}(v)\phi_{i-1}(u) + \varkappa_{2,i}(v)\phi_i(u) + \varkappa_{3,i}(v)\phi_{i+1}(u),$$

where $\varkappa_{i,j}(\cdot)$ are bounded functions. From here in view of the orthonormality and the periodicity of the functions $(\phi_j)_{j\ge 1}$, it follows that for $0 \le t \le n$ and $|i-j| \ge 2$

$$\begin{aligned}\left|\int_0^t \phi_i(u+v)\phi_j(u)\,\mathrm{d}u\right| &= \left|\int_0^{\{t\}} \phi_i(u+v)\phi_j(u)\,\mathrm{d}u\right| \\ &\le \sqrt{\int_0^1 \phi_i^2(u+v)\,\mathrm{d}u} = 1,\end{aligned}$$

where $\{t\}$ is the fractional part of t. Therefore $\varpi_{i,j}^* \le 1$ if $|i-j| \ge 2$. Thus, $\varpi_{i,j}^* \le n\mathbf{1}_{\{|i-j|\le 1\}} + \mathbf{1}_{\{|i-j|\ge 2\}}$. Note now that

$$\sum_{i,j\ge 1} |x_i||x_j| = \left(\sum_{i\ge 1}|x_i|\right)^2 \le \#(x)\left(\sum_{i\ge 1} x_i^2\right) \le n.$$

Moreover,

$$\begin{aligned}\sum_{i,j\ge 3} \mathbf{1}_{\{|i-j|\le 1\}}|x_i||x_j| &= \sum_{i\ge 1} x_i^2 + 2\sum_{i\ge 3}|x_i||x_{i-1}| \\ &\le \sum_{i\ge 1} x_i^2 + 2\left(\sum_{i\ge 1} x_i^2\right)^2 \le 3.\end{aligned}$$

By making use of these estimates in (7.7) one gets

$$\sum_{i,j\ge 3} |x_i||x_j|\widetilde{\kappa}_{i,j} \le \sum_{i,j\ge 3} |x_i||x_j|\big(3 + \varpi_{i,j}^*\big) \le 7n.$$

From here and the inequalities (7.5)–(7.7), it follows that $\mathbf{L}_{2,n}(Q) \le 84\mathbf{M}_Q$. By the definition of the distribution family in (2.5) $\mathbf{M}_Q \le \mathbf{M}^*$, where $\mathbf{M}^*$ is given in (3.24). Hence Theorem 3.5.

7.2. *Proof of Theorem* 3.6

By making use of (2.7), (4.1) and applying Ito's formula one obtains that for any square integrable $\mathbb{R}_+ \to \mathbb{R}$ functions f and g

$$\mathbf{E} I_t(f) I_t(g) = \int_0^t f(u) g(u) \widetilde{\varrho}(u)\,\mathrm{d}u, \tag{7.8}$$

where $\widetilde{\varrho}(u) = \varrho_1^2(u) + \lambda \varrho_2^2(u)$. Therefore condition (1.3) holds with $\sigma_Q = \|\varrho_1^2 + \lambda\varrho_2^2\|_{*,n}$. Further we will show that the proxy variance ς_Q in (C_1) can be defined as

$$\varsigma_Q = \frac{1}{n}\int_0^n \widetilde{\varrho}(u)\,\mathrm{d}u. \tag{7.9}$$

From (3.4) and (7.8) it follows that $\mathbf{E}\xi_{1,n}^2 = \varsigma_Q$, and for $j \geq 2$,

$$\mathbf{E}\xi_{j,n}^2 = \varsigma_Q + \frac{(-1)^j}{n}\int_0^n \cos(2\gamma_j u)\widetilde{\varrho}(u)\,\mathrm{d}u.$$

Integrating by parts and taking into account (2.10) one comes to the estimate

$$\left|\mathbf{E}\xi_{j,n}^2 - \varsigma_Q\right| \leq \frac{1}{4\gamma_j^2}\left(2\varrho_*' + \varrho_*''\right) \leq \frac{1}{4j^2}\left(2\varrho_*' + \varrho_*''\right).$$

Therefore, $\mathbf{L}_{1,n}(Q) \leq \varrho_*' + \varrho_*''/2 = \mathbf{L}_{1,n}^*$ and the condition (C_1) holds. Moreover, in view of (2.10), the quantity (7.9) satisfies also the condition (H_0) with the sequences ς_*, σ_n^* and ς_n^* given in the theorem. It remains to verify the conditions (C_2) and (H_1). We have

$$\mathbf{E}\left(\sum_{j\leq 1} x_j \widetilde{\xi}_{j,n}\right)^2 = \frac{1}{n}\sum_{i\leq 1}\sum_{j\leq 1} x_i x_j \mathbf{E}\widetilde{I}_n(\phi_i)\widetilde{I}_n(\phi_j), \tag{7.10}$$

where $\widetilde{I}_n(f)$ is defined in (4.10). By Ito's formula one can calculate that for any bounded $\mathbb{R}_+ \to \mathbb{R}$ functions f and g

$$\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g) = 2\widetilde{\tau}_{f,g}^2(n) + \lambda \mathbf{E} Y_1^4 \int_0^n f^2(t) g^2(t) \varrho_2^4(t)\,\mathrm{d}t, \tag{7.11}$$

where $\widetilde{\tau}_{f,g}(t) = \int_0^t f(u) g(u)\widetilde{\varrho}(u)\,\mathrm{d}u$. Integrating by parts yelds

$$\widetilde{\tau}_{f,g}(n) = \widetilde{\varrho}(0)\int_0^n f(u) g(u)\,\mathrm{d}u + \int_0^n \frac{\mathrm{d}\widetilde{\varrho}(u)}{\mathrm{d}u}\left(\int_u^n f(s) g(s)\,\mathrm{d}s\right)\mathrm{d}u.$$

Therefore, for $i \neq j$

$$\widetilde{\tau}_{\phi_i,\phi_j}(n) = -\int_0^n \frac{\mathrm{d}\widetilde{\varrho}(u)}{\mathrm{d}u}\left(\int_0^{\{u\}} \phi_i(s)\phi_j(s)\,\mathrm{d}s\right)\mathrm{d}u,$$

where $\{u\}$ is the fractional part of u. By the condition (2.10) $|\widetilde{\tau}_{\phi_i,\phi_j}(n)| \leq l_{1,n}$, where the $l_{1,n}$ is defined in (3.25). If $i = j$, one has

$$\widetilde{\tau}_{\phi_i,\phi_i}(n) = \widetilde{\varrho}(0) n + \int_0^n \frac{\mathrm{d}\widetilde{\varrho}(u)}{\mathrm{d}u}\left(\int_u^n \phi_i^2(s)\,\mathrm{d}s\right)\mathrm{d}u.$$

This, in view of (2.10), implies that $|\widetilde{\tau}_{\phi_i,\phi_i}(n)| \leq l_{2,n} n$, where $l_{2,n}$ is given in (3.25). Substituting (7.11) in (7.10) and using these estimates for $\widetilde{\tau}_{\phi_i,\phi_i}(n)$ one comes to the inequality $\mathbf{L}_{2,n}(Q) \leq \mathbf{L}_{2,n}^*$, where $\mathbf{L}_{2,n}^*$ is given in the theorem. Hence Theorem 3.6.

Appendix

A.1. *Technical lemmas*

Lemma A.1. *For any bounded left-continuous* $[0,\infty)\to\infty$ *functions* f *and* g *and* $-\infty<a\le 0$

$$\|\tau_{f,g}\|_{*,n}\le 4\varpi^*_{f,g}. \tag{A.1}$$

Proof. First, we note that

$$\int_0^t f(s)\varepsilon_g(s)\,\mathrm{d}s = a\int_0^t \mathrm{e}^{av}\left(\int_0^{t-v} f(s+v)g(s)\bigl(1+\mathrm{e}^{2as}\bigr)\,\mathrm{d}s\right)\mathrm{d}v,$$

where $\varepsilon_g(s)$ is defined in (4.2). Integrating by parts yields

$$\int_0^{t-v} f(s+v)g(s)\bigl(1+\mathrm{e}^{2as}\bigr)\,\mathrm{d}s = \bigl(1+\mathrm{e}^{2a(t-v)}\bigr)\int_0^{t-v} f(s+v)g(s)\,\mathrm{d}s - 2a\int_0^{t-v}\mathrm{e}^{2as}\left(\int_0^s f(z+v)g(z)\,\mathrm{d}z\right)\mathrm{d}s.$$

Taking into account the definition (4.11), this integral can be estimated as

$$\left|\int_0^{t-v} f(s+v)g(s)\bigl(1+\mathrm{e}^{2as}\bigr)\,\mathrm{d}s\right|\le 3\varpi^*_{f,g}.$$

Therefore,

$$\left|\int_0^t f(s)\varepsilon_g(s)\,\mathrm{d}s\right|\le 3\varpi^*_{f,g}\quad\text{and}\quad\left|\int_0^t \varepsilon^*_{f,g}(s)\,\mathrm{d}s\right|\le 6\varpi^*_{f,g}.$$

This implies the inequality (A.1). Hence Lemma A.1. □

Lemma A.2. *For any bounded left-continuous* $[0,+\infty)\to\mathbb{R}$ *functions* f *and* g *the equality* (4.24) *holds.*

Proof. Similarly, we find

$$\mathbf{E}\bigl[K(f),K(g)\bigr]_t=\int_0^t\bigl(\widetilde{\varrho}^2\tau^*_{f,g}(s)+\varrho_3 f(s)g(s)\bigr)\,\mathrm{d}s, \tag{A.2}$$

where the function $\tau^*_{f,g}(s)$ is defined in (4.5). Taking into account (4.17) and applying the Ito's formula one gets

$$\mathrm{d}\mathbf{E}V_t(f)V_t(g)=2a\mathbf{E}V_t(f)V_t(g)\,\mathrm{d}t+\bigl(\widetilde{\varrho}^2\tau^*_{f,g}(t)+\varrho_3 f(t)g(t)\bigr)\,\mathrm{d}t + a\bigl(g(t)\mathbf{E}V_t(f)\widetilde{\xi}_t+f(t)\mathbf{E}V_t(g)\widetilde{\xi}_t\bigr)\,\mathrm{d}t.$$

To calculate $\mathbf{E}V_t(f)\widetilde{\xi}_t$, we put $g=1$ in this equality. Since according to (4.23) $a\widetilde{\varrho}^2\tau^*_{f,1}(t)+a\varrho_3 f(t)=\kappa_f(t)$, therefore one gets

$$a\mathbf{E}V_t(f)\widetilde{\xi}_t=\int_0^t \mathrm{e}^{3a(t-s)}\bigl(f(s)a^2\mathbf{E}\widetilde{\xi}^2_s+\kappa_f(s)\bigr)\,\mathrm{d}s=A_f(t). \tag{A.3}$$

Thus

$$\mathbf{E}V_t(f)V_t(g)=\int_0^t \mathrm{e}^{2a(t-s)}\bigl(g(s)A_f(s)+f(s)A_g(s)\bigr)\,\mathrm{d}s + \int_0^t \mathrm{e}^{2a(t-s)}\bigl(\widetilde{\varrho}^2\tau^*_{f,g}(s)+\varrho_3 f(s)g(s)\bigr)\,\mathrm{d}s.$$

Hence Lemma A.2. □

Further we will need the following result.

Lemma A.3. *Let υ be a continuously differentiable $\mathbb{R}\to\mathbb{R}$ function. Then, for $n\ge 1$, $\alpha>0$ and any integrated $\mathbb{R}\to\mathbb{R}$ function Ψ,*

$$\sup_{0\le t\le n}\left|\int_0^t \mathrm{e}^{-\alpha(t-s)}\Psi(s)\upsilon(s)\,\mathrm{d}s\right|\le \varpi_{1,\Psi}\left(2\|\upsilon\|_{*,n}+\frac{\|\dot\upsilon\|_{*,n}}{\alpha}\right).$$

Proof. One obtains this inequality integrating by parts. □

Lemma A.4. *For any mesurable $[0,+\infty)\to\mathbb{R}$ functions f and g with $\|f\|_{*,n}\le 1$ and $\|g\|_{*,n}\le 1$, for $-\infty<a\le 0$ and $n\ge 1$*

$$\sup_{0\le t\le n}\left|a\int_0^t \mathrm{e}^{2a(t-s)}g(s)A_f(s)\,\mathrm{d}s\right|\le 3\widetilde{\varrho}^2\varpi^*_{f,g}+\varrho_3.$$

Proof. One can represent the function $A_f(t)$ as

$$A_f(t)=\int_0^t \mathrm{e}^{3a(t-s)}f(s)\upsilon(s)\,\mathrm{d}s+\widetilde{\varrho}^2\int_0^t \mathrm{e}^{3a(t-s)}\varepsilon_f(s)\,\mathrm{d}s, \tag{A.4}$$

where $\upsilon(s)=a^2\mathbf{E}\widetilde{\xi}_s^2+\widetilde{\varrho}^2(\mathrm{e}^{2as}-1)+a\varrho_3$. From here and (4.22) we have

$$\upsilon(s)=a\upsilon_1(s)+\upsilon_2(s) \tag{A.5}$$

with

$$\upsilon_1(s)=\frac{\varrho_3}{4}\left(\mathrm{e}^{4at}+3\right)\quad\text{and}\quad \upsilon_2(s)=\frac{\widetilde{\varrho}^2}{2}\mathrm{e}^{4at}-\frac{\widetilde{\varrho}^2}{2}.$$

It will be noted that $\|\upsilon_1\|_{*,n}\le\varrho_3$ and

$$\sup_{-\infty<a\le 0}\left(2\|\upsilon_2\|_{*,n}+\frac{\|\dot\upsilon_2\|_{*,n}}{2|a|}\right)\le 2\widetilde{\varrho}^2. \tag{A.6}$$

Further we have

$$a\int_0^t \mathrm{e}^{2a(t-s)}g(s)A_f(s)\,\mathrm{d}s=J_1(t)+J_2(t)+\widetilde{\varrho}^2J_3(t),$$

where

$$J_1(t)=a^2\int_0^t \mathrm{e}^{2a(t-s)}g(s)\left(\int_0^s \mathrm{e}^{3a(s-u)}f(u)\upsilon_1(u)\,\mathrm{d}u\right)\mathrm{d}s,$$
$$J_2(t)=a\int_0^t \mathrm{e}^{2a(t-s)}g(s)\left(\int_0^s e^{3a(s-u)}f(u)\upsilon_2(u)\,\mathrm{d}u\right)\mathrm{d}s,$$
$$J_3(t)=a\int_0^t \mathrm{e}^{2a(t-s)}g(s)\left(\int_0^s \mathrm{e}^{3a(s-u)}\varepsilon_f(u)\,\mathrm{d}u\right)\mathrm{d}s.$$

From here it follows that

$$\|J_1\|_{*,n}\le\|\upsilon_1\|_{*,n}\le\varrho_3.$$

The second integral $J_2(t)$ can be rewritten as

$$J_2(t) = a \int_0^t \mathrm{e}^{3au} \left(\int_0^{t-u} \mathrm{e}^{2a(t-u-s)} g(s+u) f(s) \upsilon_2(s) \, \mathrm{d}s \right) \mathrm{d}u.$$

By Lemma A.3 and (A.6) we obtain that for any $0 \le z \le n$ and $0 \le u \le n - z$

$$\left| \int_0^z \mathrm{e}^{2a(z-s)} \upsilon_2(s) g(s+u) f(s) \, \mathrm{d}s \right| \le 2\widetilde{\varrho}^2 \varpi^*_{f,g}.$$

Therefore, $\|J_2\|_{*,n} \le 2\widetilde{\varrho}^2 \varpi^*_{f,g}/3$. Similarly, one gets $\|J_3\|_{*,n} \le 2\varpi^*_{g,\varepsilon_f}/3$. To estimate the quantity ϖ_{g,ε_f} defined in (4.11), we note that for $0 \le v \le n$ and $0 \le t \le n - v$

$$\int_0^t g(s+v) \varepsilon_f(s) \, \mathrm{d}s = a \int_0^t \mathrm{e}^{ax} \Theta_{g,f}(t-x, v+x) \, \mathrm{d}x, \tag{A.7}$$

where $\Theta_{g,f}(t,v) = \int_0^t g(s+v) f(s)(1+\mathrm{e}^{2as}) \, \mathrm{d}s$. Denoting

$$\Upsilon_{g,f}(s,u) = \int_0^s g(r+u) f(r) \, \mathrm{d}r, \tag{A.8}$$

we represent the function $\Theta_{g,f}(t,v)$ as

$$\Theta_{g,f}(t,v) = \big(1+\mathrm{e}^{2at}\big) \Upsilon_{g,f}(t,v) - 2a \int_0^t \mathrm{e}^{2as} \Upsilon_{g,f}(s,v) \, \mathrm{d}s.$$

Therefore

$$\max_{0 \le v \le n} \max_{0 \le t \le n-v} \big| \Theta_{g,f}(t,v) \big| \le 3\varpi^*_{f,g}.$$

In view of (A.7), one gets $\varpi_{g,\varepsilon_f} \le 3\varpi^*_{f,g}$ and $\|J_3\|_{*,n} \le 2\varpi^*_{f,g}$. Hence Lemma A.4. $\square$

Lemma A.5. *For any mesurable* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g *with* $\|f\|_{*,n} \le 1$ *and* $\|g\|_{*,n} \le 1$, *for* $-\infty < a \le 0$ *and* $n \ge 1$

$$\|\widetilde{H}_{f,g}\|_{*,n} \le 16\widetilde{\varrho}^2 \varpi^*_{f,g} + 5\varrho_4, \tag{A.9}$$

where $\varrho_4 = \varrho_3 + \widetilde{\varrho}^2$.

Proof. First we represent the function $\widetilde{H}_{f,g}$ as

$$\widetilde{H}_{f,g}(t) = \widetilde{H}^{(1)}_{f,g}(t) + \widetilde{\varrho}^2 \widetilde{H}^{(2)}_{f,g}(t) + \varrho_3 \widetilde{H}^{(3)}_{f,g}(t),$$

where $\widetilde{H}^{(1)}_{f,g}(t) = 2a \int_0^t \mathrm{e}^{2a(t-s)}(g(s)A_f(s) + f(s)A_g(s)) \, \mathrm{d}s$, $\widetilde{H}^{(2)}_{f,g}(t) = 2a \int_0^t \mathrm{e}^{2a(t-s)} \tau^*_{f,g}(s) \, \mathrm{d}s$ and $\widetilde{H}^{(3)}_{f,g}(t) = 2a \int_0^t \mathrm{e}^{2a(t-s)} f(s) g(s) \, \mathrm{d}s$. Lemma A.4 implies directly

$$\|\widetilde{H}^{(1)}_{f,g}\|_{*,n} \le 12\widetilde{\varrho}^2 \varpi^*_{f,g} + 4\varrho_3 \|f\|_{*,n} \|g\|_{*,n}.$$

The next summand can be represented as

$$\widetilde{H}^{(2)}_{f,g}(t) = \widetilde{\tau}_{f,g}(t) + \widetilde{\epsilon}_{f,g}(t) + 2a \int_0^t \mathrm{e}^{2a(t-s)} f(s) g(s) \big(\mathrm{e}^{2as} - 1\big) \, \mathrm{d}s,$$

where $\widetilde{\tau}_{f,g}(t) = 2a\int_0^t e^{2a(t-s)}\tau_{f,g}(s)\,ds$ and $\widetilde{\epsilon}_{f,g}(t) = 2a\int_0^t e^{2a(t-s)}\epsilon^*_{f,g}(s)\,ds$. From (A.1) it follows that $\|\widetilde{\tau}_{f,g}\|_{*,n} \le 4\varpi^*_{f,g}$. Now taking into account that $\|\epsilon^*_f\|_{*,n} \le 2\|f\|_{*,n}$, we obtain that $\|\widetilde{\epsilon}_{f,g}\|_{*,n} \le 4\|f\|_{*,n}\|g\|_{*,n}$. Therefore, $\|\widetilde{H}^{(2)}_{f,g}\|_{*,n} \le 4\varpi^*_{f,g} + 5\|f\|_{*,n}\|g\|_{*,n}$ and $\|\widetilde{H}^{(3)}_{f,g}\|_{*,n} \le \|f\|_{*,n}\|g\|_{*,n}$. Hence Lemma A.5. □

Lemma A.6. *For any mesurable* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g *with* $\|f\|_{*,n} \le 1$ *and* $\|g\|_{*,n} \le 1$, *for* $-\infty < a \le 0$ *and* $n \ge 1$

$$\sup_{0\le t\le n}\left|a\int_0^t e^{a(t-s)} f(s)\mathcal{H}_g(s)\,ds\right| \le 8\widetilde{\varrho}^2\varpi^*_{1,g} + 2\varrho_4. \tag{A.10}$$

Proof. It is obviously, that for $0 \le t \le n$

$$\left|a\int_0^t e^{a(t-s)} f(s)\mathcal{H}_g(s)\,ds\right| \le \|f\|_{*,n}\|\mathcal{H}_g\|_{*,n} \le \|\widetilde{H}_{1,g}\|_{*,n}/2.$$

Now applying Lemma A.5 one comes to the inequality (A.10). Hence Lemma A.6. □

Lemma A.7. *For any mesurable* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g *with* $\|f\|_{*,n} \le 1$ *and* $\|g\|_{*,n} \le 1$, *for* $-\infty < a \le 0$ *and* $n \ge 1$

$$\sup_{0\le t\le n}\left|a\int_0^t e^{a(t-s)} f(s)\widetilde{U}_{1,g}(s)\,ds\right| \le 8\widetilde{\varrho}^2\varpi^*_{f,g} + \varrho_3/2.$$

Proof. The function $U_{1,g}(t)$ can be represented as

$$U_{1,g}(t) = \frac{2\widetilde{\varrho}^2}{a} g(t)\varepsilon_g(t) + \varrho_3 g^2(t).$$

From here one obtains

$$a\int_0^t e^{a(t-s)} f(s)\widetilde{U}_{1,g}(s)\,ds = 2a\widetilde{\varrho}^2 I_1(t) + \varrho_3 a^2 I_2(t),$$

where

$$I_1(t) = \int_0^t e^{a(t-s)} f(s)\left(\int_0^s e^{2a(s-r)} g(r)\varepsilon_g(r)\,dr\right) ds,$$
$$I_2(t) = \int_0^t e^{a(t-s)} f(s)\left(\int_0^s e^{2a(s-r)} g^2(r)\,dr\right) ds.$$

Denoting $\Gamma_{f,g}(t,v) = \int_0^t e^{a(t-s)} f(s+v)g(s)\varepsilon_g(s)\,ds$, one has

$$I_1(t) = \int_0^t e^{2av}\Gamma_{f,g}(t-v,v)\,dv.$$

Since $|\varepsilon_g(t)| \le 2\|g\|_{*,n} \le 2$ and

$$\dot{\varepsilon}_g(t) = a\varepsilon_g(t) + af(t)\bigl(1 + e^{2at}\bigr),$$

Lemma A.3 one can apply to estimate the function $\Gamma_{f,g}(t,x)$ as

$$\sup_{0\le t\le n}\sup_{0\le v\le t}\bigl|\Gamma_{f,g}(t-v,v)\bigr| \le 8\varpi^*_{f,g}.$$

Therefore, $\|aI_1\|_{*,n} \le 4\varpi^*_{f,g}$ and $a^2\|I_2\|_{*,n} \le \|f\|_{*,n}\|g\|^2_*/2 \le 1/2$. Hence Lemma A.7. □

Lemma A.8. *For any mesurable* $[0, +\infty) \to \mathbb{R}$ *functions* f *and* g *with* $\|f\|_{*,n} \le 1$ *and* $\|g\|_{*,n} \le 1$, *for* $-\infty < a \le 0$ *and* $n \ge 1$

$$\|\widetilde{U}_{f,g}\|_{*,n} \le \widetilde{\varrho}^2 \big(\varpi^*_{f,g} + \varpi^*_{1,g}\big) + \varrho_3/2. \tag{A.11}$$

Proof. This result follows from the estimate

$$\|\widetilde{U}_{f,g}\|_{*,n} \le \|U_{f,g}\|_{*,n}/2 \le \widetilde{\varrho}^2 \big(\|\tau_{f,g}\|_{*,n} + \|\tau_{1,g}\|_{*,n}\big) + \varrho_3/2,$$

and Lemma A.1. □

A.2. Property of the Fourier coefficients

Lemma A.9. *Suppose that the function* S *in* (1.1) *is differentiable and satisfies the condition* (3.21). *Then the Fourier coefficients* (3.2) *satisfy the inequality*

$$\sup_{l \ge 2} l \sum_{j=l}^{\infty} \theta_j^2 \le 4|\dot{S}|_1^2.$$

Proof. In view of (3.1), one has

$$\theta_{2p} = -\frac{1}{\sqrt{2}\pi p} \int_0^1 \dot{S}(t) \sin(2\pi p t)\, \mathrm{d}t$$

and

$$\begin{aligned}\theta_{2p+1} &= \frac{1}{\sqrt{2}\pi p} \int_0^1 \dot{S}(t)\big(\cos(2\pi p t) - 1\big)\, \mathrm{d}t \\ &= -\frac{\sqrt{2}}{\pi p} \int_0^1 \dot{S}(t) \sin^2(\pi p t)\, \mathrm{d}t, \qquad p \ge 1.\end{aligned}$$

From here, it follows that $\theta_j^2 \le 2|\dot{S}|_1^2/j^2$ for any $j \ge 2$. Therefore, taking into account that $\sup_{l \ge 2} l \sum_{j \ge l} j^{-2} \le 2$, we arrive at the desired result. □

Acknowledgments

The authors are grateful to the anonymous referee for very helpful comments and for suggesting to add an example with nonstandard rate of convergence.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19** (1974) 716–723. MR0423716

[2] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial mathematics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2001) 167–241. MR1841412

[3] A. Barron, L. Birgé and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** (1999) 301–415. MR1679028

[4] J. Bertoin. *Lévy Processes.* Cambridge Univ. Press, Cambridge, 1996. MR1406564

[5] J. Brua. Asymptotically efficient estimators for nonparametric heteroscedastic regression models. *Stat. Methodol.* **6** (2009) 47–60. MR2655538

[6] L. Delong and C. Klüppelberg. Optimal investment and consumption in a Black–Scholes market with Lévy driven stochastic coefficients. *Ann. Appl. Probab.* **18** (2008) 879–908. MR2418232

[7] D. Fourdrinier and S. M. Pergamenshchikov. Improved selection model method for the regression with dependent noise. *Ann. Inst. Statist. Math.* **59** (2007) 435–464. MR2388801

[8] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics*. Springer, New York, 2006. MR2229687

[9] L. I. Galtchouk and S. M. Pergamenshchikov. Nonparametric sequential estimation of the drift in diffusion processes. *Math. Methods Statist.* **13** (2004) 25–49. MR2078310

[10] L. I. Galtchouk and S. M. Pergamenshchikov. Asymptotically efficient estimates for non parametric regression models. *Statist. Probab. Lett.* **76** (2006) 852–860. MR2266100

[11] L. I. Galtchouk and S. M. Pergamenshchikov. Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *J. Nonparametr. Stat.* **21** (2009) 1–16. MR2483856

[12] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *J. Korean Statist. Soc.* **38** (2009) 305–322. MR2582478

[13] L. Galtchouk and S. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection. Preprint, 2009. Available at http://hal.archives-ouvertes.fr/hal-00326910/fr/.

[14] S. M. Goldfeld and R. E. Quandt. *Nonlinear Methods in Econometrics*. *Contributions to Economic Analysis* **77**. North-Holland, London, 1972. With a contribution by Dennis E. Smallwood. MR0343521

[15] A. Kneip. Ordered linear smoothers. *Ann. Statist.* **22** (1994) 835–866. MR1292543

[16] V. V. Konev and S. M. Pergamenshchikov. Sequential estimation of the parameters in a trigonometric regression model with the Gaussian coloured noise. *Statist. Inference Stoch. Process.* **6** (2003) 215–235. MR2018818

[17] V. V. Konev and S. M. Pergamenshchikov. General model selection estimation of a periodic regression with a Gaussian noise. *Ann. Inst. Statist. Math.* **62** (2010) 1083–1111. MR2729155

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

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

[20] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall, London, 1996. MR1422250

[21] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes* **1**. Springer, New York, 1987. MR0959133

[22] C. L. Mallows. Some comments on C_p. *Technometrics* **15** (1973) 661–675.

[23] P. Massart. A non-asymptotic theory for model selection. In *European Congress of Mathematics*. Eur. Math. Soc., Zürich, 2005. MR2185752

[24] M. Nussbaum. Spline smoothing in regression models and asymptotic efficiency in $\mathbf{L}_2$. *Ann. Statist.* **13** (1985) 984–997. MR0803753

[25] M. S. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inf. Transm.* **16** (1980) 52–68. MR0624591