



HAL
open science

Robust model selection for a semimartingale continuous time regression from discrete data

Victor Konev, Serguei Pergamenchtchikov

► **To cite this version:**

Victor Konev, Serguei Pergamenchtchikov. Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic Processes and their Applications*, 2015, 125, pp.294 - 326. 10.1016/j.spa.2014.08.003 . hal-02334868

HAL Id: hal-02334868

<https://hal.science/hal-02334868>

Submitted on 29 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust model selection for a semimartingale continuous time regression from discrete data ^{*}

Konev Victor [†] and Pergamenchtchikov Serguei [‡]

Abstract

The paper considers the problem of estimating a periodic function in a continuous time regression model observed under a general semimartingale noise with an unknown distribution in the case when continuous observation cannot be provided and only discrete time measurements are available. Two specific types of noises are studied in detail: a non - Gaussian Ornstein - Uhlenbeck process and a time - varying linear combination of a Brownian motion and compound Poisson process. We develop new analytical tools to treat the adaptive estimation problems from discrete data. A lower bound for the frequency sampling, needed for the efficiency of the procedure constructed by discrete observations, has been found. Sharp non-asymptotic oracle inequalities for the robust quadratic risk have been derived. New convergence rates for the efficient procedures have been obtained. An example of the regression with a martingale noise exhibits that the minimax robust convergence rate may be both higher or lower as compared with the minimax rate for the "white noise" model. The results of Monte-Carlo simulations are given.

Keywords: Semimartingale regression; Estimation from discrete data; Robust risk estimation; Model selection; Sharp oracle inequalities.

AMS 2000 Subject Classifications: Primary: 62G08; Secondary: 62G05

^{*}This study was partially supported by Russian Science Fondation (research project No. 14-49-00079) and by the National Research Tomsk State University.

[†]Department of Applied Mathematics and Cybernetics, Tomsk State University and Department of High Mathematics and Mathematical Physics Tomsk Polytechnical University, Tomsk, Russia, e-mail: vvkonev@mail.tsu.ru

[‡]Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen, France and National Research University - Higher School of Economics, Laboratory of Quantitative Finance, Moscow, Russia, email: serge.pergamenchtchikov@univ-rouen.fr.

1 Introduction

Consider a regression model in continuous time

$$dy_t = S(t)dt + d\xi_t, \quad 0 \leq t \leq n, \quad (1.1)$$

where S is an unknown function which belongs to the linear space \mathcal{V}_1 of 1 - periodic $\mathbb{R} \rightarrow \mathbb{R}$ cadlag functions; $\xi = (\xi_t)_{t \geq 0}$ is an unobservable semimartingale noise with the values in the Skorokhod space $\mathcal{D}[0, n]$ such that for each function f from $\mathcal{L}_2[0, n]$ the stochastic integral

$$I_n(f) = \int_0^n f(s)d\xi_s \quad (1.2)$$

is well defined and has the properties:

$$\mathbf{E}_Q I_n(f) = 0 \quad \text{and} \quad \mathbf{E}_Q I_n^2(f) \leq \varkappa_Q \int_0^n f^2(s)ds, \quad (1.3)$$

where $\varkappa_Q > 0$ is some positive constant depending on the noise distribution Q on the space $\mathcal{D}[0, n]$; \mathbf{E}_Q denotes the expectation under the distribution Q . The noise distribution Q is unknown and assumed to belong to some distribution family \mathcal{Q}_n specified below. All necessary tools concerning the stochastic calculus can be found, for example, in [18].

The problem is to estimate the unknown function S in the model (1.1) on the basis of observations

$$(y_{t_j})_{0 \leq j \leq np}, \quad t_j = j\Delta, \quad \Delta = \frac{1}{p}, \quad (1.4)$$

where $p \geq 3$ is an odd number depending on n . Such an assumption about discrete time observations arises if the continuous observation of the process (1.1) cannot be provided. There are many papers devoted to the similar nonparametric estimation problems for the regression model (1.1) and other continuous time processes on the basis of the discrete observations (1.4). Gobet, Hoffmann and Reiss [14] and Comte, Genon-Catalot and Rozenholc [5] studied the problem of estimating the coefficients of a diffusion process from discrete data. Hoffmann, Munk and Schmidt-Hieber [15] investigate nonparametric estimation for diffusion coefficients from discrete data, when the observation are blurred by an additive noise. Comte and Genon-Catalot

[4] studied nonparametric estimation problem for pure jumps Lévy process in the model (1.1) based on discrete time observations.

In this paper we consider the estimation problem in an adaptive setting, i.e. when the regularity of S is unknown. The quality of an estimate \tilde{S}_n (any one periodic $\mathbb{R} \rightarrow \mathbb{R}$ function constructed from discrete data (1.4)) will be measured with the quadratic risk

$$\mathcal{R}_Q(\tilde{S}_n, S) = \mathbf{E}_Q \|\tilde{S}_n - S\|^2, \quad (1.5)$$

where $\|f\|^2 = \int_0^1 f^2(t)dt$. Since the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

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

which enables one to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate \tilde{S}_n for all distributions in the family \mathcal{Q}_n . It will be noted that similar criteria in the problems of nonparametric estimation were used in a number of papers (see, for example, the papers by Galtchouk and Pergamenchtchikov in [10], by the authors [24] and the references therein).

The goal of this paper is to develop a model selection method for estimating a continuous time semimartingale regression (1.1) from discrete data (1.4). The origin of the model selection method goes back to early seventies with the pioneering papers by Akaike [1] and Mallows [25] who suggested to use penalizing in a log-likelihood type criterion. Barron, Birgé, Massart [3], Massart [26] and Kneip [19] developed a non-asymptotic model selection method which enables one to derive non-asymptotic oracle inequalities for the nonparametric regression models with Gaussian disturbances. The model selection procedures for the regression schemes with dependent noises and unknown distributions were studied by Comte and Genon-Catalot [4], Fourdrinier and Pergamenchtchikov [8], Galtchouk and Pergamenchtchikov [9], and by the authors (see, [23], [21] for details and further references). The interest to the model selection procedures can be explained by the fact that they provide adaptive solutions to the non-asymptotic regression models by the technique of sharp non-asymptotic oracle inequalities. As is known (see, for example, [3]), the oracle inequality gives an upper bound for the risk via the minimal risk corresponding to a chosen family of estimates. In constructing an adaptive model selection procedure, in this paper, we will use

the approach close to that of the papers [12], [11], [13], developed for a heteroscedastic regression model in discrete time. The key idea of the method in [12] is to combine the Barron-Birgé-Massart non-asymptotic penalization method [3] and the Pinsker weighted least squares method of minimizing the asymptotic risk [27], [28]. The advantage of this method is two-fold: it deals with a robust risk and enables one to construct asymptotically efficient procedures.

The rest of the paper is organized as follows. Section 2 gives two specific examples of the noise in the model (1.1). First, a non - Gaussian Ornstein - Uhlenbeck process driven by a finite intensity Lévy process and second - a deterministic but time - varying linear combination of a Brownian motion and a compound Poisson process. In Section 3 we construct a model selection procedure based on the weighted least squares estimators and, under general moment conditions on the distribution ξ in (1.1), obtain sharp non-asymptotic oracle inequalities for the risks (1.5) and (1.6). Moreover, we check these conditions for the examples given in Section 2. Section 4 illustrates the performance of the proposed model selection procedure through numerical simulations. It turns out that the sample convergence rate is close to that obtained in [12] for heteroscedastic regression models in the discrete time. Section 5 deals with the asymptotic properties of the proposed procedure under the additional assumption that the unknown function S in (1.1) belongs to a Sobolev ball. In this section we obtain an asymptotic sharp lower bound for the robust risk (1.6) and we calculate the Pinsker constant for the model (1.1). Then we find the lower bound for the frequency sampling p which is needed for constructing efficient estimation procedures from discrete data. It is shown that the limit of the risk (1.6) for the proposed model selection procedure, normalized by the minimax rate, equals the Pinsker constant, i.e. the procedure is efficient. Section 6 gives the proofs of all the theorems. In the Appendix some technical results are established.

2 Examples

2.1 Non-Gaussian Ornstein-Uhlenbeck process

First we consider an example of the disturbance $(\xi_t)_{t \geq 0}$ in (1.1) given by a non-Gaussian Ornstein-Uhlenbeck process with the Lévy subordinator. Such processes are used in the financial Black-Scholes market models with stepwise

randomly evolving coefficients (see [2] for details and further references). Let the noise process in (1.1) obey the equation

$$d\xi_t = a\xi_t dt + du_t, \quad \xi_0 = 0, \quad (2.1)$$

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 of the form

$$z_t = \sum_{j=1}^{N_t} Y_j \quad (2.2)$$

where $(N_t)_{t \geq 0}$ 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

$$\mathbf{E}Y_1 = 0, \quad \mathbf{E}Y_1^2 = 1 \quad \text{and} \quad \mathbf{E}Y_1^4 < \infty. \quad (2.3)$$

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

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

The parameters λ , a , ϱ_1 and ϱ_2 of the process (2.1) are assumed to satisfy the following conditions

$$\left\{ \begin{array}{l} -a_* \leq a \leq 0, \quad \lambda \geq \lambda_*, \\ \varrho_{min} \leq \varrho_1 \leq \varrho_{max}, \quad \varrho_{min} \leq \varrho^* \leq \varrho_{max}, \end{array} \right. \quad (2.4)$$

where $\varrho^* = \varrho_1^2 + \lambda\varrho_2^2$. For this noise model the family \mathcal{Q}_n consists of all distributions on the Skorokhod space $\mathcal{D}[0, n]$ of the process (2.1) with the parameters satisfying the inequalities (2.4) with unknown bounds λ_* , a_* , ϱ_{min} and ϱ_{max} .

Remark 2.1. *The estimation problem from the continuous data for the models of type (1.1), (2.1) has been studied in a parametric setting in [6], [16], [17] and [20].*

2.2 Martingale noise

Next we consider a martingale noise in (1.1) obeying the equation

$$d\xi_t = \sqrt{\varrho_1(t)}dw_t + \sqrt{\varrho_2(t)}dz_t, \quad (2.5)$$

where ϱ_1 and ϱ_2 are positive two times continuously differentiable $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ nonrandom functions; the process $(z_t)_{t \geq 0}$ is as defined in (2.2)–(2.3). We assume that $\lambda \geq \lambda_*$ for some unknown $\lambda_* > 0$. Moreover, denoting

$$\tilde{\varrho}(t) = \varrho_1(t) + \lambda \varrho_2(t), \quad (2.6)$$

we assume that, there exist $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ continuous functions $\varrho_{min}(\cdot)$ and $\varrho_{max}(\cdot)$ such that for $t \geq 0$

$$\varrho_{min}(t) \leq \tilde{\varrho}(t) \leq \varrho_{max}(t) \quad (2.7)$$

and, for any $\delta > 0$,

$$\lim_{t \rightarrow \infty} t^\delta \varrho_{min}(t) = +\infty \quad \text{and} \quad \lim_{t \rightarrow \infty} \frac{\varrho_{max}(t)}{t^\delta} = 0. \quad (2.8)$$

In addition, let the first two derivatives of the functions $\tilde{\varrho}(\cdot)$ satisfy the following conditions for some positive constants ϱ'_* and ϱ''_*

$$\sup_{t \geq 0} (t+1) \left| \frac{d}{dt} \tilde{\varrho}(t) \right| \leq \varrho'_*, \quad \sup_{t \geq 0} \left| \frac{d^2}{dt^2} \tilde{\varrho}(t) \right| \leq \varrho''_. \quad (2.9)$$

In this case \mathcal{Q}_n is the family of all distributions of the process (2.5) on $\mathcal{D}[0, n]$ satisfying the conditions (2.7) and (2.9) for some fixed unknown parameters $\lambda_* > 0$, $\varrho_{min}(\cdot)$, $\varrho_{max}(\cdot)$, ϱ'_* and ϱ''_* .

3 Oracle inequalities

In this section we construct a model selection procedure for estimating a function S in (1.1) by the discrete time observations (1.4) and establish the oracle inequalities for its risk. Let $(\phi_j)_{j \geq 1}$ be the standard trigonometric basis in $\mathcal{L}_2[0, 1]$ defined as

$$\phi_1 = 1, \quad \phi_j(x) = \begin{cases} \sqrt{2} \cos(\varpi_j x) & \text{for even } j \geq 2; \\ \sqrt{2} \sin(\varpi_j x) & \text{for odd } j \geq 3, \end{cases} \quad (3.1)$$

where $\varpi_j = 2\pi[j/2]$, $[x]$ denotes the integer part of x . The restrictions of the functions $\{\phi_1, \dots, \phi_p\}$, $p \geq 3$, on the sampling lattice

$$\mathcal{T}_p = \{t_1, \dots, t_p\}, \quad t_j = \frac{j}{p},$$

form an orthonormal basis in the Hilbert space $\mathbb{R}^{\mathcal{T}_p}$ with the inner product

$$(x, y)_p = \frac{1}{p} \sum_{j=1}^p x(t_j)y(t_j) \quad \text{for } x, y \in \mathbb{R}^{\mathcal{T}_p} \quad (3.2)$$

and the norm $\|x\|_p = \sqrt{(x, x)_p}$. It is clear that the space $\mathbb{R}^{\mathcal{T}_p}$ is isometric to \mathbb{R}^p . One can check directly that for any odd p

$$(\phi_j, \phi_i)_p = \chi_{\{i=j\}}. \quad (3.3)$$

This implies that the function S on the lattice \mathcal{T}_p coincides with its discrete Fourier transformation, i.e.

$$S(t) = \sum_{j=1}^p \theta_{j,p} \phi_j(t), \quad \text{if } t \in \mathcal{T}_p, \quad (3.4)$$

where $\theta_{j,p} = (S, \phi_j)_p$. The first step in constructing the model selection procedure consists in estimating the coefficients $\theta_{j,p}$ for S in (1.1) from the discrete data by the formulae

$$\widehat{\theta}_{j,p} = \frac{1}{n} \int_0^n \psi_{j,p}(t) dy_t, \quad \psi_{j,p} = \mathbf{D}_p(\phi_j). \quad (3.5)$$

Here \mathbf{D}_p stands for the linear mapping given by the equation

$$\mathbf{D}_p(f)(t) = \sum_{k=1}^{np} f(t_k) \chi_{\{t_{k-1} < t \leq t_k\}}. \quad (3.6)$$

Furthermore, we note that the system of the functions $(\psi_{j,p})_{1 \leq j \leq p}$ is orthonormal in $\mathcal{L}_2[0, 1]$ because $(\psi_{j,p}, \psi_{i,p}) = \int_0^1 \psi_{j,p}(t) \psi_{i,p}(t) dt = (\phi_j, \phi_i)_p = \chi_{\{i=j\}}$. In the sequel we need the Fourier coefficients for the function S with respect to these functions which can be written as

$$\bar{\theta}_{j,p} = (S, \psi_{j,p}) = \int_0^1 S(t) \psi_{j,p}(t) dt = \theta_{j,p} + h_{j,p}, \quad (3.7)$$

where $h_{j,p} = h_{j,p}(S) = \sum_{l=1}^p \int_{t_{l-1}}^{t_l} \phi_j(t_l)(S(t) - S(t_l)) dt$. Substituting (1.1) in (3.5) yields

$$\widehat{\theta}_{j,p} = \bar{\theta}_{j,p} + \frac{1}{\sqrt{n}} \xi_{j,p} \quad \text{and} \quad \xi_{j,p} = \frac{I_n(\psi_{j,p})}{\sqrt{n}}. \quad (3.8)$$

Besides we will impose the additional conditions on the distribution of the noise ξ_t in (1.1). For any $x = (x_1, \dots, x_p)$ from \mathbb{R}^p , we set

$$\#(x) = \sum_{j=1}^p \mathbf{1}_{\{|x_j|>0\}} \quad \text{and} \quad |x|^2 = \sum_{j=1}^p x_j^2. \quad (3.9)$$

We need the following conditions.

C₁) *There exists a variance proxy $\sigma_Q \geq 0$, which may depend on $n \geq 1$, such that the sequence $\zeta_{j,p} = \mathbf{E}_Q \xi_{j,p}^2 - \sigma_Q$ satisfies the following inequality*

$$\mathbf{L}_{1,Q}(n) = \sup_{p \geq 3} \sup_{x \in \mathcal{H}_1} \left| \sum_{j=1}^p x_j \zeta_{j,p} \right| < \infty$$

where $\mathcal{H}_1 = \{x \in [-1, 1]^p : \#(x) \leq n\}$.

C₂) *Assume that*

$$\mathbf{L}_{2,Q}(n) = \sup_{p \geq 3} \sup_{x \in \mathcal{H}_2} \mathbf{E}_Q \left(\sum_{j=1}^p x_j \tilde{\xi}_{j,p} \right)^2 < \infty$$

where $\mathcal{H}_2 = \{x \in \mathbb{R}^p : |x| \leq 1, \#(x) \leq n\}$ and $\tilde{\xi}_{j,p} = \xi_{j,p}^2 - \mathbf{E}_Q \xi_{j,p}^2$.

Remark 3.1. *As is shown in the paper [24], which considers the estimation problem for the model (1.1) under continuous observations, one needs some stability conditions of the noise variances in (3.8) and the boundedness of their deviations to derive the oracle inequalities for the robust risk. In the estimation problem from discrete data the noise properties depend highly on the frequency of observations as well. The additional supremum in the Conditions in **C₁)** and **C₂)** over $p \geq 3$ assumes that the stability of the noise variances and their deviations hold uniformly in p . Further restrictions on the functionals $\mathbf{L}_{1,Q}(n)$ and $\mathbf{L}_{2,Q}(n)$ will be imposed in Section 3.2.*

As will be shown in the Section 3.3, the conditions **C₁)** and **C₂)** are satisfied for the model (1.1) with the noises (2.1) and (2.5). Now we introduce a weight least squares estimate for $S(t)$ as

$$\widehat{S}_\gamma(t) = \sum_{j=1}^p \gamma(j) \widehat{\theta}_{j,p} \psi_{j,p}(t), \quad (3.10)$$

where $\gamma = (\gamma(1), \dots, \gamma(p))' \in [0, 1]^p$ is the vector of weight coefficients. The model selection procedure will be chosen from a finite family of such estimates $(\widehat{S}_\gamma)_{\gamma \in \Gamma}$. The set of weight sequences Γ will be given below. We will need the following characteristic of this set and the corresponding system of weight sequences

$$\nu_p = \text{card} \Gamma \quad \text{and} \quad \mu_p = \max_{\gamma \in \Gamma} \#(\gamma). \quad (3.11)$$

We assume that

$$\mu_p \leq n. \quad (3.12)$$

Remark 3.2. *The last inequality means that we cannot use more than n nonzero terms in the estimate (3.10). We need this condition because the noise variance in (3.8) tends to zero as n^{-1} .*

Now note that in view of the definition (3.7), the empirical squared error of the estimator (3.10) can be represented as

$$\text{Err}(\gamma) = \|\widehat{S}_\gamma - S\|^2 = \sum_{j=1}^p \gamma^2(j) \widehat{\theta}_{j,p}^2 - 2 \sum_{j=1}^p \gamma(j) \widehat{\theta}_{j,p} \bar{\theta}_{j,p} + \|S\|^2. \quad (3.13)$$

Since the Fourier coefficients $(\bar{\theta}_{j,p})_{j \geq 1}$ are unknown, the weight coefficients $(\gamma(j))_{1 \leq j \leq p}$ cannot be determined by minimizing this quantity. To circumvent this difficulty we replace the terms $\widehat{\theta}_{j,p} \bar{\theta}_{j,p}$ by

$$\tilde{\theta}_{j,p} = \widehat{\theta}_{j,p}^2 - \frac{\widehat{\sigma}_n}{n}, \quad (3.14)$$

where $\widehat{\sigma}_n$ is an estimator for the variance proxy σ_Q in the condition \mathbf{C}_1). We will need the following characteristics of this estimate

$$\mathbf{r}_Q(\widehat{\sigma}_n) = \mathbf{E}_Q |\widehat{\sigma}_n - \sigma_Q| \quad \text{and} \quad \mathbf{r}_n^*(\widehat{\sigma}_n) = \sup_{Q \in \mathcal{Q}_n} \mathbf{E}_Q |\widehat{\sigma}_n - \sigma_Q|. \quad (3.15)$$

For replacing the terms $\widehat{\theta}_{j,p} \bar{\theta}_{j,p}$ by the estimates (3.14) on the right-hand side of the empirical squared error (3.13), one has to pay some penalty. Thus, one comes to the cost function of the form

$$J(\gamma) = \sum_{j=1}^p \gamma^2(j) \widehat{\theta}_{j,p}^2 - 2 \sum_{j=1}^p \gamma(j) \tilde{\theta}_{j,p} + \rho \widehat{P}(\gamma) \quad (3.16)$$

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

$$\widehat{P}(\gamma) = \frac{\widehat{\sigma}_n |\gamma|^2}{n}, \quad (3.17)$$

where $|\gamma|^2 = \sum_{j=1}^p \gamma^2(j)$. If σ_Q is known, one can put $\widehat{\sigma}_n = \sigma_Q$ and

$$\widehat{P}(\gamma) = P_Q(\gamma) = \frac{\sigma_Q |\gamma|^2}{n}. \quad (3.18)$$

Substituting in (3.10) the values of the weight coefficients $\gamma(j)$, minimizing the cost function (3.16), that is

$$\widehat{\gamma} = \operatorname{argmin}_{\gamma \in \Gamma} J(\gamma), \quad (3.19)$$

yields the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\widehat{\gamma}}. \quad (3.20)$$

Our first goal is to obtain the oracle inequalities for the quadratic risk of the estimate (3.20) defined in (1.5). To state the result we introduce the sequence

$$\Psi_Q(n, p) = 2 \left(3\mathbf{L}_{1,Q}(n) + \nu_p \mathbf{L}_{2,Q}(n) (\sigma_Q)^{-1} + \nu_p \varkappa_Q \right). \quad (3.21)$$

Proposition 3.1. *Let the conditions (1.3), \mathbf{C}_1) and \mathbf{C}_2) imposed on the noise distribution $(\xi_t)_{t \geq 0}$ in (1.1) hold. Then, for any $n \geq 2$, $p \geq 3$, $0 < \rho < 1/3$ and any set Γ with the property (3.12), the estimator (3.20) satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_Q(\widehat{S}_*, S) &\leq \frac{1+3\rho}{1-3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_Q(\widehat{S}_\gamma, S) + \frac{1}{n} \frac{\Psi_Q(n, p)}{(1-3\rho)\rho} \\ &\quad + \frac{6\mu_p \mathbf{r}_Q(\widehat{\sigma}_n)}{n(1-3\rho)}. \end{aligned} \quad (3.22)$$

The proof of Proposition 3.1 is given in Section 6.

If the proxy variance σ_Q in the Condition \mathbf{C}_1) is known, then one comes to the following result.

Corollary 3.2. *Suppose that the conditions of Proposition 3.1 hold with known $\sigma_Q > 0$. Then, for any $n \geq 2$, $p \geq 3$, $0 < \rho < 1/3$ and any set Γ with the property (3.12), the estimator (3.20) satisfies the oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\rho}{1-3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_Q(\widehat{S}_\gamma, S) + \frac{1}{n} \frac{\Psi_Q(n, p)}{(1-3\rho)\rho}. \quad (3.23)$$

Remark 3.3. *It should be noted that Comte and Genon-Catalot in [4] use the oracle inequality of type (3.22) to find an adaptive convergence rate for the density estimation problem for the Lévy processes from discrete data. In this paper we will use this inequality to study efficiency properties for the procedure (3.19).*

3.1 Estimation of σ_Q

In this section we will consider the case of an unknown proxy variance σ_Q in the condition \mathbf{C}_1) and derive the oracle inequalities for the continuous time estimate (3.10). We additionally assume that the unknown function $S(t)$ in (1.1) has an absolutely integrable derivative. First we have to estimate σ_Q and find an upper bound for $\mathbf{r}_Q(\widehat{\sigma}_n)$ in (3.22). One can use the following estimate for σ_Q

$$\widehat{\sigma}_n = \frac{n}{\check{p}} \sum_{j=l_0}^{\check{p}} \widehat{\theta}_{j,p}^2 \quad \text{and} \quad \check{p} = \min(p, n), \quad (3.24)$$

where $l_0 \geq 1$ will be specified later. We set $\widehat{\sigma}_n = 0$ for $l_0 > \check{p}$.

Lemma 3.3. *Assume that the conditions (1.3), \mathbf{C}_1) and \mathbf{C}_2) hold and the unknown function $S(t)$ is differentiable for $0 \leq t \leq 1$ and such that for some $r > 0$*

$$\|S\|^2 + \|\dot{S}\|^2 = \int_0^1 (|S(t)|^2 + |\dot{S}(t)|^2) dt \leq r. \quad (3.25)$$

Then, for any $n \geq 1$ and $p \geq 3$, the quantity $\mathbf{r}_Q(\widehat{\sigma}_n)$, defined in (3.15), satisfies the inequality

$$\mathbf{r}_Q(\widehat{\sigma}_n) \leq \mathbf{K}_Q(n) \left(\frac{n}{\check{p}l_0} + \frac{l_0}{\check{p}} + \frac{1}{\sqrt{\check{p}}} \right), \quad (3.26)$$

where $\mathbf{K}_Q(n) = 16r + 8\sqrt{r\kappa_Q} + \mathbf{L}_{1,Q}(n) + \sigma_Q + \sqrt{\mathbf{L}_{2,Q}(n)}$.

The proof of Lemma 3.3 is given in Section 6.

Minimizing the right hand side of the last inequality with respect to l_0 we find the appropriate value for l_0 in (3.24), namely

$$l_0 = \lceil \sqrt{n} \rceil, \quad (3.27)$$

which yields

$$\mathbf{r}_Q(\widehat{\sigma}_n) \leq 3\mathbf{K}_Q(n) \left(\frac{\sqrt{n}}{\check{p}} + \frac{1}{\sqrt{\check{p}}} \right) := 3\mathbf{K}_Q(n) \mathbf{g}_{p,n}^*. \quad (3.28)$$

Combining Proposition 3.1 and Lemma 3.3 leads to the following result.

Proposition 3.4. *Assume that the conditions (1.3), \mathbf{C}_1) and \mathbf{C}_2) hold and the function S satisfies the inequality (3.25) for some $r > 0$. Then, for any $n \geq 1$, $p \geq 3$, $0 < \rho < 1/3$ and any set Γ with the property (3.12), the estimate (3.20) obeys the oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1 + 3\rho}{1 - 3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_Q(\widehat{S}_\gamma, S) + \frac{\mathcal{B}_Q(n, p)}{n(1 - 3\rho)\rho}, \quad (3.29)$$

where $\mathcal{B}_Q(n, p) = \Psi_Q(n, p) + 18\mathbf{K}_Q(n)\mu_p \mathbf{g}_{p,n}^*$.

Remark 3.4. *Note that the oracle inequality (3.29) involves the term $\mathcal{B}_Q(n, p)$ which depends on n and p . Our goal is to find conditions on the noise distribution Q and the frequency p providing the boundedness of this term by any power of n as $n \rightarrow \infty$. As will be seen later, this property implies the efficiency of model selection procedure (3.20). To this end we need to study the robust risks for this estimate.*

3.2 Robust estimation

In order to obtain the oracle inequality for the robust risk (1.6) of the estimate (3.10) we will impose additional conditions on the distribution family \mathcal{Q}_n of the noise ξ in the equation (1.1). Actually these conditions are stipulated also by the further studies of the asymptotic properties of the procedure (3.20) provided that both the number of observation periods n and the observation frequency p tend to infinity. Let \mathcal{P}_n denote the class of all probability measures on the space $\mathcal{D}[0, n]$ and \mathcal{P}_n^* be its subclass defined as

$$\mathcal{P}_n^* = \left\{ Q \in \mathcal{P}_n : \mathbf{L}_{1,Q}(n) \leq \mathbf{L}_{1,n}^*, \mathbf{L}_{2,Q}(n) \leq \mathbf{L}_{2,n}^* \right\}, \quad (3.30)$$

where $\mathbf{L}_{1,Q}(n)$ and $\mathbf{L}_{2,Q}(n)$ are functionals from the conditions \mathbf{C}_1 , \mathbf{C}_2) and $\mathbf{L}_{1,n}^*$ and $\mathbf{L}_{2,n}^*$ are numerical sequences such that for any $\delta > 0$

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

\mathbf{H}_1) Assume that each distribution in the family \mathcal{Q}_n enters the class (3.30), i.e. $\mathcal{Q}_n \subseteq \mathcal{P}_n^*$ for each $n \geq 1$. Besides, the constant \varkappa_Q in (1.3) and the variance proxy σ_Q from the condition \mathbf{C}_1) are such that for $n = 1, 2, \dots$

$$\begin{cases} \varkappa^*(n) := \sup_{Q \in \mathcal{Q}_n} \varkappa_Q < \infty, & \sigma^*(n) := \sup_{Q \in \mathcal{Q}_n} \sigma_Q < \infty, \\ \sigma_*(n) := \inf_{Q \in \mathcal{Q}_n} \sigma_Q > 0 \end{cases} \quad (3.31)$$

and for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\varkappa^*(n)}{n^\delta} = 0, \quad \lim_{n \rightarrow \infty} \frac{\sigma^*(n)}{n^\delta} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^\delta \sigma_*(n) = +\infty. \quad (3.32)$$

Taking into account these notations, we get the following upper bounds for the functions $\Psi_Q(\rho)$ and $\mathbf{K}_Q(n)$ in (3.21) and (3.26)

$$\Psi_n^*(p) = 2 \left(2\mathbf{L}_{1,n}^* + \nu_p \mathbf{L}_{2,n}^* (\sigma_*(n))^{-1} + \nu_p \varkappa^*(n) \right)$$

and

$$\mathbf{K}_n^* = 16r + 8\sqrt{r\varkappa^*(n)} + \mathbf{L}_{1,n}^* + \sqrt{\mathbf{L}_{2,n}^* + \sigma^*(n)}.$$

Now proceeding from Theorem 3.4 we come to the following result.

Theorem 3.5. Assume that the conditions (1.3) and \mathbf{H}_1) hold and the function S satisfies the inequality (3.25) for some $r > 0$. Then, for any $n \geq 1$, $p \geq 3$, $0 < \rho < 1/3$ and any set Γ with the property (3.12) the estimate (3.20) satisfies the oracle inequality for the robust risk (1.6), i.e.

$$\mathcal{R}_n^*(\widehat{S}_*, S) \leq \frac{1+3\rho}{1-3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_n^*(\widehat{S}_\gamma, S) + \frac{\mathcal{B}_n^*(p)}{n(1-3\rho)\rho}, \quad (3.33)$$

where $\mathcal{B}_n^*(p) = \Psi_n^*(p) + 18\mathbf{K}_n^* \mu_p \mathbf{g}_{p,n}^*$.

3.3 Specification of weights in the model selection

Now we will specify the weight coefficients $(\gamma(j))_{1 \leq j \leq p}$ in the way proposed in [12]. Consider a numerical grid of the form

$$\mathcal{A} = \{1, \dots, k^*\} \times \{t_1, \dots, t_m\}, \quad (3.34)$$

where $t_i = i\varepsilon$ and $m = \lceil 1/\varepsilon^2 \rceil$. We assume that both parameters $k^* \geq 1$ and $n^{-2} \leq \varepsilon \leq 1$ are functions of n , i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} k^*(n) = +\infty, \quad \lim_{n \rightarrow \infty} \frac{k^*(n)}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \sigma^*(n)\varepsilon(n) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^\delta \varepsilon(n) = +\infty \end{array} \right. \quad (3.35)$$

for any $\delta > 0$. One can take, for example, for $n \geq 2$

$$\varepsilon(n) = \frac{1}{(\sigma^*(n) \ln n) \wedge n^2} \quad \text{and} \quad k^*(n) = k_0^* + \sqrt{\ln n}, \quad (3.36)$$

where $a \wedge b = \min(a, b)$ and $k_0^* \geq 0$ is some fixed constant. For each $\alpha = (\beta, t) \in \mathcal{A}$, we introduce the weight sequence

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

with the elements

$$\gamma_\alpha(j) = \mathbf{1}_{\{1 \leq j < j_*\}} + (1 - (j/\omega_\alpha)^\beta) \mathbf{1}_{\{j_* \leq j \leq \omega_\alpha\}} \quad (3.37)$$

where $j_* = \lceil \omega_\alpha / \ln(n+1) \rceil$, $\omega_\alpha = (\tau_\beta t n)^{1/(2\beta+1)}$ and $\tau_\beta = (\beta+1)(2\beta+1)/\pi^{2\beta}\beta$. Now we specify the set Γ in (3.19) as

$$\Gamma = \{\gamma_\alpha, \alpha \in \mathcal{A}\}. \quad (3.38)$$

Note that for any $p \geq 1$ the number of elements in Γ equals $\nu_p = k^*m$. Therefore, in view of the assumptions (3.35), for any $\delta > 0$, one has

$$\lim_{n \rightarrow \infty} \frac{k^*m}{n^\delta} = 0. \quad (3.39)$$

Taking into account that $\tau_\beta < 1$ for $\beta \geq 1$ and that $\varepsilon \geq n^{-2}$, we obtain

$$\omega^* = \sup_{\alpha \in \mathcal{A}} \omega_\alpha \leq (n/\varepsilon)^{1/3} \leq n. \quad (3.40)$$

Therefore, for any $p \geq 1$ and $n \geq 1$

$$\mu_p = \sup_{\alpha \in \mathcal{A}} \sum_{j=1}^p \mathbf{1}_{\{\gamma_\alpha(j) > 0\}} \leq \omega^* \leq n,$$

i.e. the condition (3.12) holds. Moreover, the last condition in (3.35) yields

$$\lim_{n \rightarrow \infty} \frac{\sup_{p \geq 1} \mu_p}{n^{1/3+\delta}} = 0 \quad \text{for any } \delta > 0.$$

Our main goal is to bound asymptotically (as $n \rightarrow \infty$) the term $\mathcal{B}_n^*(p)$ in (3.33) by any power of n . To this end we note that if

$$p \geq p_* = 2[n^{5/6}/2] + 3, \quad (3.41)$$

then the upper bound (3.28) can be estimated as $\mathbf{g}_{p,n}^* \leq 2n^{-1/3}$, i.e.

$$\lim_{n \rightarrow \infty} \frac{1}{n^\delta} \sup_{p \geq p_*} \mu_p \mathbf{g}_{p,n}^* = 0, \quad \forall \delta > 0. \quad (3.42)$$

Therefore, using these properties in (3.33) one comes to the following result.

Theorem 3.6. *Assume that the conditions (1.3) and \mathbf{H}_1) hold, the function S satisfies the inequality (3.25) for some $r > 0$. Then, for any $n \geq 3$, $p \geq 3$ and $0 < \rho < 1/3$, the estimate (3.20) with the weight coefficients (3.38) satisfies the oracle inequality (3.33), in which, for any $\delta > 0$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n^\delta} \sup_{p \geq p_*} \mathcal{B}_n^*(p) = 0. \quad (3.43)$$

Remark 3.5. *It will be noted that the assertion of Theorem 3.6 holds true, generally speaking, for any random process $(\xi_t)_{t \geq 0}$ in (1.1) for which the stochastic integral (1.2) is well defined and satisfies the conditions (1.3) and \mathbf{H}_1). The class of square integrable semimartingales, in our opinion, is the most appropriate one for the estimation problems with the quadratic risks (1.5) and (1.6). As an alternative, one could consider, for example, non-semimartingale regression models based on the fractional Brownian motions with the Hurst parameter $H \neq 1/2$. However, in the latter case, the stochastic integral is defined not for functions from $\mathcal{L}_2[0, n]$, but for some special spaces (see [7] for details). This does not agree with the the definition (1.5) and, moreover, the first limiting relation in (3.32) fails to hold.*

Now we consider the model (1.1)–(2.1) with the parameters satisfying the conditions (2.4) and derive the oracle inequality for the model selection procedure (3.20). To this end we set

$$\begin{cases} \mathbf{M}^* &= \left(116 + 33 \frac{\mathbf{E}Y_1^4}{\lambda_*}\right) \varrho_{max}^2; \\ \tau_* &= a_* (1 + e^{a_*} + 2(\sqrt{2}a_*e^{a_*} + 3)^2 e^{2a_*}). \end{cases} \quad (3.44)$$

Theorem 3.7. *Let the model be given by (1.1)–(2.1) with the parameters satisfying the conditions (2.4). Then, for any $n \geq 3$, $p \geq 3$ and $0 < \rho < 1/3$, the estimate (3.20) with the weight coefficients (3.38) satisfies the oracle inequality (3.33) in which $\varkappa^*(n) = 3\varrho_{max}$, $\sigma^*(n) = \varrho_{max}$, $\sigma_*(n) = \varrho_{min}$, $\mathbf{L}_{1,n}^* = 3\varrho_{max}\tau_*$, $\mathbf{L}_{2,n}^* = 96\mathbf{M}^*$ and the term $\mathcal{B}_n^*(p)$ has the property (3.43) and can be represented as*

$$\mathcal{B}_n^*(p) = \mathbf{c}_1 k^* m + \mathbf{c}_2 + \mathbf{c}_3 \mu_p \mathbf{g}_{p,n}^* \quad (3.45)$$

for some positive fixed constants \mathbf{c}_1 and \mathbf{c}_2 .

When considering the estimation problem for the model (1.1) with the martingale noise (2.5), we set

$$\mathbf{L}_{1,n}^* = \varkappa^*(n) + \varrho_*'/2 + \varrho_*'' \quad (3.46)$$

and

$$\mathbf{L}_{2,n}^* = 8\varkappa^*(n) + 11(\ln 2)^2 \left(\varrho_*'\right)^2 + 8 \frac{(\varkappa^*(n))^2 \mathbf{E}Y_1^4}{\lambda_*}, \quad (3.47)$$

where the parameters ϱ_*' , ϱ_*'' are defined in (2.7) and $\varkappa^*(n) = \sup_{0 \leq t \leq n} |\varrho_{max}(t)|$. Moreover, we set $\sigma_*(n) = n^{-1} \int_0^n \varrho_{min}(t) dt$ and

$$\sigma^*(n) = n^{-1} \int_0^n \varrho_{max}(t) dt. \quad (3.48)$$

Theorem 3.8. *Let the model be given by (1.1)–(2.5) with the parameters satisfying the conditions (2.7), (2.8) and (2.9). Then, for any $n \geq 3$, $p \geq 3$ and $0 < \rho < 1/3$, the estimate (3.20) with the weight coefficients (3.38) satisfies the oracle inequality (3.33) with the parameters (3.46) – (3.48), in which the term $\mathcal{B}_n^*(p)$ possesses the property (3.43) and has the form*

$$\mathcal{B}_n^*(p) = \mathbf{c}_1 \left(1 + \varkappa^*(n) + \frac{1 + (\varkappa^*(n))^2}{\sigma_*(n)} k^* m \right) + \mathbf{c}_2 (1 + \varkappa^*(n)) \mu_p \mathbf{g}_{p,n}^*$$

for some constants $\mathbf{c}_1 > 0$ and $\mathbf{c}_2 > 0$.

Remark 3.6. *It will be noted that the equation for $\mathcal{B}_n^*(p)$ in the Theorem 3.8 reduces to the form (3.45) provided that the functions $\varrho_{\min}(t)$ and $\varrho_{\max}(t)$ are constant.*

Now we write the oracle inequality for the “white noise” regression model, i.e. for the process (1.1), (2.5) with $\varrho_1 \equiv 1$ and $\varrho_2 \equiv 0$. In this case the distribution family \mathcal{Q}_n consists of a single distribution Q .

Theorem 3.9. *Let the model be given by (1.1)–(2.5) with $\varrho_1 \equiv 1$ and $\varrho_2 \equiv 0$. Then, for any $n \geq 3$, $p \geq 3$ and $0 < \rho < 1/3$, the estimate (3.20) with the weight coefficients (3.38) satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1 + 3\rho}{1 - 3\rho} \min_{\gamma \in \Gamma} \mathcal{R}_Q(\widehat{S}_\gamma, S) + \frac{1}{n} \frac{6k^*m}{(1 - 3\rho)\rho}.$$

The proofs of Theorems 3.7– 3.8 and Theorem 3.9 are given in Section 6.

4 Monte Carlo simulations

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (3.20). We define S in (1.1) as 1 - periodic function with $S(t) = t \sin(2\pi t) + t^2(1 - t) \cos(4\pi t)$ on $[0, 1]$ and then simulate the model as

$$dy_t = S(t)dt + d\xi_t, \tag{4.1}$$

where ξ_t is a non-Gaussian Ornstein–Uhlenbeck process obeying the equation

$$d\xi_t = -\xi_t dt + 0.5 dw_t + 0.5 dz_t.$$

Here z_t is a compound Poisson process (2.2) with the intensity $\lambda = 0.1$ and a Gaussian $\mathcal{N}(0, 1)$ sequence $(Y_j)_{j \geq 1}$.

The frequency of observations in (1.4) per period equals $p = 100001$. We use the weight sequence (3.37) in the procedure (3.20) as proposed in [12]

for a discrete time model: $k^* = 100 + \sqrt{\ln n}$, $\varepsilon = 1/\ln n$ and $m = \lceil 1/\varepsilon^2 \rceil$ and $\sigma^*(n) = 1$. We used the cost function (3.16) with $\rho = (3 + \ln n)^{-2}$, which satisfies the condition (5.4) below. We define the empirical risk as

$$\bar{\mathbf{R}} = \frac{1}{p} \sum_{j=1}^p \widehat{\mathbf{E}} \left(\tilde{S}_n(t_j) - S(t_j) \right)^2, \quad (4.2)$$

and the relative quadratic risk

$$\bar{\mathbf{R}}_* = \bar{\mathbf{R}}/\|S\|_p^2. \quad (4.3)$$

For the model (4.1) the empirical norm $\|S\|_p^2 = 0.18836$. The expectations in (4.2) was taken as an average over $N = 10000$ replications, i.e.

$$\widehat{\mathbf{E}} \left(\tilde{S}_n(\cdot) - S(\cdot) \right)^2 = \frac{1}{N} \sum_{t=1}^N \left(\tilde{S}_n^t(\cdot) - S(\cdot) \right)^2.$$

The following table gives the values for the sample risks (4.2) and (4.3) for different numbers of observation period n .

n	$\bar{\mathbf{R}}$	$\bar{\mathbf{R}}_*$
20	0.0465	0.2471(24.7%)
200	0.0078	0.0415(4.2%)
1000	0.0028	0.0148(1.5%)

Table : Empirical risks

The Figures 1–3 show the behavior of the regression function and its estimates by the model selection procedure (3.20) depending on the values of observation periods n .

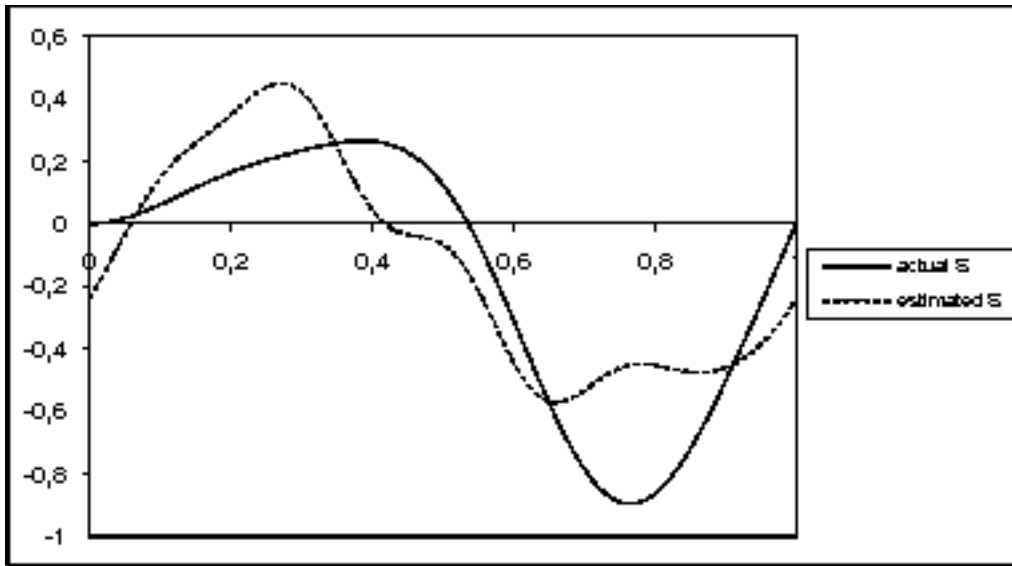


Figure 1: $n=20$

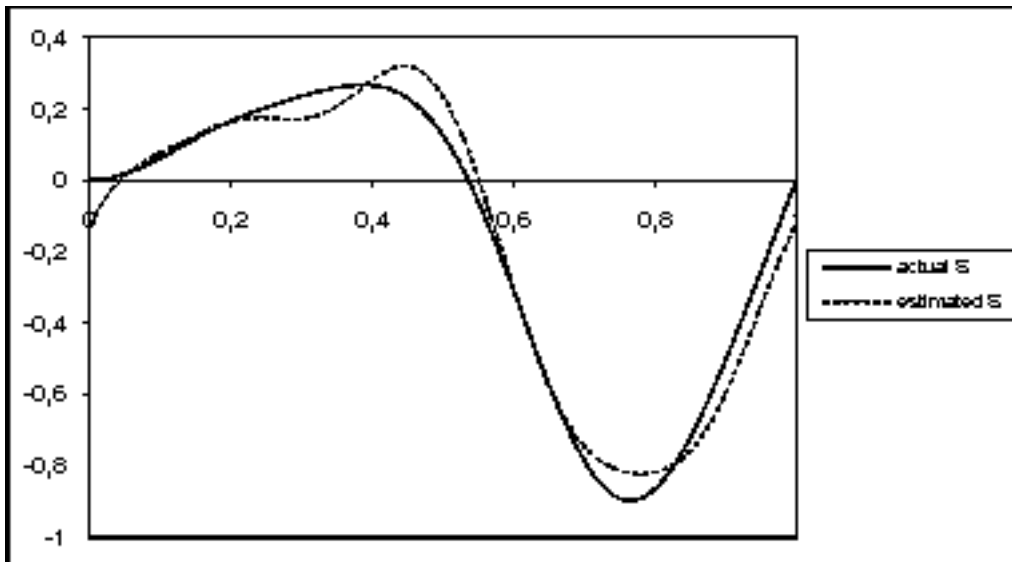


Figure 2: $n=200$

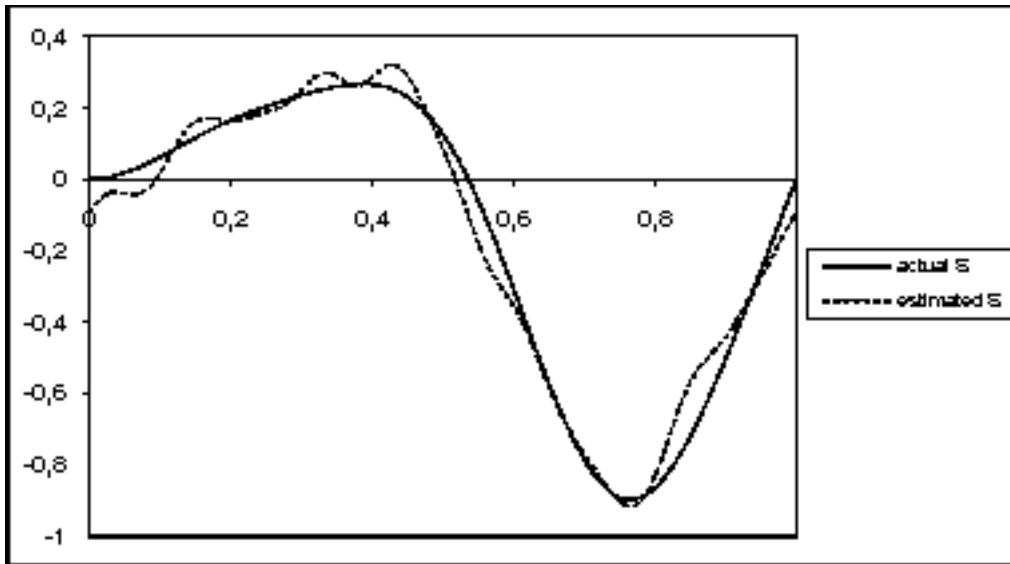


Figure 3: $n=1000$

Remark 4.1. *In the simulations we took $k_0^* = 100$ in (3.36). Note that the parameter $k^*(n)$ is the upper bound for the regularity order of the function S . So, it will be reasonable to choose maximum possible values for this upper bound for non large n , but not too large since we need to minimize the cost function in (3.19).*

5 Robust asymptotic efficiency

In this section we assume that the unknown function S in the model (1.1) belongs to the Sobolev ball W_r^k defined as

$$W_r^k = \left\{ f \in \mathcal{C}_{per}^k[0, 1] : \sum_{j=1}^{\infty} a_j \theta_j^2 \leq r \right\}, \quad \theta_j = \int_0^1 f(s) \phi_j(s) ds, \quad (5.1)$$

where $\mathcal{C}_{per}^k[0, 1]$ is the set of k times differentiable $[0, 1] \rightarrow \mathbb{R}$ functions f such that the $k - 1$ derivative $f^{(k-1)}$ is absolutely continuous and $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \leq i \leq k$; $a_j = \sum_{i=0}^k (\varpi_j)^{2i}$, ϖ_j is given in (3.1) and $(\theta_j)_{j \geq 1}$ are the Fourier coefficients of the function f . It is well known that the optimal

(minimax) estimation convergence rate for the functions from the set (5.1) is $n^{2k/(2k+1)}$ (see, for example, [28], [27]). We will see that asymptotically the robust risk (1.6) normalized by this rate is bounded from below by the Pinsker constant R_k^* which in this case has the following form

$$R_k^* = ((2k + 1)r)^{1/(2k+1)} \left(\frac{k}{(k + 1)\pi} \right)^{2k/(2k+1)}. \quad (5.2)$$

To this end we denote by Π_n the set of all estimators \widehat{S}_n measurable with respect to the sigma-algebra $\sigma\{y_t, 0 \leq t \leq n\}$ generated by the process (1.1). Moreover, we denote by Q_0 the distribution of Wiener process with the scale parameter $\sigma^*(n)$ defined in (3.31).

H₂) Assume that the distribution Q_0 belongs to the family \mathcal{Q}_n . Through this condition we obtain the following lower bound.

Theorem 5.1. Under the condition **H₂**)

$$\liminf_{n \rightarrow \infty} (d_n)^{2k/(2k+1)} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_n, S) \geq R_k^*, \quad (5.3)$$

where $d_n = n/\sigma^*(n)$.

Proof of this theorem follows directly from Theorem 3.2 in [22].

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^* .

H₃) Assume that there exists $\check{\delta} > 0$ such that

$$\lim_{n \rightarrow \infty} \frac{n^{5/6+\check{\delta}}}{p} = 0.$$

Furthermore, we assume that in the procedure (3.20) the parameter $\rho = \rho(n)$ is a function of n such that

$$\lim_{n \rightarrow \infty} \rho(n) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^\delta \rho(n) = \infty \quad (5.4)$$

for all $\delta > 0$.

Theorem 5.2. Assume that the conditions **H₁)** and **H₃)** hold. Then the robust risk (1.6) for the model selection procedure \widehat{S}_* , defined in (3.20) with the weight coefficients (3.38), admits the following asymptotic upper bound

$$\limsup_{n \rightarrow \infty} (d_n)^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_*, S) \leq R_k^*. \quad (5.5)$$

Theorem 5.1 and Theorem 5.2 imply the following result

Corollary 5.3. *Under the conditions \mathbf{H}_1), \mathbf{H}_2) and \mathbf{H}_3)*

$$\lim_{n \rightarrow \infty} (d_n)^{2k/(2k+1)} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_n, S) = R_k^*. \quad (5.6)$$

Remark 5.1. *It should be noted that the equality (5.6) means the robust efficiency holds with the convergence rate $(d_n)^{2k/(2k+1)}$. In view of the conditions (3.32) the sequence σ_n^* is slowly changing, i.e. asymptotically (as $n \rightarrow \infty$) $n^{-\delta} \leq \sigma_n^* \leq n^\delta$ for any $\delta > 0$. So, if the $\sigma_n^* = \text{const}$, then one obtains the usual minimax rate for the Sobolev ball (see, for example, [28], [27]). Furthermore, this rate becomes worse if the maximal value of the noise intensity $\sigma^*(n) \rightarrow \infty$ as $n \rightarrow \infty$, and it improves if $\sigma^*(n) \rightarrow 0$ as $n \rightarrow \infty$. For the regression model (1.1) with the martingale noise (2.5), the parameter $\sigma^*(n)$ is given in (3.48). If, for example, $\rho_{\max}(t) = \ln(t+2)$, then $\sigma^*(n) \rightarrow \infty$ as $n \rightarrow \infty$. In the case when $\rho_{\max}(t) = 1/\ln(t+2)$, one has $\sigma^*(n) \rightarrow 0$ as $n \rightarrow \infty$.*

To show Theorem 5.2 we need to estimate the function S from W_r^k in the nonadaptive setting, i.e. in the case when the parameters $k \geq 1$, $r > 0$ and $\sigma^*(n)$ are known. To this end we set

$$\gamma_0 = \gamma_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, t_0), \quad (5.7)$$

where $t_0 = \lceil \bar{r}/\varepsilon \rceil \varepsilon$, $\bar{r} = r/\sigma^*(n)$ and ε satisfies the conditions in (3.35). Note that for sufficiently large n the pair α_0 belongs to the set (3.34).

Theorem 5.4. *The estimator \widehat{S}_{γ_0} satisfies the following asymptotic upper bound*

$$\limsup_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \leq R_k^*. \quad (5.8)$$

Now taking into account the upper bound (5.8) in the oracle inequality (3.33) and using the property (3.43) we come to Theorem 5.2.

6 Proofs

6.1 Proof of Proposition 3.1

For each estimate (3.10) with $\gamma \in \Gamma$, combining (3.13) and (3.16) yields

$$\text{Err}(\gamma) = J(\gamma) + 2 \sum_{j=1}^p \gamma(j) \theta'_{j,p} + \|S\|^2 - \rho \widehat{P}(\gamma), \quad (6.1)$$

where $\theta'_{j,p} = \widetilde{\theta}_{j,p} - \bar{\theta}_{j,p} \widehat{\theta}_{j,p}$. From (3.8) and (3.14), one finds

$$\theta'_{j,p} = \frac{1}{\sqrt{n}} \bar{\theta}_{j,p} \xi_{j,p} + \frac{\sigma_Q - \widehat{\sigma}_n}{n} + \frac{\zeta_{j,p}}{n} + \frac{\widetilde{\xi}_{j,p}}{n},$$

where the sequences $\zeta_{j,p}$ and $\widetilde{\xi}_{j,p}$ are defined in the conditions \mathbf{C}_1) and \mathbf{C}_2) respectively. Furthermore, for any $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ we set

$$\begin{aligned} \mathbf{U}(x) &= \sum_{j=1}^p x_j, & \mathbf{V}(x) &= \frac{1}{\sqrt{n}} \sum_{j=1}^p x_j \bar{\theta}_{j,p} \xi_{j,p}, \\ \mathbf{B}_1(x) &= \sum_{j=1}^p x_j \zeta_{j,p} & \text{and} & \quad \mathbf{B}_2(x) = \sum_{j=1}^p x_j \widetilde{\xi}_{j,p}. \end{aligned} \quad (6.2)$$

Taking this into account, we rewrite (6.1) as

$$\begin{aligned} \text{Err}(\gamma) &= J(\gamma) + 2\mathbf{V}(\gamma) + \frac{2(\sigma_Q - \widehat{\sigma}_n)}{n} \mathbf{U}(\gamma) + \frac{2\mathbf{B}_1(\gamma)}{n} \\ &\quad + 2\sqrt{P_Q(\gamma)} \frac{\mathbf{B}_2(e(\gamma))}{\sqrt{n\sigma_Q}} + \|S\|^2 - \rho \widehat{P}(\gamma) \end{aligned} \quad (6.3)$$

where $e(\gamma) = \gamma/|\gamma|$. Now let γ_0 be a fixed weight vector in the set Γ and $\widehat{\gamma}$ be as in (3.19). Substituting γ_0 and $\widehat{\gamma}$ in (6.3), we arrive at the equation

$$\begin{aligned} \text{Err}(\widehat{\gamma}) - \text{Err}(\gamma_0) &= J(\widehat{\gamma}) - J(\gamma_0) + \frac{2(\sigma_Q - \widehat{\sigma}_n)}{n} \mathbf{U}(\widehat{x}) + \frac{2}{n} \mathbf{B}_1(\widehat{x}) \\ &\quad + 2\sqrt{P_Q(\widehat{\gamma})} \frac{\mathbf{B}_2(\widehat{e})}{\sqrt{n\sigma_Q}} - 2\sqrt{P_Q(\gamma_0)} \frac{\mathbf{B}_2(e_0)}{\sqrt{n\sigma_Q}} \\ &\quad + \rho \widehat{P}(\gamma_0) - \rho \widehat{P}(\widehat{\gamma}) + 2\mathbf{V}(\widehat{x}) \end{aligned} \quad (6.4)$$

where $\hat{x} = \hat{\gamma} - \gamma_0$, $\hat{e} = e(\hat{\gamma})$ and $e_0 = e(\gamma_0)$. From (3.11) it follows that for each $\gamma \in \Gamma$,

$$|\mathbf{U}(\hat{x})| \leq \sum_{j=1}^p (|\hat{\gamma}(j)| + |\gamma_0(j)|) \leq 2\mu_p.$$

Similarly, in view of the condition \mathbf{C}_1) we obtain

$$|\mathbf{B}_1(\hat{x})| \leq |\mathbf{B}_1(\hat{\gamma})| + |\mathbf{B}_1(\gamma_0)| \leq 2\mathbf{L}_{1,Q}(n).$$

Therefore, by the definition (3.19) one gets

$$\begin{aligned} \text{Err}(\hat{\gamma}) - \text{Err}(\gamma_0) &\leq \frac{4|\hat{\sigma}_n - \sigma_Q|}{n} \mu_p + \frac{4\mathbf{L}_{1,Q}(n)}{n} + 2\sqrt{P_Q(\hat{\gamma})} \frac{\mathbf{B}_2(\hat{e})}{\sqrt{\sigma_Q n}} - \rho\hat{P}(\hat{\gamma}) \\ &\quad + \rho\hat{P}(\gamma_0) - 2\sqrt{P_Q(\gamma_0)} \frac{\mathbf{B}_2(e_0)}{\sqrt{n\sigma_Q}} + 2\mathbf{V}(\hat{x}). \end{aligned} \quad (6.5)$$

Further, by applying the elementary inequality

$$2|ab| \leq \varepsilon a^2 + \varepsilon^{-1}b^2 \quad (6.6)$$

with $\varepsilon = \rho$, one has

$$2\sqrt{P_Q(\hat{\gamma})} \frac{|\mathbf{B}_2(e(\hat{\gamma}))|}{\sqrt{n\sigma_Q}} \leq \rho P_Q(\hat{\gamma}) + \frac{\mathbf{B}_2^2(e(\hat{\gamma}))}{n\sigma_Q\rho}.$$

Since $0 < \rho < 1$, this inequality and (6.5) imply that

$$\begin{aligned} \text{Err}(\hat{\gamma}) &\leq \text{Err}(\gamma_0) + \frac{4\mathbf{L}_{1,Q}(n)}{n} + \frac{2\mathbf{B}_2^*}{n\sigma_Q\rho} + \frac{1}{n}|\hat{\sigma}_n - \sigma_Q| (|\hat{\gamma}|^2 + |\gamma_0|^2 + 4\mu_p) \\ &\quad + 2\rho P_Q(\gamma_0) + 2\mathbf{V}(\hat{x}), \end{aligned}$$

where $\mathbf{B}_2^* = \sup_{\gamma \in \Gamma} \mathbf{B}_2^2(e(\gamma))$. Moreover, taking into account here that $\sup_{\gamma \in \Gamma} |\gamma|^2 \leq \mu_p$, we get

$$\begin{aligned} \text{Err}(\hat{\gamma}) &\leq \text{Err}(\gamma_0) + \frac{4\mathbf{L}_{1,Q}(n)}{n} + \frac{2\mathbf{B}_2^*}{n\sigma_Q\rho} + \frac{6\mu_p}{n} |\hat{\sigma}_n - \sigma_Q| \\ &\quad + 2\rho P_Q(\gamma_0) + 2\mathbf{V}(\hat{x}). \end{aligned} \quad (6.7)$$

By the condition \mathbf{C}_2), one has

$$\mathbf{E}_Q \mathbf{B}_2^* \leq \sum_{\gamma \in \Gamma} \mathbf{E}_Q \mathbf{B}_2^2(e(\gamma)) \leq \nu_p \mathbf{L}_{2,Q}(n), \quad (6.8)$$

where $\nu_p = \text{card}(\Gamma)$ is defined in (3.11). Now we examine the last term in the right side of (6.7). To this end for any $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ we set

$$S_x(t) = \sum_{j=1}^p x_j \bar{\theta}_{j,p} \phi_j(t) \quad \text{and} \quad \widehat{S}_x(t) = \sum_{j=1}^p x_j \widehat{\theta}_{j,p} \phi_j(t). \quad (6.9)$$

Using (1.3), one obtains that for any non-random vector x from \mathbb{R}^p

$$\mathbf{E}_Q \mathbf{V}^2(x) \leq \varkappa_Q \frac{1}{n} \sum_{j=1}^p x_j^2 \bar{\theta}_{j,p}^2 = \frac{\varkappa_Q}{n} \|S_x\|^2. \quad (6.10)$$

Let denote $Z^* = \max_{x \in \Gamma_1} n \mathbf{V}^2(x) / \|S_x\|^2$ and $\Gamma_1 = (\Gamma - \gamma_0) \setminus \{0\}$. The average of this quantity can be estimated as

$$\mathbf{E}_Q Z^* \leq \sum_{x \in \Gamma_1} \frac{n \mathbf{E}_Q \mathbf{V}^2(x)}{\|S_x\|^2} \leq \sum_{x \in \Gamma_1} \varkappa_Q \leq \varkappa_Q \nu_p. \quad (6.11)$$

By applying (6.6) one gets

$$\max_{x \in \Gamma_1} (2|\mathbf{V}(x)| - \rho \|S_x\|^2) \leq \frac{Z^*}{n\rho}. \quad (6.12)$$

Now we need to find an upper bound for $\|S_x\|^2$. To this end note that for each $x \in \Gamma_1$, one has the following estimate

$$\|S_x\|^2 - \|\widehat{S}_x\|^2 = \sum_{j=1}^p x_j^2 (\bar{\theta}_{j,p}^2 - \widehat{\theta}_{j,p}^2) \leq -2\mathbf{V}(x^2),$$

where $x^2 = (x_1^2, \dots, x_p^2)$. Since $|x_j| \leq 1$, one gets, for each $x \in \Gamma_1$,

$$\mathbf{E}_Q \mathbf{V}^2(x^2) \leq \varkappa_Q \frac{\sum_{j=1}^p x_j^4 \bar{\theta}_{j,p}^2}{n} \leq \varkappa_Q \frac{\sum_{j=1}^p x_j^2 \bar{\theta}_{j,p}^2}{n} = \varkappa_Q \frac{\|S_x\|^2}{n}.$$

Denoting $Z_1^* = \max_{x \in \Gamma_1} n \mathbf{V}^2(x^2) / \|S_x\|^2$, one has

$$\mathbf{E}_Q Z_1^* \leq \varkappa_Q \nu_p. \quad (6.13)$$

By the same argument as in (6.12), one shows that for any $x \in \Gamma_1$

$$2|\mathbf{V}(x^2)| \leq \rho \|S_x\|^2 + \frac{Z_1^*}{n\rho}.$$

From here provided that $0 < \rho < 1$, it follows that

$$\|S_x\|^2 \leq \frac{\|\widehat{S}_x\|^2}{1-\rho} + \frac{Z_1^*}{n\rho(1-\rho)}. \quad (6.14)$$

Substituting this bound in (6.12) yields

$$\max_{x \in \Gamma_1} \left(2|\mathbf{V}(x)| - \frac{\rho \|\widehat{S}_x\|^2}{1-\rho} \right) \leq \frac{Z^*(1-\rho) + \rho Z_1^*}{n\rho(1-\rho)} \leq \frac{Z^* + Z_1^*}{n\rho(1-\rho)}.$$

Since the range of the random vector \widehat{x} is included in Γ_1 , the last inequality implies

$$2|\mathbf{V}(\widehat{x})| \leq \frac{\rho \|\widehat{S}_{\widehat{x}}\|^2}{1-\rho} + \frac{Z^* + Z_1^*}{n\rho(1-\rho)}.$$

Note that the norm $\|\widehat{S}_{\widehat{x}}\|^2$ can be estimated as

$$\|\widehat{S}_{\widehat{x}}\|^2 = \|(\widehat{S}_{\widehat{\gamma}} - S) - (S - \widehat{S}_{\gamma_0})\|^2 \leq 2(\text{Err}(\widehat{\gamma}) + \text{Err}(\gamma_0)),$$

i.e.

$$2\mathbf{V}(\widehat{x}) \leq \frac{2\rho(\text{Err}(\widehat{\gamma}) + \text{Err}(\gamma_0))}{1-\rho} + \frac{Z^* + Z_1^*}{n\rho(1-\rho)}.$$

From here and (6.7), it follows that for $0 < \rho < 1/3$

$$\begin{aligned} \text{Err}(\widehat{\gamma}) &\leq \frac{1+\rho}{1-3\rho} \text{Err}(\gamma_0) + \frac{4\mathbf{L}_{1,Q}(n)}{(1-3\rho)n} + \frac{2B_2^*}{\rho\sigma_Q(1-3\rho)n} \\ &\quad + \frac{6\mu_p|\widehat{\sigma}_n - \sigma_Q|}{(1-3\rho)n} + \frac{2\rho}{1-3\rho} P_Q(\gamma_0) + \frac{Z^* + Z_1^*}{\rho(1-3\rho)n}. \end{aligned}$$

Taking the expectation and making use of (6.8), (6.11) and (6.13) we get

$$\begin{aligned} \mathbf{E}_Q \|\widehat{S}_* - S\|^2 &\leq \frac{1+\rho}{1-3\rho} \mathbf{E}_Q \text{Err}(\gamma_0) + \frac{4\mathbf{L}_{1,Q}(n)}{(1-3\rho)n} + \frac{2\nu_p \mathbf{L}_{2,Q}(n)}{\rho\sigma_Q(1-3\rho)n} \\ &\quad + \frac{6\mu_p \mathbf{r}_Q(\widehat{\sigma}_n)}{(1-3\rho)n} + \frac{2\rho}{1-3\rho} P_Q(\gamma_0) + \frac{2\nu_p \varkappa_Q}{\rho(1-3\rho)n}, \end{aligned}$$

where $\mathbf{r}_Q(\widehat{\sigma}_n)$ is defined in (3.15). Lemma A.1 yields

$$\begin{aligned} \mathbf{E}_Q \|\widehat{S}_* - S\|^2 &\leq \frac{1+3\rho}{1-3\rho} \mathbf{E}_Q \text{Err}(\gamma_0) + \frac{2(2+\rho)\mathbf{L}_{1,Q}(n)}{(1-3\rho)n} + \frac{2\nu_p \mathbf{L}_{2,Q}(n)}{\rho\sigma_Q(1-3\rho)n} \\ &\quad + \frac{6\mu_p \mathbf{r}_Q(\widehat{\sigma}_n)}{(1-3\rho)n} + \frac{2\nu_p \varkappa_Q}{\rho(1-3\rho)n}. \end{aligned}$$

From here, in view of the definition of the function $\Psi_Q(n, p)$ in (3.21), one comes to the inequality (3.22). Hence Proposition 3.1. \square

6.2 Proof of Lemma 3.3

It is clear that the inequality (3.26) holds for $l_0 > \check{p}$. Let now $l_0 \leq \check{p}$. Setting $x'_j = \mathbf{1}_{\{l_0 \leq j \leq \check{p}\}}$ and substituting (3.8) in (3.24) yields

$$\widehat{\sigma}_n = \frac{n}{\check{p}} \sum_{j=l_0}^{\check{p}} (\bar{\theta}_{j,p})^2 + \frac{2n}{\check{p}} \mathbf{V}(x') + \frac{1}{\check{p}} \sum_{j=l_0}^{\check{p}} \xi_{j,p}^2. \quad (6.15)$$

Furthermore, putting $x''_j = \check{p}^{-1/2} \mathbf{1}_{\{l_0 \leq j \leq \check{p}\}}$, one can write the last term on the right of (6.15) as

$$\frac{1}{\check{p}} \sum_{j=l_0}^{\check{p}} \xi_{j,p}^2 = \frac{1}{\sqrt{\check{p}}} \mathbf{B}_2(x'') + \frac{1}{\check{p}} \mathbf{B}_1(x') + \frac{\check{p} - l_0 + 1}{\check{p}} \sigma_Q,$$

where the functions $\mathbf{V}(\cdot)$, $\mathbf{B}_1(\cdot)$ and $\mathbf{B}_2(\cdot)$ are defined in (6.2). Note now that through the Cauchy inequality the condition (3.25) implies $|S_1^2| \leq r$ and

$|\dot{S}|_1^2 \leq r$. Hence, Lemma A.3 yields

$$\begin{aligned} \mathbf{E}_Q |\hat{\sigma}_n - \sigma_Q| &\leq \frac{8rn}{\check{p}} \sum_{j=l_0}^{\infty} j^{-2} + \frac{2n}{\check{p}} \mathbf{E}_Q |\mathbf{V}(x')| + \frac{\sqrt{\mathbf{E}_Q |\mathbf{B}_2(x'')|^2}}{\sqrt{\check{p}}} \\ &\quad + \frac{|\mathbf{B}_1(x')|}{\check{p}} + \frac{l_0 - 1}{\check{p}} \sigma_Q. \end{aligned}$$

Taking into account that $\sum_{j \geq l_0} j^{-2} \leq 2/l_0$ and the conditions $\mathbf{C}_1) - \mathbf{C}_2)$, one comes to the following upper bound

$$\begin{aligned} \mathbf{E}_Q |\hat{\sigma}_n - \sigma_Q| &\leq \frac{16rn}{l_0 \check{p}} + \frac{2n}{\check{p}} \mathbf{E}_Q |\mathbf{V}(x')| + \frac{\sqrt{\mathbf{L}_{2,Q}(n)}}{\sqrt{\check{p}}} \\ &\quad + \frac{\mathbf{L}_{1,Q}(n) + (l_0 - 1)\sigma_Q}{\check{p}}. \end{aligned} \tag{6.16}$$

Moreover, in view of the inequality (6.10), one gets

$$\mathbf{E}_Q |\mathbf{V}(x')| \leq \left(\frac{\varkappa_Q}{n} \sum_{j=l_0}^{\check{p}} \bar{\theta}_{j,p}^2 \right)^{1/2} \leq \frac{4(r\varkappa_Q)^{1/2}}{\sqrt{n l_0}} \leq \frac{4(r\varkappa_Q)^{1/2}}{l_0}.$$

This implies immediately Lemma 3.3 . \square

6.3 Proof of Theorem 3.7

One has to check the conditions (1.3), $\mathbf{C}_1)$ and $\mathbf{C}_2)$ for the model(1.1)-(2.1) and apply Theorem 3.6. The property (1.3) for this model is checked in [24] with $\varkappa_Q = 3\varrho^*$ and $\varrho^* = \varrho_1^2 + \lambda\varrho_2^2$. Therefore, from the definition of the distribution family in (2.4) we obtain $\varkappa_Q = 3\varrho_{max}$. A straightforward but tedious calculation shows that

$$\mathbf{E}_Q \xi_{j,p}^2 = \varrho^* (1 + \tau_j), \tag{6.17}$$

where $\tau_j = n^{-1} \int_0^n \psi_{j,p}(t) \varepsilon_j(t) dt$ and $\varepsilon_j(t) = a \int_0^t e^{a(t-s)} \psi_{j,p}(s) (1 + e^{2as}) ds$. It is easy to check that $|\tau_1| \leq 2$. For $j \geq 2$ we represent τ_j and $\varepsilon_j(t)$ as

$$\tau_j = \tau_{1,j} + \tau_{2,j} \quad \text{and} \quad \varepsilon_j(t) = \varepsilon_{1,j}(t) + \varepsilon_{2,j}(t),$$

where $\varepsilon_{1,j}(t) = a \int_0^t e^{a(t-s)} \psi_{j,p}(s) ds$, $\varepsilon_{2,j}(t) = a \int_0^t e^{a(t+s)} \psi_{j,p}(s) ds$

$$\tau_{1,j} = \frac{1}{n} \int_0^n \psi_{j,p}(t) \varepsilon_{1,j}(t) dt \quad \text{and} \quad \tau_{2,j} = \frac{1}{n} \int_0^n \psi_{j,p}(t) \varepsilon_{2,j}(t) dt. \quad (6.18)$$

In view of (3.6) $\tau_{1,j}$ can be represented as

$$\tau_{1,j} = \frac{a}{n} \sum_{k=1}^{np} \sum_{l=1}^k \phi_j(t_k) \phi_j(t_l) \varkappa_{k,l},$$

where

$$\varkappa_{k,l} = \int_{t_{l-1}}^{t_l} \left(\int_{t_{k-1}}^{t_k} (e^{a(t-s)} \mathbf{1}_{\{s \leq t\}}) dt \right) ds.$$

Integration yields $\varkappa_{k,k} = \varkappa_0^*(\Delta) = (e^{a\Delta} - 1 - a\Delta)a^{-2}$ and for $1 \leq l < k$ the coefficients $\varkappa_{k,l} = \varkappa_1^*(\Delta) e^{at_{k-l}}$ with $\varkappa_1^*(\Delta) = a^{-2}(e^{-a\Delta} - 1)(1 - e^{a\Delta})$. Therefore,

$$\tau_{1,j} = a\varkappa_0^*(\Delta)p + a\varkappa_1^*(\Delta)p^2 A_{j,p}, \quad (6.19)$$

where

$$A_{j,p} = \frac{1}{np^2} \sum_{l=1}^{np-1} v_{j,l}, \quad v_{j,l} = \sum_{k=l+1}^{np} \phi_j(t_k) \phi_j(t_l) e^{at_{k-l}}.$$

It should be noted that

$$0 \leq \varkappa_0^*(\Delta) \leq \frac{\Delta^2}{2} \quad \text{and} \quad 0 \leq \varkappa_1^*(\Delta) \leq \Delta^2 e^{a^*}. \quad (6.20)$$

From the definition of the functions $(\phi_j)_{2 \leq j \leq p}$ it follows that

$$v_{j,l} = \sum_{k=1}^{np-l} \cos(\varpi_j t_k) e^{at_k} + (-1)^j \sum_{k=1}^{np-l} \cos(\varpi_j t_{k+2l}) e^{at_k} := \tilde{v}_{j,l} + (-1)^j \hat{v}_{j,l}.$$

So, setting

$$\tilde{A}_{j,p} = \frac{1}{np^2} \sum_{l=1}^{np-1} \tilde{v}_{j,l} \quad \text{and} \quad \hat{A}_{j,p} = \frac{1}{np^2} \sum_{l=1}^{np-1} \hat{v}_{j,l},$$

we can represent $A_{j,p}$ in the form $A_{j,p} = \tilde{A}_{j,p} + (-1)^j \hat{A}_{j,p}$, in which

$$\tilde{A}_{j,p} = \frac{1}{p} \mathcal{R}e \frac{\tilde{\mathbf{q}}_j}{1 - \tilde{\mathbf{q}}_j} - \tilde{B}_{j,p} \quad \text{and} \quad \hat{A}_{j,p} = \frac{1 - e^{an}}{np^2} \hat{B}_j. \quad (6.21)$$

Here

$$\tilde{B}_{j,p} = \frac{1 - e^{an}}{np^2} \mathcal{R}e \frac{\tilde{\mathbf{q}}_j}{(1 - \tilde{\mathbf{q}}_j)^2} \quad \text{and} \quad \hat{B}_j = \mathcal{R}e \frac{\tilde{\mathbf{q}}_j \hat{\mathbf{q}}_j}{(1 - \tilde{\mathbf{q}}_j)(1 - \hat{\mathbf{q}}_j)},$$

where $\tilde{\mathbf{q}}_j = e^{a\Delta + i\varpi_j \Delta}$ and $\hat{\mathbf{q}}_j = e^{-a\Delta + i\varpi_j \Delta}$. First, note that

$$\mathcal{R}e \frac{1}{1 - \tilde{\mathbf{q}}_j} = \frac{1 - e^{a\Delta} \cos(\varpi_j \Delta)}{1 - 2e^{a\Delta} \cos(\varpi_j \Delta) + e^{2a\Delta}} = \frac{1 - e^{a\Delta} + 2e^{a\Delta} \sin^2(\varpi_j \Delta/2)}{(1 - e^{a\Delta})^2 + 4e^{a\Delta} \sin^2(\varpi_j \Delta/2)}.$$

Taking into account that

$$\sin(x) \geq 2x/\pi, \quad 0 \leq x \leq \pi/2,$$

and

$$\varpi_j \geq 2\pi j/3, \quad 2 \leq j \leq p,$$

we obtain for $2 \leq j \leq p$

$$2 \sin(\varpi_j \Delta/2) \geq j\Delta. \quad (6.22)$$

Therefore,

$$\frac{1}{p} \left| \mathcal{R}e \frac{1}{1 - \tilde{\mathbf{q}}_j} \right| \leq \frac{|a|\Delta}{4pe^{a\Delta} \sin^2(\varpi_j \Delta/2)} + \frac{1}{2p} \leq \frac{a_* e^{a_*}}{j^2} + \frac{1}{2p}, \quad (6.23)$$

where a_* is given in (2.4). Now we estimate $\tilde{B}_{j,p}$. We have

$$\begin{aligned} \tilde{B}_{j,p} &= \frac{(1 - e^{an})e^{a\Delta} (\cos(\varpi_j \Delta) - 2e^{a\Delta} + e^{2a\Delta} \cos(\varpi_j \Delta))}{np^2 (1 - 2e^{a\Delta} \cos(\varpi_j \Delta) + e^{2a\Delta})^2} \\ &= \frac{(1 - e^{an})e^{a\Delta} \left((1 - e^{a\Delta})^2 - 2(1 + e^{2a\Delta}) \sin^2(\varpi_j \Delta/2) \right)}{np^2 \left((1 - e^{a\Delta})^2 + 4e^{a\Delta} \sin^2(\varpi_j \Delta/2) \right)^2}. \end{aligned}$$

Applying the inequality (6.22) yields

$$\begin{aligned} |\tilde{B}_{j,p}| &\leq \frac{|a|^2 \Delta^2}{16p^2 e^{a\Delta} \sin^4(\varpi_j \Delta/2)} + \frac{1}{4p^2 e^{a\Delta} \sin^2(\varpi_j \Delta/2)} \\ &\leq e^{a_*} (a_*^2 j^{-4} + j^{-2}). \end{aligned} \quad (6.24)$$

By making use of the estimates (6.23) and (6.24) in (6.21) we obtain

$$\sup_{n \geq 1} |\tilde{A}_{j,p}| \leq \frac{1}{2p} + \frac{(a_* + 1)e^{a_*}}{j^2} + \frac{a_*^2 e^{a_*}}{j^4}. \quad (6.25)$$

To estimate $\hat{A}_{j,p}$ we represent \hat{B}_j as

$$\hat{B}_j = \frac{\cos(\varpi_j \Delta)}{2(\cos(\varpi_j \Delta) - \text{ch}(a\Delta))},$$

where $\text{ch}(x) = (e^x + e^{-x})/2$. From here and (6.22), it follows that

$$\begin{aligned} |\hat{A}_{j,p}| &\leq \frac{1}{2np^2} \left(\frac{1}{\text{ch}(a\Delta) - \cos(\varpi_j \Delta)} \right) \leq \frac{1}{2np^2} \left(\frac{1}{1 - \cos(\varpi_j \Delta)} \right) \\ &= \frac{1}{4np^2 \sin^2(\varpi_j \Delta/2)} \leq \frac{1}{nj^2}. \end{aligned} \quad (6.26)$$

Combining (6.25) and (6.26) yields

$$p^{-1} \sup_{n \geq 1} |A_{j,p}| \leq \mathbf{A}_* (p^{-1} + j^{-2}), \quad \mathbf{A}_* = 1 + (1 + a_* + a_*^2)e^{a_*}.$$

This and (6.19), in view of (6.20), implies that

$$j^{-2} \sup_{n \geq 1} |\tau_{1,j}| \leq \tau_1^* (p^{-1} + j^{-2}), \quad \tau_1^* = a_* (1 + e^{a_*} \mathbf{A}_*). \quad (6.27)$$

It remains to estimate $\tau_{2,j}$ in (6.18). First we note that

$$\tau_{2,j} = \frac{a}{2n} \iota_{j,p}^2 \quad \text{and} \quad \iota_{j,p} = \int_0^n \psi_{j,p}(t) e^{at} dt.$$

It is easy to check that for $2 \leq j \leq p$

$$\iota_{j,p} = \frac{\sqrt{2}(e^{-a\Delta} - 1)}{|a|} (1 - e^{an}) \left(\Upsilon_j \left(\frac{1}{1 - \tilde{\mathbf{q}}_j} \right) - 1 \right), \quad (6.28)$$

where $\Upsilon_j(z) = \Re e(z)$ for even j and $\Upsilon_j(z) = \Im m(z)$ for odd j . For even j , in view of (6.23), one gets the inequality

$$|\iota_{j,p}| \leq \iota_* (p^{-1} + j^{-2}), \quad \iota_* = \sqrt{2} a_* e^{2a_*} + 3 e^{a_*}. \quad (6.29)$$

For odd $3 \leq j \leq p$ one has the estimate

$$\left| \Im m \frac{1}{1 - \tilde{\mathbf{q}}_j} \right| \leq \frac{1}{j\Delta},$$

which implies that

$$|\iota_{j,p}| \leq \sqrt{2} e^{a_*} (p^{-1} + j^{-1}) \leq \iota_* (p^{-1} + j^{-1}). \quad (6.30)$$

Therefore,

$$\sup_{n \geq 1} |\tau_{2,j}| \leq a_* \iota_*^2 (p^{-2} + j^{-2}), \quad 2 \leq j \leq p.$$

From here and the definition of τ_j in (6.17), we obtain

$$\sup_{n \geq 1} \sum_{j=1}^p |\tau_j| \leq \tau_* \left(1 + \sum_{j \geq 1} j^{-2} \right) \leq 3\tau_*,$$

where τ_* is defined in (3.44). This, in view of (6.17), shows that the condition \mathbf{C}_1) holds with $\sigma_Q \leq \varrho_{max}$. Since, for any Q from the noise distribution family \mathcal{Q}_n defined in (2.4), $\sigma_Q \leq \varrho_{max}$, we obtain the upper bound for the functional $\mathbf{L}_{1,Q}(n)$, i.e. $\mathbf{L}_{1,Q}(n) \leq 3\varrho^* \tau_* \leq 3\varrho_{max} \tau_* := \mathbf{L}_{1,n}^*$. Now we check the condition \mathbf{C}_2) and find the upper bound $\mathbf{L}_{2,n}^*$ for $\mathbf{L}_{2,Q}(n)$. We have

$$\mathbf{E}_Q \left(\sum_{j=1}^p x_j \tilde{\xi}_{j,p} \right)^2 \leq 2\mathbf{E}_Q \tilde{\xi}_{1,p}^2 + 2\mathbf{E}_Q \left(\sum_{j=2}^p x_j \tilde{\xi}_{j,p} \right)^2$$

and

$$\mathbf{E}_Q \left(\sum_{j=2}^p x_j \tilde{\xi}_{j,p} \right)^2 = \frac{1}{n^2} \sum_{j=2}^p \sum_{l=2}^p x_j x_l \mathbf{E}_Q \tilde{I}_n(\psi_{j,p}) \tilde{I}_n(\psi_{l,p}), \quad (6.31)$$

where $\tilde{I}_n(f) = I_n^2(f) - \mathbf{E}_Q I_n^2(f)$. To study this term we will use the correlation measure introduced in [24]. The correlation measure for two integrated $[0, +\infty) \rightarrow \mathbb{R}$ functions f and g is defined as

$$\Upsilon_{f,g} = \max_{0 \leq v \leq n} \max_{0 \leq t \leq n-v} \left| \int_0^t f(u+v)g(u)du \right|$$

and $\Upsilon_{f,g}^* = \max(\Upsilon_{f,g}, \Upsilon_{g,f})$. Denoting $v_{j,l} = \Upsilon_{\psi_{j,p}, \psi_{l,p}}$ and $v_{j,l}^* = \Upsilon_{\psi_{j,p}, \psi_{l,p}}^*$, and applying Theorem 4.4 from [24], we obtain for any Q from the distribution family \mathcal{Q}_n the following inequality

$$|\mathbf{E}_Q \tilde{I}_n(\psi_{j,p}) \tilde{I}_n(\psi_{l,p})| \leq 2n\mathbf{M}_Q \left(1 + v_{j,l}^* + v_{j,1}^* + v_{l,1}^* \right),$$

where $\mathbf{M}_Q = 116(\varrho^*)^2 + 33\lambda\varrho_2^4\mathbf{E}Y_1^4$. It is clear, that $v_{1,1}^* = n$, i.e.

$$\mathbf{E}_Q \tilde{\xi}_{1,p}^2 = \frac{1}{n^2} \mathbf{E}_Q \left(\tilde{I}_n(\psi_{1,p}) \right)^2 \leq 8\mathbf{M}_Q.$$

Furthermore, note that for any $t, v \geq 0$ and $j \geq 2$

$$\left| \int_0^t \psi_{j,p}(u+v) du \right| = \left| \int_{v+[t]}^{v+t} \psi_{j,p}(u) du \right| \leq \int_0^1 |\psi_{j,p}(u)| du \leq \|\psi_{j,p}\| = 1,$$

i.e. $\sup_{j \geq 2} v_{j,1}^* \leq 1$. Therefore, for $j, l \geq 2$

$$|\mathbf{E}_Q \tilde{I}_n(\psi_{j,p}) \tilde{I}_n(\psi_{l,p})| \leq 2n\mathbf{M}_Q \left(3 + v_{j,l}^* \right). \quad (6.32)$$

To estimate the term $v_{j,l}$ note that for any $1 \leq j, l \leq p$

$$v_{j,l} = \max_{0 \leq v \leq 1} \max_{0 \leq t \leq n-v} \left| [t] \mathbf{V}_{j,l}(v) + \int_{[t]}^t \psi_{j,p}(u+v) \psi_{l,p}(u) du \right|,$$

where $\mathbf{V}_{j,l}(v) = \int_0^1 \psi_{j,p}(u+v) \psi_{l,p}(u) du$. Therefore,

$$v_{j,l} \leq 2 + n \max_{0 \leq v \leq 1} |\mathbf{V}_{j,l}(v)|. \quad (6.33)$$

Taking into account the definition (3.5), we find that for $t_{s-1} \leq v \leq t_s$ the function $\mathbf{V}_{j,l}(v)$ can be represented as

$$\begin{aligned} \mathbf{V}_{j,l}(v) &= (t_s - v) \sum_{k=1}^p \phi_l(t_k) \phi_j(t_k + t_{s-1}) \\ &\quad + (v - t_{s-1}) \sum_{k=1}^p \phi_l(t_k) \phi_j(t_k + t_s). \end{aligned} \quad (6.34)$$

It is straightforward to verify that for any $u, v \geq 0$ the trigonometric functions ϕ_j satisfy the following equation

$$\phi_j(u + v) = \mathbf{k}_{j,1}^*(v) \phi_{j-1}(u) + \mathbf{k}_{j,2}^*(v) \phi_j(u) + \mathbf{k}_{j,3}^*(v) \phi_{j+1}(u),$$

with the bounded coefficients, because $\max_{j \geq 1} \max_{1 \leq i \leq 3} \sup_{v \geq 0} |\mathbf{k}_{j,i}^*(v)| \leq 1$. From here and (6.34), it follows that, $\mathbf{V}_{j,l}(v) = 0$ for $|l - j| \geq 2$. So, taking into account in (6.33) that $\sup_{j,l \geq 1} |\mathbf{V}_{j,l}(v)| \leq 2$, we can represent the upper bound (6.33) as $v_{j,l} \leq 2 + 2n \mathbf{1}_{\{|l-j| \leq 1\}}$ and, therefore, $v_{j,l}^* \leq 2 + 2n \mathbf{1}_{\{|l-j| \leq 1\}}$. Substituting this estimate in (6.32) yields

$$|\mathbf{E}_Q \tilde{I}_n(\psi_{j,p}) \tilde{I}_n(\psi_{l,p})| \leq 2n \mathbf{M}_Q \left(5 + 2n \mathbf{1}_{\{|l-j| \leq 1\}} \right).$$

Thus, taking into account that $\#(x) \leq n$ and $|x| \leq 1$, we estimate the right hand side of (6.31) as

$$\begin{aligned} \mathbf{E}_Q \left(\sum_{j=2}^p x_j \tilde{\xi}_{j,p} \right)^2 &\leq \frac{10 \mathbf{M}_Q}{n} \sum_{j=2}^p \sum_{l=2}^p x_j x_l \left(1 + n \mathbf{1}_{\{|l-j| \leq 1\}} \right) \\ &\leq 10 \mathbf{M}_Q \left(\frac{1}{n} \left(\sum_{j=2}^p x_j \right)^2 + \sum_{j=2}^p x_j x_{j-1} + \sum_{j=2}^p x_j^2 + \sum_{j=2}^{p-1} x_j x_{j+1} \right) \\ &\leq 40 \mathbf{M}_Q, \end{aligned}$$

which shows that $\mathbf{L}_{2,Q}(n) \leq 96 \mathbf{M}_Q$ and proves the validity of the condition \mathbf{C}_2). For each noise distribution in the family \mathcal{Q}_n , defined in (2.4), we obtain that $\mathbf{M}_Q \leq \mathbf{M}^*$. Therefore, the upper bound for $\mathbf{L}_{2,Q}(n)$ in the condition \mathbf{C}_2) is equal to $\mathbf{L}_{2,n}^* = \mathbf{M}^*$. Hence Theorem 3.7. \square

6.4 Proof of Theorem 3.8

By making use of (2.5), (2.6) and applying Ito's formula, one obtains that for any square integrable $\mathbb{R}_+ \rightarrow \mathbb{R}$ functions f and g

$$\mathbf{E} I_t(f) I_t(g) = \int_0^t f(u)g(u) \tilde{\varrho}(u) du. \quad (6.35)$$

Therefore, the condition (1.3) holds with

$$\varkappa_Q = \varkappa_Q(n) = \|\tilde{\varrho}\|_{*,n} = \sup_{0 \leq u \leq n} |\tilde{\varrho}(u)| \leq \|\varrho_{max}\|_{*,n} := \varkappa^*(n).$$

By the condition (2.8)

$$\lim_{n \rightarrow \infty} \frac{1}{n^\delta} \varkappa^*(n) = 0$$

for any $\delta > 0$. Further we will show that the proxy σ_Q in \mathbf{C}_1) can be defined as

$$\sigma_Q = \frac{1}{n} \int_0^n \tilde{\varrho}(u) du. \quad (6.36)$$

From (3.8) and (6.35), it follows that $\mathbf{E} \xi_{1,p}^2 = \sigma_Q$. Moreover, taking into account that $\phi_j^2(t) = 1 + (-1)^j \cos(2\varpi_j t)$ for $j \geq 2$, we obtain that

$\mathbf{E} \xi_{j,p}^2 = \sigma_Q + (-1)^j \mathbf{A}_{j,p}$, where $\mathbf{A}_{j,p} = n^{-1} \sum_{k=1}^{pn} \cos(2\varpi_j t_k) \int_{t_{k-1}}^{t_k} \tilde{\varrho}(u) du$.

First we note that $\sup_{j \geq 1} |\mathbf{A}_{j,p}| \leq \varkappa^*(n)$. We use this estimate to check the condition \mathbf{C}_1) for $p \leq 5$, i.e.

$$\sum_{j=1}^p |\mathbf{E} \xi_{j,p}^2 - \varsigma_Q| \leq \sum_{j=2}^5 |\mathbf{A}_{j,p}| \leq 4 \varkappa^*(n). \quad (6.37)$$

Using the notations

$$\mathbf{A}'_{j,p} = \frac{1}{2n \sin(\varpi_j \Delta)} \sum_{k=1}^{np-1} \mathbf{b}_k \sin(\varpi_j(2t_k + \Delta))$$

and $\mathbf{b}_k = \int_{t_k}^{t_{k+1}} \tilde{\varrho}(u) du - \int_{t_{k-1}}^{t_k} \tilde{\varrho}(u) du$, one can represent $\mathbf{A}_{j,p}$ as

$$\mathbf{A}_{j,p} = \frac{1}{2n} \int_0^\Delta (\tilde{\varrho}(u + n - \Delta) - \tilde{\varrho}(u)) du - \mathbf{A}'_{j,p}.$$

This implies that $|\mathbf{A}_{j,p}'| \leq \Delta \varkappa^*(n)/n + |\mathbf{A}_{j,p}'|$. The last term can be rewritten as

$$\mathbf{A}_{j,p}' = \frac{\sum_{k=2}^{np} (\mathbf{b}_k - \mathbf{b}_{k-1}) \cos(2\varpi_j t_k) + \mathbf{b}_1 \cos(2\varpi_j \Delta) - \mathbf{b}_{np}}{4n \sin^2(\varpi_j \Delta)}.$$

In view of the conditions (2.9), we obtain $\sup_{k \geq 1} |\mathbf{b}_k| \leq \varrho_*' \Delta^2$ and $\sup_{k \geq 2} |\mathbf{b}_k - \mathbf{b}_{k-1}| \leq \varrho_*'' \Delta^3$. This yields

$$|\mathbf{A}_{j,p}'| \leq \frac{(\varrho_*' + 2\varrho_*'')\Delta^2}{4 \sin^2(\varpi_j \Delta)}.$$

Note that $0 < \varpi_j \Delta \leq \pi/2$ for $2 \leq j \leq \check{i}_* = 2[p/4] + 1$ and $\pi/2 < \varpi_j \Delta \leq \pi$ for $\check{i}_* < j \leq p$. Therefore, similarly to (6.22) one gets $\sin(\varpi_j \Delta) \geq j\Delta$ for $2 \leq j \leq \check{i}_*$. For $\check{i}_* < j \leq p-1$, we get $\sin(\varpi_j \Delta) = \sin(\pi - \varpi_j \Delta) \geq (p-j)\Delta$. Thus, estimating $|\mathbf{A}_{p,n}'|$ from above $\varkappa^*(n)$, we obtain that for $p \geq 9$

$$\begin{aligned} \sum_{j=1}^p |\mathbf{E} \xi_{j,p}^2 - \varsigma_Q| &\leq 2\varkappa^*(n) + \sum_{j=2}^{\check{i}_*} |\mathbf{A}_{j,p}'| + \sum_{j=\check{i}_*+1}^{p-1} |\mathbf{A}_{j,p}'| \\ &\leq 2\varkappa^*(n) + \check{\varrho}_* \left(\sum_{j=2}^{\check{i}_*} j^{-2} + \sum_{j=\check{i}_*+1}^{p-1} (p-j)^{-2} \right) \end{aligned}$$

where $\check{\varrho}_* = (\varrho_*' + 2\varrho_*'')/4$. From here taking into account that $\sum_{j \geq 1} j^{-2} \leq 2$, one gets

$$\sup_{p \geq 1} \sum_{j=1}^p |\mathbf{E} \xi_{j,p}^2 - \varsigma_Q| \leq \mathbf{L}_{1,n}^*.$$

Therefore, the condition $\mathbf{C}_1)$ holds. It remains to verify the conditions $\mathbf{C}_2)$. We have

$$\mathbf{E} \left(\sum_{j=1}^p x_j \tilde{\xi}_{j,p} \right)^2 = \frac{1}{n^2} \sum_{i=1}^p \sum_{j=1}^p x_i x_j \mathbf{E} \tilde{I}_n(\psi_{i,p}) \tilde{I}_n(\psi_{j,p}), \quad (6.38)$$

where $\tilde{I}_t(f) = I_t^2(f) - \mathbf{E} I_t^2(f)$. By applying Ito's formula one can calculate that for any bounded $\mathbb{R}_+ \rightarrow \mathbb{R}$ functions f and g

$$\mathbf{E} \tilde{I}_n(f) \tilde{I}_n(g) = 2\tilde{\tau}_{f,g}^2(n) + \lambda \mathbf{E} Y_1^4 \int_0^n f^2(t) g^2(t) \varrho_2^2(t) dt, \quad (6.39)$$

where $\tilde{\tau}_{f,g}(t) = \int_0^t f(u) g(u) \tilde{\varrho}(u) du$. First, we take $f = g = \psi_{i,p}$ and get

$$\mathbf{E} \tilde{I}_n^2(\psi_{j,p}) \leq 8 n^2 \varkappa^*(n) + 4 \mathbf{E} Y_1^4 n (\varkappa^*(n))^2 \lambda_*^{-1}.$$

If $f = \psi_{i,p}$ and $g = \psi_{j,p}$ then

$$\tilde{\tau}_{\psi_{i,p}, \psi_{j,p}}(n) = - \sum_{k=1}^{np-1} \int_{t_{k-1}}^{t_k} (\tilde{\varrho}(u + \Delta) - \tilde{\varrho}(u)) du \sum_{s=1}^k \phi_i(t_s) \phi_j(t_s).$$

Now, taking into account the orthogonality property (3.3) for $i \neq j$ we get that for the $k = rp + \iota$ with $1 \leq \iota < p$

$$\left| \sum_{s=1}^k \phi_i(t_s) \phi_j(t_s) \right| = \left| \sum_{s=1}^{\iota} \phi_i(t_s) \phi_j(t_s) \right| \leq p \|\phi_i\|_p \|\phi_j\|_p = p.$$

Thus, in view of the condition (2.9)

$$\left| \tilde{\tau}_{\psi_{i,p}, \psi_{j,p}}(n) \right| \leq \sum_{k=1}^{np-1} \int_{t_{k-1}}^{t_k} \frac{|\tilde{\varrho}(u + \Delta) - \tilde{\varrho}(u)|}{\Delta} du \leq 2 \int_0^n \left| \frac{d}{dt} \tilde{\varrho}(t) \right| dt \leq 2 \varrho'_* \ln(n+1).$$

Therefore, for $i \neq j$ and $n \geq 3$

$$\begin{aligned} \left| \mathbf{E} \tilde{I}_n(\psi_{i,p}) \tilde{I}_n(\psi_{j,p}) \right| &\leq 8 \left(\varrho'_* \right)^2 \ln^2(n+1) + 4 \mathbf{E} Y_1^4 n (\varkappa^*(n))^2 \lambda_*^{-1} \\ &\leq 11(\ln 2)^2 \left(\varrho'_* \right)^2 n + 4 \mathbf{E} Y_1^4 n (\varkappa^*(n))^2 \lambda_*^{-1}. \end{aligned}$$

This directly implies that the condition \mathbf{C}_2) holds with the parameter $\mathbf{L}_{2,n}^*$ given in (3.47). Hence Theorem 3.8. \square

6.5 Proof of Theorem 3.9

One can check directly that in this case for 1 periodic functions f and g

$$\mathbf{E} I_n(f) I_n(g) = n (f, g) \quad \text{and} \quad \mathbf{E} \tilde{I}_n(f) \tilde{I}_n(g) = 2n^2 (f, g)^2.$$

This implies the conditions \mathbf{C}_1) and \mathbf{C}_2) with $\sigma_Q = 1$, $\mathbf{L}_{1,Q}(n) = 0$ and $\mathbf{L}_{2,Q}(n) = 2$. Therefore, through Corollary 3.2 we obtain Theorem 3.9. \square

6.6 Proof of Theorem 5.4

First, we note that in view of (3.8) one can represent the quadratic risk with the norm (3.2) for the estimate (3.10) as

$$\mathbf{E}_Q \|\widehat{S}_{\gamma_0} - S\|_p^2 = \frac{1}{n} \sum_{j=1}^p \gamma_0^2(j) \mathbf{E}_Q \xi_{j,p}^2 + \overline{\Theta}_p,$$

where $\overline{\Theta}_p = \sum_{j=1}^p (\theta_{j,p} - \gamma_0(j) \bar{\theta}_{j,p})^2$. The first term can be estimated by the condition \mathbf{H}_1) as

$$\sup_{Q \in \mathcal{Q}_n} \mathbf{E}_Q \sum_{j=1}^p \gamma_0^2(j) \xi_{j,p}^2 \leq \sigma^*(n) \sum_{j=1}^p \gamma_0^2(j) + \mathbf{L}_{1,n}^*.$$

Therefore, taking into account that $d_n = n/\sigma^*(n)$, we get

$$\sup_{Q \in \mathcal{Q}_n} \mathbf{E}_Q \|\widehat{S}_{\gamma_0} - S\|_p^2 \leq \frac{1}{d_n} \sum_{j=1}^p \gamma_0^2(j) + \frac{\mathbf{L}_{1,n}^*}{n} + \overline{\Theta}_p.$$

To examine the first term on the right hand of this inequality, we note that

$$\lim_{n \rightarrow \infty} \frac{1}{d_n^{1/(2k+1)}} \sum_{j=1}^p \gamma_0^2(j) = \frac{2(\tau_k r)^{1/(2k+1)} k^2}{(k+1)(2k+1)}. \quad (6.40)$$

Furthermore, by the inequality (6.6) for any $0 < \tilde{\varepsilon} < 1$ we get

$$\overline{\Theta}_p \leq (1 + \tilde{\varepsilon}) \Theta_p + (1 + \tilde{\varepsilon}^{-1}) \sum_{j=1}^p h_{j,p}^2, \quad (6.41)$$

where $\Theta_p = \sum_{j=1}^p (1 - \gamma_0(j))^2 \theta_{j,p}^2$. In view of the definition (3.37), we can represent this term as $\Theta_p = \sum_{j=\iota_0}^{[\omega_0]} (1 - \gamma_0(j))^2 \theta_{j,p}^2 + \sum_{j=[\omega_0]+1}^p \theta_{j,p}^2 := \Theta_{1,p} + \Theta_{2,p}$, where $\iota_0 = j_*(\alpha_0)$, $\omega_0 = \omega_{\alpha_0} = (\tau_k t_0 n)^{1/(2k+1)}$ and $t_0 = [r/\sigma^*(n)\varepsilon] \varepsilon$. By applying Lemma A.5 and the inequality (6.6) yield

$$\Theta_{1,p} \leq (1 + \tilde{\varepsilon}) \sum_{j=\iota_0}^{[\omega_0]} (1 - \gamma_0(j))^2 \theta_j^2 + 4\pi^2 r (1 + \tilde{\varepsilon}^{-1}) \omega_0^3 p^{-2}.$$

Similarly, through Lemma A.4

$$\Theta_{2,p} \leq (1 + \tilde{\varepsilon}) \sum_{j \geq [\omega_0] + 1} \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) r p^{-2}.$$

Hence,

$$\Theta_p \leq (1 + \tilde{\varepsilon}) \Theta_{\iota_0}^* + (1 + \tilde{\varepsilon}^{-1}) (4\pi^2 r \omega_0^3 + r) p^{-2},$$

where $\Theta_{\iota}^* = \sum_{j \geq \iota} (1 - \gamma_0(j))^2 \theta_j^2$.

To estimate the last term on the right side of (6.41) we note that

$$\sup_{S \in W_r^1} \max_{1 \leq j \leq p} h_{j,p}^2 \leq \|\dot{S}\|^2 p^{-2} \leq r p^{-2}.$$

Moreover, $W_r^k \subseteq W_r^2$ for any $k \geq 2$. From here and Lemma A.6 one has for any $k \geq 1$ the inequality

$$\sup_{S \in W_r^k} \sum_{j=1}^p h_{j,p}^2 \leq r \left(p^{-1} \mathbf{1}_{\{k=1\}} + 3p^{-2} \mathbf{1}_{\{k \geq 2\}} \right).$$

Moreover, in view of the conditions (3.32) and \mathbf{H}_3)

$$\lim_{n \rightarrow \infty} d_n^{2k/(2k+1)} \left(p^{-1} \mathbf{1}_{\{k=1\}} + \omega_0^3 p^{-2} \right) = 0.$$

So, $\limsup_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \bar{\Theta}_p \leq \limsup_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \Theta_{\iota_0}^*$. To estimate the term $\Theta_{\iota_0}^*$ we set $v_n = d_n^{2k/(2k+1)} \sup_{j \geq \iota_0} (1 - \gamma_0(j))^2 / a_j$, where the sequence $(a_j)_{j \geq 1}$ is defined in (5.1). This leads to the inequality

$$\sup_{S \in W_r^1} d_n^{2k/(2k+1)} \Theta_{\iota_0}^* \leq v_n \sum_{j \geq 1} a_j \theta_j^2 \leq v_n r.$$

The conditions (3.35) imply $\lim_{n \rightarrow \infty} \sigma^*(n) t_0 = r$, and therefore

$$\limsup_{n \rightarrow \infty} v_n \leq \pi^{-2k} (\tau_k r)^{-2k/(2k+1)},$$

where the coefficient τ_k is given in (3.37). This implies immediately that

$$\limsup_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \bar{\Theta}_p \leq \frac{r^{1/(2k+1)}}{\pi^{2k} (\tau_k)^{2k/(2k+1)}}. \quad (6.42)$$

Taking into account that

$$R_k^* = \frac{2(\tau_k r)^{1/(2k+1)} k^2}{(k+1)(2k+1)} + \frac{r^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2k/(2k+1)}}$$

and applying (6.40) and (6.42)

$$\lim_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \sup_{Q \in \mathcal{Q}_n} \mathbf{E}_Q \|\widehat{S}_{\gamma_0} - S\|_p^2 \leq R_k^*. \quad (6.43)$$

Moreover, Lemma A.2 yields that for any $\tilde{\varepsilon} > 0$

$$\sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \leq (1 + \tilde{\varepsilon}) \sup_{S \in W_r^k} \sup_{Q \in \mathcal{Q}_n} \mathbf{E}_Q \|\widehat{S}_{\gamma_0} - S\|_p^2 + (1 + \tilde{\varepsilon}^{-1})r p^{-2}.$$

So, in view of the condition \mathbf{H}_3), we derive the desired inequality

$$\lim_{n \rightarrow \infty} d_n^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \leq R_k^*.$$

Hence Theorem 5.4. \square

7 Conclusion

In this paper, we have developed a model selection method for estimating a continuous time regression with semimartingale noises from discrete data. The discreteness of observations is a key factor which seriously complicates the estimation problem because the efficient estimation procedures proposed in papers [21] – [24] for the continuously observed regression models do not work in this case. Moreover we are faced with the new problem of having to specify the sampling frequency which allows one to construct a robust adaptive efficient procedure. To this end we have derived non-asymptotic oracle inequalities for the robust quadratic risks under some general conditions \mathbf{C}_1) and \mathbf{C}_2). These conditions have been verified in two important examples.

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

8 Appendix

A.1 Upper bound for the penalty term (3.18)

Lemma A.1. *Assume that the condition \mathbf{C}_1) holds. Then for any $n \geq 1$ and $\gamma \in \Gamma$,*

$$P_Q(\gamma) \leq \mathbf{E}_Q \text{Err}(\gamma) + \frac{\mathbf{L}_{1,Q}(n)}{n}.$$

Proof. Taking into account (3.8) and the condition \mathbf{C}_1) in (3.13) we obtain

$$\mathbf{E}_Q \text{Err}(\gamma) \geq \frac{1}{n} \sum_{j=1}^p \gamma^2(j) \mathbf{E} \xi_{j,p}^2 \geq P_Q(\gamma) - \frac{\mathbf{L}_{1,Q}(n)}{n}.$$

Hence Lemma A.1. \square

A.2 Relations between the norms $\|\cdot\|$ and $\|\cdot\|_p$.

Lemma A.2. *Let f be an absolutely continuous $[0, 1] \rightarrow \mathbb{R}$ function with $\|\dot{f}\| < \infty$ and g be a simple $[0, 1] \rightarrow \mathbb{R}$ function of the form $g(t) = \sum_{j=1}^p c_j \chi_{(t_{j-1}, t_j]}(t)$, where c_j are some constants. Then $\forall \tilde{\varepsilon} > 0$ the function $\Delta = f - g$ satisfies the following inequalities*

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

Proof. Applying the inequality (6.6) we obtain

$$\begin{aligned} \|\Delta\|^2 &= \|\Delta\|_p^2 + \sum_{l=1}^p \int_{t_{l-1}}^{t_l} (2\Delta(t_l) (\Delta(t) - \Delta(t_l)) + (\Delta(t) - \Delta(t_l))^2) dt \\ &\leq (1 + \tilde{\varepsilon}) \|\Delta\|_p^2 + (1 + \tilde{\varepsilon}^{-1}) \sum_{l=1}^p \int_{t_{l-1}}^{t_l} [\Delta(t_l) - (\Delta(t))]^2 dt \\ &= (1 + \tilde{\varepsilon}) \|\Delta\|_p^2 + (1 + \tilde{\varepsilon}^{-1}) \sum_{l=1}^p \int_{t_{l-1}}^{t_l} |f(t_l) - f(t)|^2 dt. \end{aligned}$$

Noting that, for $t_{l-1} < t \leq t_l$, one has the estimate

$$|f(t_l) - f(t)|^2 \leq \left(\int_{t_{l-1}}^{t_l} |\dot{f}(u)| du \right)^2 \leq p^{-1} \int_{t_{l-1}}^{t_l} |\dot{f}(u)|^2 du,$$

one comes to the first inequality. Similarly, one can verify the second inequality. Hence Lemma A.2. \square

A.3 Properties of the Fourier coefficients.

Lemma A.3. *Let the function $S(t)$ in (1.1) be absolutely continuous and have an absolutely integrable derivative. Then the coefficients $(\bar{\theta}_{j,p})_{1 \leq j \leq p}$ defined in (3.8) satisfy the inequalities*

$$|\bar{\theta}_{1,p}| \leq |S|_1 \quad \text{and} \quad \max_{2 \leq j \leq p} j |\bar{\theta}_{j,p}| \leq 2\sqrt{2} |\dot{S}|_1, \quad (\text{A.1})$$

where $|S|_1 = \int_0^1 |S(t)| dt$.

Proof. We note, that if $j = 1$, then $|\bar{\theta}_{1,p}| = |\int_0^1 S(t) dt| \leq |S|_1$. If $j \geq 2$, then the coefficients $\bar{\theta}_{j,p}$ can be represented as $\bar{\theta}_{j,p} = \sqrt{2} \Upsilon_j \left(\sum_{k=1}^p \mathbf{q}_j^k \int_{t_{k-1}}^{t_k} S(t) dt \right)$ where $\Upsilon_j = \text{Re}(z)$ for even j and $\Upsilon_j = \text{Im}(z)$ for odd j , $\mathbf{q}_j = e^{i\varpi_j \Delta}$, $\Delta = 1/p$. Taking into account that the function S is 1 - periodic we transform the sum as

$$\sum_{k=1}^p \mathbf{q}_j^k \int_{t_{k-1}}^{t_k} S(t) dt = \mathbf{q}_j (1 - \mathbf{q}_j)^{-1} \sum_{k=1}^p \mathbf{q}_j^k \mathbf{a}_k,$$

where $\mathbf{a}_k = \mathbf{a}_k(S) = \int_{t_k}^{t_{k+1}} S(t) dt - \int_{t_{k-1}}^{t_k} S(t) dt$. Moreover, noting that $|\mathbf{a}_k| \leq \Delta \int_{t_{k-1}}^{t_{k+1}} |\dot{S}(u)| du$, we obtain

$$|\bar{\theta}_{j,p}| \leq \sqrt{2} |1 - \mathbf{q}_j|^{-1} \sum_{k=1}^p |\mathbf{a}_k| \leq 2\sqrt{2} \Delta |\dot{S}|_1 |1 - \mathbf{q}_j|^{-1}.$$

Applying the inequalities (6.22) yields $|1 - \mathbf{q}_j| = 2 \sin(\varpi_j \Delta / 2) \geq j \Delta$ and we obtain the second upper bound in (A.1). Hence Lemma A.3. \square

Lemma A.4. 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 W_r^1 satisfy, for any $\tilde{\varepsilon} > 0$, the following inequality

$$\sum_{j=N}^p \theta_{j,p}^2 \leq (1 + \tilde{\varepsilon}) \sum_{j \geq N} \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) r p^{-2}. \quad (\text{A.2})$$

Proof. Note that $\sum_{j=N}^p \theta_{j,p}^2 = \min_{x_1, \dots, x_{N-1}} \|S - \sum_{j=1}^{N-1} x_j \phi_j\|_p^2 \leq \|f_N\|_p^2$, where $f_N(t) = \sum_{j \geq N} \theta_j \phi_j(t)$. By applying Lemma A.2 and taking into account that $\|f_N\|^2 = \sum_{j \geq N} \theta_j^2$ and $\|\dot{f}_N\|^2 \leq r$, we obtain the bound (A.2). Hence Lemma A.4 \square

Lemma A.5. For any $p \geq 2$ and $r > 0$, the coefficients $(\theta_{j,p})_{1 \leq j \leq p}$ of functions S from the class W_r^1 satisfy the following inequality

$$\max_{1 \leq j \leq p} \sup_{S \in W_r^1} (|\theta_{j,p} - \theta_j| - 2\pi\sqrt{r} j p^{-1}) \leq 0. \quad (\text{A.3})$$

Proof. First of all we note that $\theta_{1,p} = \theta_1$. Therefore, it suffices to check that the upper bound (A.3) holds for $2 \leq j \leq p$. Moreover, note that if $S \in W_r^1$ then the function S is absolutely continuous. Therefore, the coefficients (3.4) and (5.1) satisfy the inequalities

$$|\theta_{j,p} - \theta_j| \leq p^{-1} \int_0^1 \left| (S(u)\phi_j(u))' \right| du \leq \frac{2\varpi_j}{p} \sqrt{\int_0^1 (|S'(u)|^2 + |S(u)|^2) du}.$$

Using here that $\int_0^1 (|S'(u)|^2 + |S(u)|^2) du \leq r$, we obtain (A.3). \square

Lemma A.6. For any $p \geq 2$ and $r > 0$ the correction coefficients $(h_{j,p})_{1 \leq j \leq p}$ for the functions S from the class W_r^2 satisfy the following inequality

$$\sup_{S \in W_r^2} \sum_{j=1}^p h_{j,p}^2 \leq 3r p^{-2}. \quad (\text{A.4})$$

Proof. By making use of the definition of $h_{j,p}$ in (3.7) we can represent it as $h_{j,p} = -\theta'_{j,p}/(2p) + H_{j,p}$, where $\theta'_{j,p} = (S', \phi_j)_p$, $H_{j,p} = \sum_{l=1}^p \phi_j(t_l) v_l$ and

$v_l = \int_{t_{l-1}}^{t_l} \left(\int_t^{t_l} (S'(t_l) - S'(z)) dz \right) dt$. Since the derivative S' belongs to W_r^1 , we can get through the upper bound (A.2) with $\tilde{\varepsilon} = 1$, that for $p \geq 2$

$$\sum_{j=1}^p \left(\theta'_{j,p} \right)^2 = \|S'\|_p^2 \leq 2\|S'\|^2 + 2\|S''\|^2 p^{-2} \leq 2r + 2r p^{-2} < 3r.$$

Since S' is absolutely continuous, i.e. $S'(t_l) - S'(z) = \int_z^{t_l} S''(u) du$, we can rewrite $H_{j,p}$ as $H_{j,p} = 2^{-1} \sum_{l=1}^p \phi_j(t_l) \int_{t_{l-1}}^{t_l} (t - t_{l-1})^2 S''(t) dt$. Therefore, $|H_{j,p}| \leq 2^{-1/2} p^{-2} \int_0^1 |S''(t)| dt \leq 2^{-1/2} p^{-2} \sqrt{r}$ and we get the inequality (A.4). Hence Lemma A.6. \square

References

- [1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19:716–723, 1974.
- [2] O.E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):167–241, 2001.
- [3] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.
- [4] F. Comte and V. Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *The Annals of Statistics*, 39(2):803–837, 2011.
- [5] F. Comte, V. Genon-Catalot, and Y. Rozenholc. Nonparametric adaptive estimation for integrated diffusions. *Stochastic Processes and their Applications*, 119(3):811–834, 2009.
- [6] H. Dehling, B. Franke, and T. Kott. Drift estimation for a periodic mean reversion process. *Stat. Inference Stoch. Process.*, 13(3):175–192, 2010.
- [7] T. E. Duncan, Y. Hu, and B. Pasik-Duncan. Stochastic calculus for fractional brownian motion i. theory. *SIAM, J. Control Optim.*, 38(2):582–612, 2000.

- [8] D. Fourdrinier and S. M. Pergamenshchikov. Improved model selection method for a regression function with dependent noise. *Annals of the Institute of Statistical Mathematics*, 59(3):435–464, 2007.
- [9] L. I. Galtchouk and S. M. Pergamenshchikov. Nonparametric sequential estimation of the drift in diffusion processes via model selection. *Mathematical Methods of Statistics*, 13(1):25–49, 2004.
- [10] L. I. Galtchouk and S. M. Pergamenshchikov. Asymptotically efficient estimates for nonparametric regression models. *Statistics and Probability Letters*, 76(8):852–860, 2006.
- [11] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic non-parametric regression. *Journal of Korean Statistical Society*, 38(4):305–322, 2009.
- [12] L. I. Galtchouk and S. M. Pergamenshchikov. Sharp non-asymptotic oracle inequalities for non-parametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, 21(1):1–18, 2009.
- [13] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic non-parametric regression via model selection. <http://hal.archives-ouvertes.fr/hal-00326910/fr/>, 2013.
- [14] E. Gobet, M. Hoffmann, and M. Reiss. Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, 32(5):2223–2253, 2004.
- [15] M. Hoffmann, A. Munk, and J. Schmidt-Hieber. Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(4):1186–1216, 2012.
- [16] R. Höpfner and Yu. A. Kutoyants. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion. *Statist. Decisions*, 27(4):309 – 326, 2009.
- [17] R. Höpfner and Yu. A. Kutoyants. Estimating discontinuous periodic signals in a time inhomogeneous diffusion. *Statistical Inference for Stochastic Processes*, 13(3):193 – 230, 2010.

- [18] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer, New York, 1984.
- [19] A. Kneip. Ordered linear smoothers. *The Annals of Statistics*, 22(2):835–866, 1994.
- [20] V. V. Konev and S. M. Pergamenshchikov. Sequential estimation of the parameters in a trigonometric regression model with the Gaussian coloured noise. *Statistical Inference for Stochastic Processes*, 6(3):215–235, 2003.
- [21] V. V. Konev and S. M. Pergamenshchikov. Nonparametric estimation in a semimartingale regression model. Part 1. Oracle inequalities. *Journal of Mathematics and Mechanics, Tomsk State University*, 3(7):23–41, 2009.
- [22] V. V. Konev and S. M. Pergamenshchikov. Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Journal of Mathematics and Mechanics, Tomsk State University*, 4(8):31–45, 2009.
- [23] V. V. Konev and S. M. Pergamenshchikov. General model selection estimation of a periodic regression with a Gaussian noise. *Annals of the Institute of Statistical Mathematics*, 62(6):1083–1111, 2010.
- [24] V. V. Konev and S. M. Pergamenshchikov. Efficient robust nonparametric estimation in a semimartingale regression model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(4):1217–1244, 2012.
- [25] C. L. Mallows. Some comments on C_p . *Technometrics*, 15(4):661–675, 1973.
- [26] P. Massart. A non-asymptotic theory for model selection. In Ari Laptev, editor, *European Congress of Mathematics*, pages 309–323. Eur. Math. Soc., Zürich, 2005.
- [27] M. Nussbaum. Spline smoothing in regression models and asymptotic efficiency in L_2 . *The Annals of Statistics*, 13(3):984–997, 1985.
- [28] M. S. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission*, 16(2):52–68, 1980.