Optimal algorithms of data processing for complex information and telecommunication systems in an uncertain environment

Slim Beltaief

Normandie Université

THESE

Pour obtenir le diplôme de doctorat

Spécialité (Mathématiques Appliquées)

Préparée au sein de « l'Université de Rouen Normandie »
laboratoire de mathématiques Raphaël Salem

Algorithmes optimaux de traitement de données pour des systèmes complexes d'informations et de télécommunication dans un environnement incertain

Présentée et soutenue par
Slim BELTAIEF

Thèse soutenue publiquement le (8 septembre 2017) devant le jury composé de		
M. Dominique Fourdrinier	Professeur, Université de Rouen Normandie	Président de jury
M. Yury A. Kutoyants	Professeur, Université du Maine	Rapporteur
M. Nicolaos Limnios	Professeur, Université de Technologie de Compiègne	Rapporteur
M. Igor Nikiforov	Professeur, Université de Technologie de Troyes	Examinateur
Mme. Ghislaine Gayraud	Professeur, Université de Technologie de Compiègne	Examinatrice
M. Serguei Pergamenshchikov	Professeur, Université de Rouen Normandie	Directeur de thèse
M. Vlad Stefan BARBU	Maître de conférences, Université de Rouen Normandie	Codirecteur de thèse

UNIVERSITÉ DE ROUEN NORMANDIE

LMRS

Résumé

Ce travail est consacré au problème d'estimation non paramétrique dans des modèles de regression en temps continu. On considère le problème d'estimation d'une fonction inconnue S supposée périodique. Cette estimation est basée sur des observations générées par un processus stochastique; ces observations peuvent êtres en temps continu ou discret. Pour ce faire, nous construisons une série d'estimateurs par projection et nous approchons la fonction inconnue S par une série de Fourier finie. Dans cette thèse, nous considérons le problème d'estimation dans le cadre adaptatif, c'est-à-dire le cas où la régularité de la fonction S est inconnue. Pour ce problème, nous développons une nouvelle méthode d'adaptation basée sur la procédure de sélection de modèle proposée par Konev et Pergamenshchikov (2012). Tout d'abord, cette procédure nous donne une famille d'estimateurs; après nous choisissons le meilleur estimateur possible en minimisant une fonction coût. Nous donnons également une inégalité d'Oracle pour le risque de nos estimateurs et nous donnons la vitesse de convergence minimax.

Mots Clés : Estimation non asymptotique, sélection de modèle, inégalité d'Oracle pointue, risque robuste , efficacitée asymptotique.

Abstract

This thesis is devoted to the problem of non parametric estimation for continuous-time regression models. We consider the problem of estimating an unknown periodoc function S. This estimation is based on observations generated by a stochastic process; these observations may be in continuous or discrete time. To this end, we construct a series of estimators by projection and thus we approximate the unknown function S by a finite Fourier series. In this thesis we consider the estimation problem in the adaptive setting, i.e. in situation when the regularity of the fonction S is unknown. In this way, we develop a new adaptive method based on the model selection procedure proposed by Konev and Pergamenshchikov (2012). Firstly, this procedure give us a family of estimators, then we choose the best possible one by minimizing a cost function. We give also an oracle inequality for the risk of our estimators and we give the minimax convergence rate.

Key Words: Non asymptotic estimation, Robust risk, Model selection, Sharp oracle inequality, Assymptotic efficiency.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problem of non parametric estimation in regression models has an important role in theorical and applied statistics. In this thesis, we consider regression models in continuous time of the form

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

where S is an unknown periodic function from $\mathbf{L}_2[0,n]$, wich we want to estimate on the basis of observations y_t. This observations can be in continuous time or in discrete time. This problem was considered in many frameworks, for example, in the framework of the "signal+white noise" models (see, for example, [9] or [47]). Later, in order to study dependent observations in continuous time, were introduced "signal+color noise" regressions based on Ornstein-Uhlenbeck processes (cf. [11], [12], [13], [16]).

Moreover, to include jumps in such models, the papers [17] and [18] used non Gaussian Ornstein-Uhlenbeck processes introduced in [4] for modeling the risky assets in the stochastic volatility financial markets. Unfortunately, the dependence of the stable Ornstein-Uhlenbeck type decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain very quickly the same "signal+white noise" model.

The main goal of this thesis is to develop the model (1.1) for the noise process with large dependence. This allows us to consider the signal estimation problem for signals observed with noises of complex structure "against-signal". To achieve this goal, we use semi-Markov processes to model the dependent noises, because it is well known that such processes keep the dependence for a long time.

In our work we use the robust estimation approach introduced in [17] for such problems. To this end, we denote by Q the distribution of $(\xi_t)_{0\le t\le n}$ in the Skorokhod space $\mathcal{D}[0,n]$. We assume that Q is unknown and belongs to some distribution family $\mathcal{Q}_n$ wich will be specified later. In this thesis we use the quadratic risk

$$\mathcal{R}_Q(\widetilde{S}_n, S) = \mathbf{E}_{Q,S}\, \|\widetilde{S}_n - S\|^2\,, \tag{1.2}$$

where $\|f\|^2 = \int_0^1 f^2(s)\mathrm{d}s$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1.1) corresponding to the noise distribution Q. Since the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

$$\mathcal{R}^*_n(\widetilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widetilde{S}_n, S)\,, \tag{1.3}$$

which enables us to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate $\widetilde{S}_n$ for all distributions in the family $\mathcal{Q}_n$.

In order to estimate the function S belonging to $\mathbf{L}_2[0,n]$, one can consider a projection estimator and thus approximate S by a finite Fourier series. Following Pinsker's approach (1981), we use the weighted least square estimators which provide efficient estimation, but, in order to obtain efficient estimation, one needs to use regularity properties of the function S. Our approach is to consider the estimation problem in the adaptive setting, i.e. in situation when the regularity of the function S is unknown. In this way, we develop a new adaptive method based on the model selection procedure proposed by Konev and Pergamenshchikov (2012). To show the efficiency, one needs to obtain the corresponding sharp oracle inequality; note that this is a crucial tool in order to be able to show the adaptive efficiency. The "sharp" oracle inequality means that the upper bound for the risk has the coefficient of the main term close to one.

We recall that the main term is the minimal risk over the family of estimators that we consider. To obtain the oracle inequality one needs to develop renewal theory for the model (1.1). In our thesis we obtain a new asymptotic development for the renewal density. In fact, this result is a version of Goldie's theorem (1991). Unfortunately, we cannot use directly the Goldie's theorem, since in that result there is a singular component of the renewal distribution, which makes the use of that result impossible for the estimation purposes. In our work we give sufficient conditions for having an asymptotic development for the renewal density without a singular component.

The effeciency of the estimator will be also proved. To this end, we assume that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W^k_{\mathbf{r}} = \{f \in \mathcal{C}^k_{per}[0,1] \, : \, \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq \mathbf{r}\}\,, \tag{1.4}$$

where $\mathbf{r} > 0$ and $k \geq 1$ are some unknown parameters, $\mathcal{C}^k_{per}[0,1]$ 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 $0 \leq i \leq k$. Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (1.3) is given by

$$\mathbf{r}^*_k = \, ((2k+1)\mathbf{r})^{1/(2k+1)} \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)} . \tag{1.5}$$

Note that this is the well-known Pinsker's constant obtained for the nonadaptive filtration problem in "signal + small white noise" model (see, for example, [47]), generated by the process (1.1).

Main results of the thesis

In this thesis we study three types of regression models in continuous time, the observations are generated mainly by a semi-Markov process and Lévy process. In this way, our model is

capable to take into account "small" jumps, thanks to the Lévy process, as well as "big" jumps, thanks to the semi-Markov process.

1.1 Semi-Markov process

In our work, we consider a pure jump process $(z_t)_{t\geq 0}$ as a semi-Markov process with the following form

$$z_t = \sum_{i=1}^{N_t} Y_i, \tag{1.6}$$

where $(Y_i)_{i\geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E}\, Y_i = 0\,, \quad \mathbf{E}\, Y_i^2 = 1 \quad \text{and} \quad \mathbf{E}\, Y_i^4 < \infty\,.$$

Here N_t is a general counting process (see, for example, [22]) defined as

$$N_t = \sum_{k=1}^{\infty} 1_{\{T_k \leq t\}} \quad \text{and} \quad T_k = \sum_{l=1}^{k} \tau_l\,, \tag{1.7}$$

where $(\tau_l)_{l\geq 1}$ is an i.i.d. sequence of positive integrated random variables with distribution η and mean $\check{\tau} = \mathbf{E}\,\tau_1 > 0$. We assume that the processes $(N_t)_{t\geq 0}$ and $(Y_i)_{i\geq 1}$ are independent. Note that the process $(z_t)_{t\geq 0}$ is a special case of a semi-Markov process (see, e.g., [5] and [20]). It should be noted that if τ_j are exponential random variables, then $(N_t)_{t\geq 0}$ is a Poisson process and, in this case, the process $(\xi_t)_{t\geq 0}$ given in (1.14) is a Lévy process for which this model has been studied in [14], [15] and [17]. But, in the general case when the process (1.6) is not a Lévy process, this process has a memory and cannot be treated in the framework of semi-martingales with independent increments. In this case, we need to develop new tools based on renewal theory arguments, what we do in Chapter.5. This tools will be intensively used in the proofs of our main results.
Let ρ be the density of the renewal measure $\check{\eta}$ defined as

$$\check{\eta} = \sum_{l=1}^{\infty} \eta^{(l)}\,, \tag{1.8}$$

where $\eta^{(l)}$ is the lth convolution power of η. To study this renewal measure we assume that the measure η has a density g which satisfies the following conditions.

$\mathbf{H}_1$) *Assume that, for any $x \in \mathbb{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and, for any $K > 0$, *there exists* $\delta = \delta(K) > 0$ *for which*

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

$\mathbf{H}_2$) *For any* $\gamma > 0$,

$$\sup_{z\ge 0} z^\gamma |2g(z) - g(z-) - g(z+)| \, < \, \infty.$$

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

Remark 1.1. It should be noted that the condition $\mathbf{H}_3$) means that there exists an exponential moment for the random variable $(\tau_j)_{j\ge 1}$, i.e. these random variables are not too large. This is a natural constraint since these random variables define the intervals between jumps, i.e., the frequency of the jumps. So, to study the influence of the jumps in the model (1.13) one needs to consider the noise process (1.14) with "small" interval between jumps or large jump frequency.

For the next condition we need to introduce the Fourier transform of any function f from $\mathbf{L}_1(\mathbb{R})$, $f : \mathbb{R} \to \mathbb{R}$, defined as

$$\widehat{f}(\theta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\theta x}\, f(x)\, \mathrm{d}x. \tag{1.9}$$

$\mathbf{H}_4$) *There exists* $t^* > 0$ *such that the function* $\widehat{g}(\theta - it)$ *belongs to* $\mathbf{L}_1(\mathbb{R})$ *for any* $0 \le t \le t^*$.

Remark 1.2. It is clear that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true for any continuously differentiable function g, for example for the exponential density.

Proposition 1.1. *Let* τ *be a positive random variable with the distribution* η *having a density* g *which satisfies Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$). *Then the renewal measure* (1.8) *is absolutely continuous with density* ρ, *for which*

$$\rho(x) = \frac{1}{\check{\tau}} + \Upsilon(x)\,, \tag{1.10}$$

where $\check{\tau} = \mathbf{E}\tau_1$ *and* $\Upsilon(\cdot)$ *is some function defined on* $\mathbb{R}_+$ *with values in* $\mathbb{R}$ *such that*

$$\sup_{x\ge 0} x^\gamma |\Upsilon(x)| < \infty \quad \textit{for all} \quad \gamma > 0\,.$$

It should be noted that in view of this proposition, Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) imply

$$|\rho|_* = \sup_{t\ge 0} |\rho(t)| < \infty \quad \text{and} \quad \|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)|\, \mathrm{d}x \, < \infty \tag{1.11}$$

1.2 Lévy process

In this thesis we assume that the Lévy process L_t is defined as

$$L_t = \breve{\varrho}\, w_t + \sqrt{1-\breve{\varrho}^2}\, \check{L}_t\,, \quad \check{L}_t = x * (\mu - \widetilde{\mu})_t\,, \tag{1.12}$$

where, $0 \le \breve{\varrho} \le 1$ is an unknown constant, $(w_t)_{t\ge 0}$ is a standard Brownian motion, $\mu(\mathrm{d}s, \mathrm{d}x)$ is the jump measure with the deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, where $\Pi(\cdot)$ is some positive measure on $\mathbb{R}$ (see, for example, [10, 6] for details). In this thesis, we use the usual notation

$$\Pi(|x|^m) = \int_{\mathbb{R}} |z|^m\, \Pi(\mathrm{d}z) \quad \text{for any} \quad m > 0.$$

Note that $\Pi(\mathbb{R})$ may be equal to $+\infty$.

Remark 1.3. In this thesis, we assume that the processes $(N_t)_{t\ge 0}$ and $(Y_i)_{i\ge 1}$ in (1.6) are independent between them and are also independent of $(L_t)_{t\ge 0}$.

1.3 Non-parametric estimation based on continuous data

We consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{1.13}$$

where S is an unknown 1-periodic function from $\mathbf{L}_2[0,1]$ defined on $\mathbb{R}$ with values in $\mathbb{R}$, the noise process $(\xi_t)_{t\ge 0}$ is defined as

$$\xi_t = \varrho_1 L_t + \varrho_2 z_t\,, \tag{1.14}$$

where ϱ_1 and ϱ_2 are unknown coefficients, the pure jump process $(z_t)_{t\ge 1}$ is the semi-Markov process defined in (1.6) and $(L_t)_{t\ge 0}$ is the Levy process defined in (1.12), for which we assume that

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

The problem is to estimate the unknown function S in the model (1.13) on the basis of observations $(y_t)_{0\le t\le n}$. The main goal is to consider continuous time regression models with dependent observations for which the dependence does not disappear for a sufficient large duration of observations. To this end we define the noise in the model through a semi-Markov process which keeps the dependence for any duration n. This type of models allows, for example, to estimate the signals observed under long impulse noise impact with a memory or "against signals".
To estimate the function S we use here the model selection procedure for continuous time

regression models from [17] based on the Fourrier expansion. We recall that for any function S from $\mathbf{L}_2[0,1]$ we can write

$$S(t) = \sum_{j=1}^{\infty} \theta_j \, \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j) = \int_0^1 S(t)\phi_j(t)\mathrm{d}t \, ,$$

where $(\phi_j)_{j\geq 1}$ is an orthonormal uniformly bounded basis in $\mathbf{L}_2[0,1]$, i.e., for some constant $\phi_* \geq 1$, which may be depend on n,

$$\sup_{0\leq j\leq n} \sup_{0\leq t\leq 1} |\phi_j(t)| \, \leq \, \phi_* < \infty \, .$$

We extend the functions $\phi_j(t)$ by periodicity, i.e., we set $\phi_j(t) := \phi_j(\{t\})$, where $\{t\}$ is the fractional part of $t \geq 0$. For example, we can take the trigonometric basis $(\mathrm{Tr}_j)_{j\geq 1}$ in $\mathbf{L}_2[0,1]$ defined as $\mathrm{Tr}_1 \equiv 1$ and, for $j \geq 2$,

$$\mathrm{Tr}_j(x) = \sqrt{2} \begin{cases} \cos(2\pi[j/2]x) & \text{for even} \quad j, \\ \\ \sin(2\pi[j/2]x) & \text{for odd} \quad j, \end{cases} \tag{1.15}$$

where $[x]$ denotes the integer part of x.
Now, for obtaining efficient estimation one needs to use weighted least square estimators defined as

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(t) \, ,$$

where the coefficients $\lambda = (\lambda(j))_{1\leq j\leq n}$ belong to some finite set Λ from $[0,1]^n$. In this thesis we consider the adaptive case, i.e. we assume that the regularity of the function S is unknown. In this case we chose the weight coefficients on the basis of the model selection procedure proposed in [17] for the general semi-martingale regression model in continuous time.
Now, to choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}} \, . \tag{1.16}$$

Now, in order to study the asymptotic effeciency for our procedure, we give the following oracle inequality for the robust risk defined in (1.3) and through a specific distribution family.

Theorem 1.1. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_4$) *hold and that the unknown function S is continuously differentiable. Then, for any $n \geq 1$ and $0 < \delta < 1/6$, the procedure* (1.16) *satisfy the following oracle inequality*

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

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that, under some conditions, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\le r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0.$$

It follows directly, by using the oracle inequality, that our procedure is efficient with the efficient robust rate $\upsilon_n^{2k/(2k+1)}$ with $\upsilon_n = n/\varsigma^*(n)$ such that, for any $\check{\epsilon} > 0$,

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

Corollary 1.1.

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^*\,. \tag{1.17}$$

We prove in this thesis that the robust minimax convergence rate may be faster or slower than the classical one $(n^{2k/(2k+1)})$.

1.4 Non-parametric estimation based on discrete data

In this chapter we consider the regression model (1.13) with the noise (1.14). The problem is to estimate the unknown function S in the case when continuous observation cannot be provided and only discret time measurement are available, the observations are in the forme

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

where the integer $p \ge 1$ is the observation frequency. We will use the trigonometric basis defined in (1.15). By making use of this basis we consider the discrete Fourier transformation of S

$$S(t) = \sum_{j=1}^{p} \theta_{j,p}\,\mathrm{Tr}_j(t), \quad t \in \{t_1, ..., t_p\},$$

where the Fourier coefficients are defined by

$$\theta_{j,p} = (S, \mathrm{Tr}_j)_p = \frac{1}{p}\sum_{i=1}^{p} S(t_i)\mathrm{Tr}_j(t_i).$$

Then, we estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\Psi_{j,p}(t)\,,$$

where the weight vector $\lambda = (\lambda(1), ..., \lambda(n))$ belongs to some finite set Λ from $[0,1]^n$,

$$\widehat{\theta}_{j,p} = \frac{1}{n}\int_0^n \Psi_{j,p}(t)\mathrm{d}\,y_t \quad \text{and} \quad \Psi_{j,p}(t) = \sum_{l=1}^{np} \mathrm{Tr}_j(t_l)1_{\{t_{l-1}<t\le t_l\}}\,.$$

In order to find a proper weight sequence λ in the set Λ one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. Let us consider the empirical squared error

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}} . \tag{1.18}$$

Now, To obtain the oracle inequality and to prove the effeciency of our procedure, we obtain in this thesis the sufficient condition for the frequency observations p.
$\mathbf{H}_5$) We assume that there exists $\check{\delta} > 0$ such that for any $n \geq 3$

$$p \geq n^{5/6} .$$

Now, in order to study the asymptotic effeciency for our procedure, we give the following oracle inequality for the robust risk defined in (1.3) and through a specific distribution family.

Theorem 1.2. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_5$) *hold and that the unknown function* S *is continuously differentiable. Then, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, the procedure* (1.18) *satisfy the following oracle inequality*

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

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that, under some conditions, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\leq r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0.$$

1.5 Non-parametric estimation for Lévy regression models

We consider a regression model in continuous time with the Lévy noise

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \varepsilon\, \mathrm{d}\xi_t , \quad 0 \leq t \leq 1 , \tag{1.19}$$

where S is an unknown function defined on $\mathbb{R}$ with values in $\mathbb{R}$, $(\xi_t)_{0\leq t\leq 1}$ is a Lévy process and $\varepsilon > 0$ is the noise intensity. The problem is to estimate the function S based on the continuous observations $(y_t)_{0\leq t\leq 1}$ when $\varepsilon \to 0$. We assume that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^4) < \infty . \tag{1.20}$$

When constructing the sharp model selection procedures, in this model, we will use the approach close to that of the papers [14], [15], [16], [18] developed for the estimation of a 1-periodic function in continuous time on a large time interval, i.e.

$$\mathrm{d}x_t = S(t)\mathrm{d}t + \mathrm{d}\eta_t\,, \quad 0 \le t \le n\,.$$

Note that, for any $0 < t < 1$, setting $y_t = n^{-1}\sum_{j=1}^{n}(x_{t+j} - x_j)$, we can represent this model as a model with small parameter of form (1.19)

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon\,\mathrm{d}\xi_t\,,$$

where $\varepsilon = n^{-1/2}$ and $\xi_t = n^{-1/2}\sum_{j=1}^{n}(\eta_{t+j} - \eta_t)$. The main difference between this model and the original one is that the jumps are small, i.e.

$$\Delta\xi_t = \mathrm{O}(n^{-1/2}) = \mathrm{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0\,,$$

but we have not such property in the model (1.19). Therefore, unfortunately, we cannot use directly the method developed for the estimation problem on the large time interval to the model (1.19). So, the main goal of this chapter is to develop a new sharp model selection method for the estimation problem of the function S as $\varepsilon \to 0$. Let $(\phi_j)_{j\ge 1}$ be an orthonormal basis in $\mathbf{L}_2[0,1]$ with $\phi_1 \equiv 1$. We assume that this basis is uniformly bounded, i.e. for some constant $\phi^* \ge 1$, which may depend on $\varepsilon > 0$,

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

where $n = n_\varepsilon = [1/\varepsilon^2]$ and $[x]$ denotes the integer part of x. To estimate the function S we use the following Fourier series

$$S(t) = \sum_{j\ge 1}\theta_j\,\phi_j(t).\,,$$

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

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

where $n = [1/\varepsilon^2]$, the weights $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$. To choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}\,. \tag{1.21}$$

Theorem 1.3. *Assume that for the model* (1.19) *the condition* (1.20) *holds and the unknown function S is continuously differentiable. Then, for any $0 < \delta < 1/6$, the procedure* (1.21) *satisfy the following oracle inequality*

$$\mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*_\varepsilon(\widehat{S}_\lambda, S) + \varepsilon^2 \, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta}\,, \tag{1.22}$$

where the term $\mathbf{U}^*_\varepsilon(S) > 0$ is such that under some conditions, for any $r > 0$ and $\check{\delta} > 0$

$$\lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \sup_{\|\dot{S}\|\le r} \mathbf{U}^*_\varepsilon(S) \, = 0\,. \tag{1.23}$$

As an application of the sharp model selection method in this thesis, we consider the estimation problem for the number of signals in the multi-path connexion channel. In the framework of the statistical radio - physics models we study the telecommunication system in which we observe in the multi-path channel the summarized signal with the noise on the time interval $[0,1]$,

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

where $(n_t)_{t\ge 0}$ is some noise, usually this "white noise". The energetic parameters $(\theta_j)_{j\ge 1}$ and the number of the signals q are unknown and the signals $(\phi_j)_{j\ge 1}$ are known orthonormal functions, i.e.

$$\int_0^1 \phi_i(t)\,\phi_j(t)\,\mathrm{d}t = \mathbf{1}_{\{i\ne j\}}\,.$$

The problem is to estimate q, when the signal noise ratio goes to infinity. To describe this problem in the framework of the mathematical model we use the following stochastic differential equation

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \phi_j(t)\right) \mathrm{d}t + \varepsilon \mathrm{d}w_t\,, \tag{1.24}$$

where $(w_t)_{t\ge 0}$ is the standard Brownian motion and the parameter ε goes to zero. This means that the ratio signal/noise goes to infinity. The logarithm of the likelihood ratio for the model (1.24) can be represented as

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

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

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

Therefore, the maximum likelihood estimate for $\widehat{q} = q^*$. So, if $q^* = \infty$ we obtain that $\widehat{q} = \infty$. So, this estimator does not work. For this reason we propose to study the estimation problem

for q for the process (1.24) in the nonparametric setting and to apply the model selection procedure (1.21).

Chapter 2

Non-parametric estimation for semi-Markov regression models based on continuous data

2.1 Introduction

Let us consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{2.1}$$

where S is an unknown 1-periodic function from $\mathbf{L}_2[0,1]$ defined on $\mathbb{R}$ with values in $\mathbb{R}$, the noise process $(\xi_t)_{t\ge 0}$ is defined as

$$\xi_t = \varrho_1 L_t + \varrho_2 z_t\,, \tag{2.2}$$

where ϱ_1 and ϱ_2 are unknown coefficients, the pure jump process $(z_t)_{t\ge 1}$ is the semi-Markov process defined in (1.6) and $(L_t)_{t\ge 0}$ is the Lévy process defined in (1.12), for which we assume that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^8) < \infty\,. \tag{2.3}$$

The problem is to estimate the unknown function S in the model (2.1) on the basis of observations $(y_t)_{0\le t\le n}$. Firstly, this problem was considered in the framework of the "signal+white noise" models (see, for example, [9] or [47]). Later, in order to study dependent observations in continuous time, were introduced "signal+color noise" regressions based on Ornstein-Uhlenbeck processes (cf. [11], [12], [13], [16]).

Moreover, to include jumps in such models, the papers [17] and [18] used non Gaussian Ornstein-Uhlenbeck processes introduced in [4] for modeling of the risky assets in the stochastic volatility financial markets. Unfortunately, the dependence of the stable Ornstein-Uhlenbeck type decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain very quickly the same "signal+white noise" model.

The main goal of this chapter is to consider continuous time regression models with dependent observations for which the dependence does not disappear for a sufficient large duration of observations. To this end, we define the noise in the model (2.1) through a semi-Markov process which keeps the dependence for any duration n. This type of models allows, for example, to estimate the signals observed under long impulse noise impact with a memory or "against signals".

In this chapter we use the robust estimation approach introduced in [17] for such problems. To this end, we denote by Q the distribution of $(\xi_t)_{0\le t\le n}$ in the Skorokhod space $\mathcal{D}[0,n]$. We assume that Q is unknown and belongs to some distribution family $\mathcal{Q}_n$ specified in Section 2.2. In this chapter we use the quadratic risk

$$\mathcal{R}_Q(\widetilde{S}_n, S) = \mathbf{E}_{Q,S}\,\|\widetilde{S}_n - S\|^2\,, \tag{2.4}$$

where $\|f\|^2 = \int_0^1 f^2(s)\mathrm{d}s$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (2.1) corresponding to the noise distribution Q. Since the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

$$\mathcal{R}^*_n(\widetilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widetilde{S}_n, S)\,, \tag{2.5}$$

which enables us to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate $\widetilde{S}_n$ for all distributions in the family $\mathcal{Q}_n$.

To summarize, the goal of this chapter is to develop robust efficient model selection methods for the model (2.1) with the semi-Markov noise having unknown distribution, based on the approach proposed by Konev and Pergamenshchikov in [17] and [18] for continuous time regression models with semi-martingale noises. Unfortunately, we cannot use directly this method for semi-Markov regression models, since their tool essentially uses the fact that the Ornstein-Uhlenbeck dependence decreases with geometrical rate and the "white noise" case is obtained sufficiently quickly.

Thus in the thesis we propose new analytical tools based on renewal methods to obtain the sharp non-asymptotic oracle inequalities. As a consequence, we obtain the robust efficiency for the proposed model selection procedures in the adaptive setting.

2.2 Model selection

Note that for any function f from $\mathbf{L}_2[0,n]$, $f:[0,n]\to\mathbb{R}$, for the noise process $(\xi_t)_{t\ge 0}$ defined in (2.2), with $(z_t)_{t\ge 0}$ given in (1.6), the integral

$$I_n(f) = \int_0^n f(s)\mathrm{d}\xi_s \tag{2.6}$$

is well defined with $\mathbf{E}_Q\, I_n(f) = 0$. Moreover, as it is shown in Lemma 2.2,

$$\mathbf{E}_Q\, I_n^2(f) \le \varkappa_Q \int_0^n f^2(s)\mathrm{d}\,s\,, \tag{2.7}$$

where $\varkappa_Q = \varrho_1^2 + \varrho_2^2\,|\rho|_*$ and $|\rho|_* = \sup_{t\ge 0}|\rho(t)| < \infty$. Let us define the family of the noise distributions for the model (2.1) which is used in the robust risk (2.5). Note that any distribution Q from $\mathcal{Q}_n$ is defined by the unknown parameters in (2.2) and (1.12). We assume that

$$\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau} \le \varsigma^*, \tag{2.8}$$

where the unknown bounds ς^* are functions of n, i.e. $\varsigma^* = \varsigma^*(n)$, such that for any $\check{\epsilon} > 0$,

$$\lim_{n\to\infty} n^{\check{\epsilon}}\,\varsigma^*(n) = +\infty \quad \text{and} \quad \lim_{n\to\infty} \frac{\varsigma^*(n)}{n^{\check{\epsilon}}} = 0\,. \tag{2.9}$$

Remark 2.1. As we will see later, the parameter σ_Q is the limit of the Fourier transform of the noise process (2.2). Such limit is called variance proxy (see [17]).

Remark 2.2. Note that, generally (but it is not necessary) the parameters ϱ_1 and ϱ_2 can be dependent on n. The conditions (2.9) mean that we consider all possible cases, i.e. these parameters may go to infinity or be constant or go to zero as well. See, for example, the conditions (3.32) in [18].

Now, let $(\phi_j)_{j\geq 1}$ be an orthonormal uniformly bounded basis in $\mathbf{L}_2[0,1]$, i.e., for some constant $\phi_* \geq 1$, which may depend on n,

$$\sup_{0\leq j\leq n}\ \sup_{0\leq t\leq 1} |\phi_j(t)| \,\leq\, \phi_* < \infty\,. \tag{2.10}$$

We extend the functions $\phi_j(t)$ by periodicity, i.e., we set $\phi_j(t) := \phi_j(\{t\})$, where $\{t\}$ is the fractional part of $t \geq 0$. For example, we can take the trigonometric basis defined in (1.15). To estimate the function S we use here the model selection procedure for continuous time regression models from [17], based on the Fourrier expansion. We recall that, for any function S from $\mathbf{L}_2[0,1]$, we can write

$$S(t) = \sum_{j=1}^{\infty} \theta_j\,\phi_j(t) \quad \text{and} \quad \theta_j = (S,\phi_j) = \int_0^1 S(t)\phi_j(t)\mathrm{d}t\,. \tag{2.11}$$

So, to estimate the function S it suffices to estimate the coefficients θ_j and to replace them in this representation by their estimators. Using the fact that the function S and ϕ_j are 1 - periodic, we can write that

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

If we replace here the differential $S(t)\mathrm{d}t$ by the stochastic observed differential $\mathrm{d}y_t$ we obtain the natural estimate for θ_j on the time interval $[0,n]$

$$\widehat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\mathrm{d}\,y_t\,, \tag{2.12}$$

which can be represented, in view of the model (2.1), as

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j)\,. \tag{2.13}$$

Now (see, for example, [9]) we can estimate the function S by the projection estimators, i.e.

$$\widehat{S}_m(t) = \sum_{j=1}^{m} \widehat{\theta}_{j,n}\,\phi_j(t)\,, \quad 0\leq t\leq 1\,, \tag{2.14}$$

for some number $m \to \infty$ as $n \to \infty$. It should be noted that Pinsker in [47] shows that the projection estimators of the form (2.14) are not efficient. For obtaining efficient estimation one needs to use weighted least square estimators defined as

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(t)\,, \tag{2.15}$$

where the coefficients $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$. As it is shown in [47], in order to obtain efficient estimators, the coefficients $\lambda(j)$ in (2.15) need to be chosen depending on the regularity of the unknown function S. In this thesis we consider the adaptive case, i.e. we assume that the regularity of the function S is unknown. In this case we chose the weight coefficients on the basis of the model selection procedure proposed in [17] for the general semi-martingale regression model in continuous time. These coefficients will be obtained later in (2.28). To this end, first we set

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = 1 + \max_{\lambda\in\Lambda} \check{L}(\lambda)\,, \tag{2.16}$$

where $\#(\Lambda)$ is the cardinal number of Λ and $\check{L}(\lambda) = \sum_{j=1}^{n}\lambda(j)$. Now, to choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

which in our case is equal to

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n}\lambda(j)\widehat{\theta}_{j,n}\theta_j + \sum_{j=1}^{\infty}\theta_j^2. \tag{2.17}$$

Since the Fourier coefficients $(\theta_j)_{j\ge1}$ are unknown, we replace the terms $\widehat{\theta}_{j,n}\theta_{j,n}$ by

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{\widehat{\sigma}_n}{n}\,, \tag{2.18}$$

where $\widehat{\sigma}_n$ is an estimate for the variance proxy σ_Q defined in (2.8). If it is known, we take $\widehat{\sigma}_n = \sigma_Q$; otherwise, we can choose it, for example, as in [17], i.e.

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \widehat{\mathbf{T}}^2_{j,n}\,, \tag{2.19}$$

where $\widehat{\mathbf{T}}_{j,n}$ are the estimators for the Fourier coefficients $(\mathbf{T}_j)_{j\ge1}$ with respect to the trigonometric basis (1.15), i.e.

$$\widehat{\mathbf{T}}_{j,n} = \frac{1}{n}\int_0^n Tr_j(t)\mathrm{d}y_t\,. \quad \text{and} \quad \mathbf{T}_j = \int_0^1 Tr_j(t)\,S(t)\mathrm{d}t\,. \tag{2.20}$$

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

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda), \tag{2.21}$$

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

$$P_n(\lambda) = \frac{\widehat{\sigma}_n |\lambda|^2}{n}. \tag{2.22}$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,, \tag{2.23}$$

where

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_n(\lambda). \tag{2.24}$$

We recall that the set Λ is finite so $\widehat{\lambda}$ exists. In the case when $\widehat{\lambda}$ is not unique, we take one of them. Let us now specify the weight coefficients $(\lambda(j))_{1\le j\le n}$. Consider, for some fixed $0 < \varepsilon < 1$, a numerical grid of the form

$$\mathcal{A} = \{1, \dots, k^*\} \times \{\varepsilon, \dots, m\varepsilon\}\,, \tag{2.25}$$

where $m = [1/\varepsilon^2]$. We assume that both parameters $k^* \ge 1$ and ε are functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0\,, \\ \\ \lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\check{\delta}}\varepsilon(n) = +\infty, \end{cases} \tag{2.26}$$

for any $\check{\delta} > 0$. One can take, for example, for $n \ge 2$,

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k^*_0 + \sqrt{\ln n}\,, \tag{2.27}$$

where $k^*_0 \ge 0$ is some fixed constant and the threshold $\varsigma^*(n)$ is introduced in (2.8). For each $\alpha = (\beta, l) \in \mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha = (\lambda_\alpha(j))_{1\le j\le n}$$

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\le j<j_*\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}, \tag{2.28}$$

where $j_* = 1 + [\ln \upsilon_n]$, $\omega_\alpha = (\mathrm{d}_\beta\, l\upsilon_n)^{1/(2\beta+1)}$,

$$\mathrm{d}_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad \upsilon_n = n/\varsigma^*. \tag{2.29}$$

Now we define the set Λ as

$$\Lambda \,=\, \{\lambda_{\alpha}\,,\, \alpha \in \mathcal{A}\}\,. \tag{2.30}$$

It will be noted that in this case the cardinal of the set Λ is

$$\check{\iota} = k^{*}m\,. \tag{2.31}$$

Moreover, taking into account that $\mathrm{d}_{\beta} < 1$ for $\beta \geq 1$ we obtain for the set (2.30)

$$|\Lambda|_{*} \,\leq\, 1 + \sup_{\alpha\in\mathcal{A}} \omega_{\alpha} \leq 1 + (\upsilon_{n}/\varepsilon)^{1/3}\,. \tag{2.32}$$

Remark 2.3. Note that the form (2.28) for the weight coefficients in (2.15) was proposed by Pinsker in [47] for the efficient estimation in the nonadaptive case, i.e. when the regularity parameters of the function S are known. In the adaptive case these weight coefficients are used in [17, 18] to show the asymptotic efficiency for model selection procedures.

2.3 Oracle inequality

In this section we obtain in Theorem 2.2 the non-asymptotic oracle inequality for the quadratic risk (2.4) for the model selection procedure (2.23) and in Theorem 2.3 the non-asymptotic oracle inequality for the robust risk (2.5) for the same model selection procedure (2.23), considered with the coefficients (2.28).

In order to prove the oracle inequality, the following conditions will be needed for the noise $(\xi_t)_{t\geq 0}$. Here we use the conditions introduced in [17] for the general semi-martingale model (2.1).

$\mathbf{C}_1$) *For all $n \geq 1$ and Q there exist a variance proxy $\sigma_Q > 0$ and the constant $\mathbf{C}_{1,Q,n} \geq 0$ such that, for any basis functions with the bound* (2.10),

$$\sup_{x\in[-1,1]^n} \left|B_{1,Q,n}(x)\right| \,\leq\, \mathbf{C}_{1,Q,n} < \infty\,,$$

where $B_{1,Q,n}(x) = \sum_{j=1}^{n} x_j \,\Big(\mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q\Big)$.

$\mathbf{C}_2$) *For all $n \geq 1$ and Q there exists a constant $\mathbf{C}_{2,Q,n} \geq 1$ such that, for any basis functions with the bound* (2.10),

$$\sup_{|x|\leq 1} \mathbf{E}_Q\, B_{2,Q,n}^2(x) \,\leq\, \mathbf{C}_{2,Q,n} < \infty,$$

where $|x|^2 = \sum_{j=1}^{n} x_j^2$ *and* $B_{2,Q,n}(x) = \sum_{j=1}^{n} x_j \,\Big(\xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2\Big)$.

Before stating the non-asymptotic oracle inequality, let us first introduce the following parameters which will be used for describing the rest term in the oracle inequalities. For the

renewal density ρ defined in (1.8) we set

$$\Upsilon(x) = \rho(x) - \frac{1}{\check{\tau}} \quad \text{and} \quad \|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)|\,\mathrm{d}x\,, \tag{2.33}$$

where $\check{\tau} = \mathbf{E}\,\tau_1$. In Proposition 5.1 we show that $|\rho|_* = \sup_{t\geq 0}|\rho(t)| < \infty$ and $\|\Upsilon\|_1 < \infty$. So, using this, we can introduce the following parameters

$$\Psi_Q = 4\varkappa_Q \check{\iota} + \left(5 + \frac{4\check{\iota}}{\sigma_Q}\right)\left(\sigma_Q\,\check{\tau}\,\phi^2_{max}\,\|\Upsilon\|_1 + \phi^4_{max}(1+\sigma_Q^2)^3\,\check{l}\right) \tag{2.34}$$

and

$$\mathbf{c}^*_Q = \sigma_Q + 2\varkappa_Q + \sigma_Q\,\check{\tau}\,\phi^2_{max}\,\|\Upsilon\|_1 + \phi^4_{max}(1+\sigma_Q^2)^2\,\check{l}\,, \tag{2.35}$$

where $\check{l} = (4\check{\tau}^2+8)\,\|\Upsilon\|_1 + 5 + 13(1+\check{\tau})^2(1+|\rho|^2_*)(\mathbf{E}Y_1^4) + 4\Pi(x^4)$. First, let us state the non-asymptotic oracle inequality for the quadratic risk (2.4) for the model selection procedure (2.23).

Theorem 2.1. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold. Then, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, the estimator of* S *given in* (2.23) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{\Psi_Q + 10|\Lambda|_*\,\mathbf{E}_S|\widehat{\sigma}_n - \sigma_Q|}{n\delta}\,. \tag{2.36}$$

Proof. First, note that we can rewrite the empirical squared error in (2.17) as follows

$$\mathrm{Err}_n(\lambda) = J_n(\lambda) + 2\sum_{j=1}^{\infty}\lambda(j)\check{\theta}_{j,n} + ||S||^2 - \delta P_n(\lambda), \tag{2.37}$$

where $\check{\theta}_{j,n} = \widetilde{\theta}_{j,n} - \theta_j\widehat{\theta}_{j,n}$. Using the definition of $\widetilde{\theta}_{j,n}$ in (2.18) we obtain that

$$\check{\theta}_{j,n} = \frac{1}{\sqrt{n}}\theta_j\xi_{j,n} + \frac{1}{n}\widetilde{\xi}_{j,n} + \frac{1}{n}\varsigma_{j,n} + \frac{\sigma_Q - \widehat{\sigma}_n}{n}\,,$$

where $\varsigma_{j,n} = \mathbf{E}_Q\xi^2_{j,n} - \sigma_Q$ and $\widetilde{\xi}_{j,n} = \xi^2_{j,n} - \mathbf{E}_Q\xi^2_{j,n}$. Putting

$$M(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\lambda(j)\theta_j\xi_{j,n} \quad \text{and} \quad P^0_n = \frac{\sigma_Q|\lambda|^2}{n}\,, \tag{2.38}$$

we can rewrite (2.37) as

$$\begin{aligned}\mathrm{Err}_n(\lambda) =& J_n(\lambda) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n}\,\check{L}(\lambda) + 2M(\lambda) + \frac{2}{n}B_{1,Q,n}(\lambda)\\ &+ 2\sqrt{P^0_n(\lambda)}\frac{B_{2,Q,n}(e(\lambda)}{\sqrt{\sigma_Q n}} + \|S\|^2 - \rho P_n(\lambda),\end{aligned} \tag{2.39}$$

where $e(\lambda) = \lambda/|\lambda|$, the function $\check{L}(\cdot)$ is defined in (2.16) and the functions $B_{1,Q,n}(\cdot)$ and $B_{2,Q,n}(\cdot)$ are given in conditions $\mathbf{C}_1$) and $\mathbf{C}_2$).

Let $\lambda_0 = (\lambda_0(j))_{1\le j\le n}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be as in (2.24). Substituting λ_0 and $\widehat{\lambda}$ in Equation (2.39), we obtain

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) - \mathrm{Err}_n(\lambda_0) =& J(\widehat{\lambda}) - J(\lambda_0) + 2\frac{\sigma_Q - \widehat{\sigma}_Q}{n}\,\check{L}(\varpi) + \frac{2}{n}B_{1,Q,n}(\varpi) + 2M(\varpi)\\ &+ 2\sqrt{P_n^0(\widehat{\lambda})}\frac{B_{2,Q,n}(\widehat{e})}{\sqrt{\sigma_Q n}} - 2\sqrt{P_n^0(\lambda_0)}\frac{B_{2,Q,n}(e_0)}{\sqrt{\sigma_Q n}}\\ &- \delta P_n(\widehat{\lambda}) + \delta P_n(\lambda_0), \end{aligned} \tag{2.40}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{e} = e(\widehat{\lambda})$ and $e_0 = e(\lambda_0)$. Note that, by (2.16),

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

Applying the inequality

$$2|ab| \le \delta a^2 + \delta^{-1}b^2 \tag{2.41}$$

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

$$2\sqrt{P_n^0(\lambda)}\frac{|B_{2,Q,n}(e(\lambda))|}{\sqrt{\sigma_Q n}} \le \delta P_n^0(\lambda) + \frac{B_{2,Q,n}^2(e(\lambda))}{\delta\sigma_Q\, n}.$$

Taking into account the bound (2.59), we get

$$\begin{aligned}\mathrm{Err}_n(\hat{\lambda}) \le& \mathrm{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta\sigma_Q\, n}\\ &+ \frac{1}{n}|\widehat{\sigma} - \sigma_Q|(|\widehat{\lambda}|^2 + |\lambda_0|^2) + 2\delta P_n(\lambda_0)\,,\end{aligned}$$

where $B_{2,Q,n}^* = \sup_{\lambda\in\Lambda} B_{2,Q,n}^2((e(\lambda))$. Moreover, noting that in view of (2.16) $\sup_{\lambda\in\Lambda}|\lambda|^2 \le |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) \le& \mathrm{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta\sigma_Q n}\\ &+ \frac{4|\Lambda|_*}{n}|\widehat{\sigma} - \sigma_Q| + 2\delta P_n(\lambda_0).\end{aligned} \tag{2.42}$$

To estimate the second term in the right side of this inequality we set

$$S_x = \sum_{j=1}^{n} x(j)\theta_j\phi_j\,, \quad x = (x(j))_{1\le j\le n} \in \mathbb{R}^n\,.$$

Thanks to (2.7) we estimate the term $M(x)$ for any $x \in \mathbb{R}^n$ as

$$\mathbf{E}_Q M^2(x) \le \varkappa_Q \frac{1}{n}\sum_{j=1}^{n} x^2(j)\theta_j^2 = \varkappa_Q \frac{1}{n}||S_x||^2. \tag{2.43}$$

To estimate this function for a random vector $x \in \mathbb{R}^n$ we set

$$Z^* = \sup_{x \varepsilon \Lambda_1} \frac{nM^2(x)}{||S_x||^2}, \quad \Lambda_1 = \Lambda - \lambda_0 .$$

So, through Inequality (2.41), we get

$$2|M(x)| \le \delta ||S_x||^2 + \frac{Z^*}{n\delta}. \tag{2.44}$$

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

$$\mathbf{E}_Q Z^* \le \sum_{x \in \Lambda_1} \frac{n\mathbf{E}_Q M^2(x)}{||S_x||^2} \le \sum_{x \in \Lambda_1} \varkappa_Q = \varkappa_Q \check{\iota} , \tag{2.45}$$

where $\check{\iota} = \mathrm{card}(\Lambda)$. Moreover, note that, for any $x \in \Lambda_1$,

$$||S_x||^2 - ||\widehat{S}_x||^2 = \sum_{j=1}^{n} x^2(j)(\theta_j^2 - \widehat{\theta}_j^2) \le -2M_1(x), \tag{2.46}$$

where $M_1(x) = n^{-1/2} \sum_{j=1}^{n} x^2(j)\theta_j \xi_{j,n}$. Taking into account that, for any $x \in \Lambda_1$ the components $|x(j)| \le 1$, we can estimate this term as in (2.43), i.e.,

$$\mathbf{E}_Q M_1^2(x) \le \varkappa_Q \frac{||S_x||^2}{n} .$$

Similarly to the previous reasoning we set

$$Z_1^* = \sup_{x \varepsilon \Lambda_1} \frac{nM_1^2(x)}{||S_x||^2}$$

and we get

$$\mathbf{E}_Q Z_1^* \le \varkappa_Q \check{\iota} . \tag{2.47}$$

Using the same type of arguments as in (2.44), we can derive

$$2|M_1(x)| \le \delta ||S_x||^2 + \frac{Z_1^*}{n\delta}. \tag{2.48}$$

From here and (2.46), we get

$$||S_x||^2 \le \frac{||\widehat{S}_x||^2}{1-\delta} + \frac{Z_1^*}{n\delta(1-\delta)} \tag{2.49}$$

for any $0 < \delta < 1$. Using this bound in (2.44) yields

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

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

$$2M(\varpi) \le \frac{2\delta(\mathrm{Err}_n(\widehat{\lambda}) + \mathrm{Err}_n(\lambda_0))}{1-\delta} + \frac{Z^* + Z_1^*}{n\delta(1-\delta)}.$$

Using this bound in (2.42) we obtain

$$\mathrm{Err}_n(\widehat{\lambda}) \leq \frac{1+\delta}{1-3\delta}\mathrm{Err}_n(\lambda_0) + \frac{Z^* + Z_1^*}{n\delta(1-3\delta)} + \frac{2\mathbf{C}_{1,Q,n}}{n(1-3\delta)} + \frac{2B^*_{2,Q,n}}{\delta(1-3\delta)\sigma_Q n}$$
$$+ \frac{(4|\Lambda|_* + 2)}{n(1-3\delta)}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).$$

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

$$\mathrm{Err}_n(\widehat{\lambda}) \leq \frac{1+\delta}{1-3\delta}\mathrm{Err}_n(\lambda_0) + \frac{2(Z^* + Z_1^*)}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4B^*_{2,Q,n}}{\delta\sigma_Q n}$$
$$+ \frac{(8|\Lambda|_* + 2)}{n}|\widehat{\sigma}_n - \sigma_Q| + \frac{2\delta}{(1-3\delta)}\, P_n^0(\lambda_0).$$

In view of Proposition 2.3 we estimate the expectation of the term $B^*_{2,Q,n}$ in (2.42) as

$$\mathbf{E}_Q\, B^*_{2,Q,n} \leq \sum_{\lambda\in\Lambda} \mathbf{E}_Q B^2_{2,Q,n}(e(\lambda)) \leq \check{\iota}\mathbf{C}_{2,Q,n}\,.$$

Taking into account that $|\Lambda|_* \geq 1$, we get

$$\mathcal{R}(\widehat{S}_*, S) \leq \frac{1+\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\iota}}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4\check{\iota}\mathbf{C}_{2,Q,n}}{\delta\sigma_Q n}$$
$$+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).$$

Using the upper bound for $P_n(\lambda_0)$ in Lemma 2.6, one obtains (2.36), that finishes the proof. □

Now we study the estimate (2.19).

Proposition 2.1. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold and that the function S is continuously differentiable. Then, for any $n \geq 2$,*

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \sigma_Q| \leq \frac{6\|\dot{S}\|^2 + \mathbf{c}_Q^*}{\sqrt{n}}\,. \tag{2.50}$$

Proof. We use here the same method as in [14]. First of all note that Definition (2.20) implies that

$$\widehat{\mathbf{T}}_{j,n} = \mathbf{T}_j + \frac{1}{\sqrt{n}}\,\eta_{j,n}\,, \tag{2.51}$$

where

$$\mathbf{T}_j = \int_0^1 S(t)\, Tr_j(t)\mathrm{d}t \quad \text{and} \quad \eta_{j,n} = \frac{1}{\sqrt{n}}\int_0^n \mathrm{Tr}_j(t)\,\mathrm{d}\xi_t\,.$$

So, we have

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j^2 + 2M_n + \frac{1}{n}\sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2\,, \tag{2.52}$$

where

$$M_n = \frac{1}{\sqrt{n}} \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j \, \eta_{j,n} \, .$$

Note that, for continuously differentiable functions (see, for example, Lemma A.6 in [14]), the Fourier coefficients $(\mathbf{T}_j)_{\{j\geq 1\}}$ satisfy the following inequality, for any $n \geq 1$,

$$\sum_{j=[\sqrt{n}]+1}^{\infty} \mathbf{T}_j^2 \leq \frac{4\left(\int_0^1 |\dot{S}(t)|\mathrm{d}t\right)^2}{\sqrt{n}} \leq \frac{4\|\dot{S}\|^2}{\sqrt{n}} \, . \tag{2.53}$$

In the same way as in (2.43) we estimate the term M_n, i.e.,

$$\mathbf{E}_Q \, M_n^2 \leq \frac{\varkappa_Q}{n} \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j^2 \leq \frac{4\varkappa_Q \|\dot{S}\|^2}{n\sqrt{n}},$$

while the absolute value of this term for $n \geq 1$ can be estimated as

$$|\mathbf{E}_Q \, M_n| \leq \frac{\varkappa_Q + \|\dot{S}\|^2}{\sqrt{n}} \, .$$

Moreover, using Propositions 2.2 and 2.3 we can represent the last term in (2.52) as

$$\frac{1}{n} \sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2 = \frac{\sigma_Q(n-\sqrt{n})}{n} + \frac{B_{1,Q,n}(x')}{n} + \frac{B_{2,Q,n}(x'')}{\sqrt{n}},$$

with $x'_j = \mathbf{1}_{\{\sqrt{n}<j\leq n\}}$ and $x''_j = \mathbf{1}_{\{\sqrt{n}<j\leq n\}}/\sqrt{n}$. Therefore,

$$\mathbf{E}_Q \left| \frac{1}{n} \sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2 - \sigma_Q \right| \leq \frac{\sigma_Q}{\sqrt{n}} + \frac{\mathbf{C}_{1,Q,n}}{n} + \frac{\sqrt{\mathbf{C}_{2,Q,n}}}{\sqrt{n}}.$$

Taking into account that $\mathbf{C}_{2,Q,n} \geq 1$, we obtain the bound (2.50) and hence the desired result. □

Theorem 2.1 and Proposition 2.1 implies the following result.

Theorem 2.2. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold and that the function* S *is continuously differentiable. Then, for any* $n \geq 1$ *and* $0 < \delta \leq 1/6$*, the procedure* (2.23), (2.19) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{60\widetilde{\Lambda}_n \, \|\dot{S}\|^2 + \widetilde{\Psi}_{Q,n}}{n\delta} \, , \tag{2.54}$$

where $\widetilde{\Psi}_{Q,n} = 10\widetilde{\Lambda}_n \mathbf{c}_Q^* + \Psi_Q$ *and* $\widetilde{\Lambda}_n = |\Lambda|_*/\sqrt{n}$.

Remark 2.4. Note that the coefficient $\varkappa_Q$ can be estimated as $\varkappa_Q \le (1+\check{\tau}|\rho|_*)\sigma_Q$. Therefore, taking into account that $\phi^4_{max} \ge 1$, the remainder term in (2.54) can be estimated as

$$\widetilde{\Psi}_{Q,n} \le \mathbf{C}_* \left(1+\sigma_Q^6+\frac{1}{\sigma_Q}\right)(1+\widetilde{\Lambda}_n)\check{\iota}\phi^4_{max}\,, \tag{2.55}$$

where $\mathbf{C}_* > 0$ is some constant which is independent of the distribution Q.

Furthermore, let us study the robust risk (2.5) for the procedure (2.23). In this case, the distribution family $\mathcal{Q}_n$ consists in all distributions on the Skorokhod space $\mathcal{D}[0,n]$ of the process (2.2) with the parameters satisfying the conditions (2.8) and (2.9).

Moreover, we assume also that the number of the weight vectors and the upper bound for the basis functions in (2.10) may depend on $n \ge 1$, i.e. $\check{\iota} = \check{\iota}(n)$ and $\phi_* = \phi_*(n)$, such that for any $\check{\epsilon} > 0$

$$\lim_{n\to\infty} \frac{\check{\iota}(n)}{n^{\check{\epsilon}}} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\phi_*(n)}{n^{\check{\epsilon}}} = 0\,. \tag{2.56}$$

The next result presents the non-asymptotic oracle inequality for the robust risk (2.5) for the model selection procedure (2.23), considered with the coefficients (2.28).

Theorem 2.3. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_4$) *hold and that the unknown function* S *is continuously differentiable. Then, for the robust risk defined in* (2.5) *through the distribution family* (2.8) – (2.9)*, the procedure* (2.23) *with the coefficients* (2.28) *for any* $n \ge 1$ *and* $0 < \delta < 1/6$*, satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{U}^*_n(S)}{n\delta}, \tag{2.57}$$

where the sequence $\mathbf{U}^*_n(S) > 0$ *is such that, under the conditions* (2.9), (2.26) *and* (2.56)*, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\le r} \frac{\mathbf{U}^*_n(S)}{n^{\check{\delta}}} = 0. \tag{2.58}$$

Proof. First note, that in view of (2.31) and (2.26)

$$\lim_{n\to\infty} \frac{\check{\iota}}{n^{\check{\epsilon}}} = \lim_{n\to\infty} \frac{k^* m}{n^{\check{\epsilon}}} = 0 \qquad \text{for any} \quad \check{\epsilon} > 0\,.$$

Furthermore, the bound (2.32) and the conditions (2.9) and (2.26) yield

$$\lim_{n\to\infty} \frac{|\Lambda|_*}{n^{1/3+\check{\epsilon}}} = 0 \quad \text{for any} \quad \check{\epsilon} > 0\,.$$

So, from here we obtain the convergence (2.58). □

Now we need to check the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) for the process (2.2)

Proposition 2.2. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold. Then Condition* $\mathbf{C}_1$) *holds with*

$$\mathbf{C}_{1,Q,n} = \sigma_Q\, \check{\tau}\, \phi^2_{max}\, \|\Upsilon\|_1. \tag{2.59}$$

Proof. First, note that from (2.93) we have

$$\xi_{j,n} = \frac{\varrho_1}{\sqrt{n}} I_n^L(\phi_j) + \frac{\varrho_2}{\sqrt{n}} I_n^z(\phi_j)\,.$$

So, using (2.95) we can write that

$$\mathbf{E}\xi_{j,n}^2 = \frac{\varrho_1^2}{n}\int_0^n \phi_j^2(t)\mathrm{d}\,t + \frac{\varrho_2^2}{n}\mathbf{E}\sum_{l=1}^{\infty}\phi_j^2(T_l)1_{\{T_l\le n\}}\,. \tag{2.60}$$

Proposition 5.1 implies

$$\begin{aligned}\mathbf{E}\sum_{l=1}^{\infty}\phi_j^2(T_l)1_{\{T_l\le n\}} &= \int_0^n \phi_j^2(x)\,\rho(x)\mathrm{d}x \\ &= \frac{1}{\breve{\tau}}\int_0^n \phi_j^2(x)\mathrm{d}\,x + \int_0^n \phi_j^2(x)\Upsilon(x)\mathrm{d}\,x\,.\end{aligned}$$

Note that $\int_0^n \phi_j^2(t)\mathrm{d}\,t = n$. So, in view of the condition (2.10), we obtain

$$\left|\mathbf{E}\xi_{j,n}^2 - \sigma_Q\right| = \frac{\varrho_2^2}{n}\left|\int_0^n \phi_j^2(x)\Upsilon(x)\mathrm{d}\,x\right| \le \frac{\varrho_2^2}{n}\,\phi_{max}^2\,\|\Upsilon\|_1\,. \tag{2.61}$$

Estimating here ϱ_2^2 by $\sigma_Q\breve{\tau}$ we obtain the inequality (2.59) and hence the conclusion follows. □

Proposition 2.3. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold. Then Condition* $\mathbf{C}_2$) *holds with*

$$\mathbf{C}_{2,Q,n} = \phi_{max}^4(1+\sigma_Q^2)^3\,\breve{l} \tag{2.62}$$

and $\breve{l}$ *is given in* (2.35).

Proof. By Ito's formula one gets

$$\mathrm{d}I_t^2(f) = 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\breve{\varrho}^2\,f^2(t)\mathrm{d}\,t + \sum_{0\le s\le t} f^2(s)(\Delta\xi_s^d)^2\,, \tag{2.63}$$

where $\xi_t^d = \varrho_3\,\breve{L}_t + \varrho_2 z_t$ and $\varrho_3 = \varrho_1\sqrt{1-\breve{\varrho}^2}$. Taking into account that the processes $(\breve{L}_t)_{t\ge0}$ and $(z_t)_{t\ge0}$ are independent and the time of jumps T_k defined in (1.7) has a density, we have $\Delta z_s\Delta\breve{L}_s = 0$ a.s. for any $s\ge0$. Therefore, we can rewrite the differential (2.63) as

$$\begin{aligned}\mathrm{d}I_t^2(f) =& 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\breve{\varrho}^2\,f^2(t)\mathrm{d}\,t \\ &+ \varrho_3^2\mathrm{d}\sum_{0\le s\le t} f^2(s)(\Delta\breve{L}_s)^2 + \varrho_2^2\mathrm{d}\sum_{0\le s\le t} f^2(s)(\Delta z_s)^2\,.\end{aligned} \tag{2.64}$$

From Lemma 2.2 it follows that

$$\mathbf{E}I_t^2(f) = \varrho_1^2\int_0^t f^2(s)ds + \varrho_2^2\int_0^t f^2(s)\rho(s)\mathrm{d}s\,.$$

Therefore, putting

$$\widetilde{I}_t(f) = I_t^2(f) - \mathbf{E} I_t^2(f)\,, \tag{2.65}$$

we obtain

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

where $\check{m}_t = \sum_{0\le s\le t}(\Delta\check{L}_s)^2 - t$ and $m_t = \sum_{0\le s\le t}(\Delta z_s)^2 - \int_0^t \rho(s)\mathrm{d}s$. For any non-random vector $x = (x_j)_{1\le j\le n}$ with $\sum_{j=1}^n x_j^2 \le 1$, we set

$$\bar{I}_t(x) = \sum_{j=1}^n x_j \widetilde{I}_t(\phi_j). \tag{2.66}$$

Denoting

$$A_t(x) = \sum_{j=1}^n x_j I_t(\phi_j)\phi_j(t) \quad \text{and} \quad B_t(x) = \sum_{j=1}^n x_j \phi_j^2(t)\,, \tag{2.67}$$

we get the following stochastic differential equation for (2.66)

$$\mathrm{d}\bar{I}_t(x) = 2A_{t-}(x)\mathrm{d}\xi_t + B_t(x)\mathrm{d}\widetilde{m}_t\,, \quad \bar{I}_0(x) = 0\,.$$

Applying the Ito's formula one obtains

$$\begin{aligned}\mathbf{E}\,\bar{I}_n^2(x) =& 2\mathbf{E}\int_0^n \bar{I}_{t-}(x)\mathrm{d}\bar{I}_t(x) + 4\varrho_1^2\check{\varrho}^2\mathbf{E}\int_0^n A_t^2(x)\mathrm{d}\,t \\ &+ \varrho_3^2\ \mathbf{E}\,\check{D}_n(x) + \varrho_2^2\,\mathbf{E}\,D_n(x)\,,\end{aligned} \tag{2.68}$$

where $\check{D}_n(x) = \sum_{0\le t\le n}\left(2A_{t-}(x)\Delta\check{L}_t + \varrho_3^2 B_t(x)(\Delta\check{L}_t)^2\right)^2$ and
$D_n(x) = \sum_{k=1}^{+\infty}\left(2A_{T_k-}(x)Y_k + \varrho_2 B_{T_k-}(x)Y_k^2\right)^2\ \mathbf{1}_{\{T_k\le n\}}$. Let us now show that

$$\left|\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x)\right| \le 2\,\varrho_2^4\phi_{max}^3\,\|\Upsilon\|_1\,n^2\,. \tag{2.69}$$

To this end, note that

$$\begin{aligned}\int_0^n \bar{I}_{t-}(x)\mathrm{d}\bar{I}_t(x) =& 2\sum_{1\le j,l\le n} x_j x_l \int_0^n \widetilde{I}_{t-}(\phi_j)\, I_{t-}(\phi_l)\phi_l(t)\mathrm{d}\xi_t \\ &+ \sum_{j=1}^n x_j \int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}\widetilde{m}_t\,.\end{aligned}$$

Using here Lemma 2.5, we get $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\, I_{t-}(\phi_i)\phi_i(t)\mathrm{d}\xi_t = 0$. Moreover, the process $(\check{m}_t)_{t\ge 0}$ is a martingale, i.e. $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}\widetilde{m}_t = 0$. Therefore,

$$\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x) = \varrho_2^2\sum_{j=1}^n x_j\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}m_t\,.$$

Taking into account here that for any non-random bounded function f

$$\mathbf{E}\int_0^n f(t)\mathrm{d}m_t = 0,$$

we obtain $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\,B_t(x)\,\mathrm{d}m_t = \mathbf{E}\int_0^n I^2_{t-}(\phi_j)\,B_t(x)\,\mathrm{d}m_t$. So, Lemma 2.4 yields

$$\left|\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\,B_t(x)\mathrm{d}m_t\right| = \left|\sum_{l=1}^n x_l\mathbf{E}\int_0^n I^2_{t-}(\phi_j)\phi^2_l(t)\mathrm{d}m_t\right|$$

$$\le\ 2\,\varrho_2^2\phi^3_{max}\,\|\Upsilon\|_1\sum_{l=1}^n |x_l|n\,.$$

Therefore,

$$\left|\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x)\right| \le 2\,\varrho_2^4\phi^3_{max}\,\|\Upsilon\|_1\,n\sum_{1\le l,j\le n}|x_l|\,|x_j|$$

$$= 2\,\varrho_2^4\phi^3_{max}\,\|\Upsilon\|_1\,n\left(\sum_{l=1}^n |x_l|\right)^2.$$

Taking into account here that $\left(\sum_{l=1}^n |x_l|\right)^2 \le\ n\sum_{l\ge 1} x_l^2 \le n$, we obtain (2.69). Reminding that $\Pi(x^2)=1$ we can calculate directly that

$$\mathbf{E}\,\check{D}_n(x) = 4\,\mathbf{E}\int_0^n A^2_t(x)\mathrm{d}t + \varrho_3^4\,\Pi(x^4)\int_0^n B^2_t(x)\mathrm{d}t\,. \tag{2.70}$$

Note that, thanks to Lemma 2.1, we obtain that

$$\mathbf{E}\int_0^n A^2_t(x)\mathrm{d}\,t = \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\mathbf{E}I_t\phi_i(t)I_t\phi_j(t)\mathrm{d}\,t$$

$$= \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\int_0^t \phi_i(v)\phi_j(v)(\varrho_1^2+\varrho_2^2\rho(v))\mathrm{d}v$$

$$= \frac{1}{2}\varrho_1^2\sum_{i,j} x_ix_j\left(\int_0^n \phi_i(t)\phi_j(t)\mathrm{d}t\right)^2 + \varrho_2^2\,A_{1,n}(x)$$

$$\le \frac{n^2}{2}\varrho_1^2 + \varrho_2^2\,A_{1,n}(x)\,,$$

where $A_{1,n}(x) = \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\,\left(\int_0^t \phi_i(v)\phi_j(v)\,\rho(v)\mathrm{d}v\right)\mathrm{d}t$. This term can be estimated through Proposition 5.1 as

$$\left|A_{1,n}(x)\right| = \left|\frac{n^2}{2\check{\tau}} + \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\,\left(\int_0^t \phi_i(v)\phi_j(v)\,\Upsilon(v)\mathrm{d}v\right)\mathrm{d}t\right|$$

$$\le \frac{n^2}{2\check{\tau}} + n\,\phi^4_{max}\,\|\Upsilon\|_1\sum_{i,j}|x_i||x_j| \le \left(\frac{1}{2\check{\tau}} + \phi^4_{max}\,\|\Upsilon\|_1\right)n^2\,.$$

So, reminding that $\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau}$ and that $\phi_{max} \geq 1$, we obtain that

$$\mathbf{E}\int_0^n A_t^2(x)\mathrm{d}\,t \leq \left(\frac{\sigma_Q}{2} + \phi_{max}^4 \|\Upsilon\|_1\right) n^2 \leq \left(\frac{1}{4} + \|\Upsilon\|_1\right) \phi_{max}^4 \,(1+\sigma_Q^2)\, n^2\,. \tag{2.71}$$

Taking into account that

$$\sup_{t\geq 0} B_t^2(x) \;\leq\; \phi_{max}^4 \left(\sum_{j=1}^n |x_j|\right)^2 \;\leq\; \phi_{max}^4\, n\,, \tag{2.72}$$

that $\phi_{max} \geq 1$ and that $\varrho_1^4 \leq \sigma_Q^2$, we estimate the expectation in (2.70) as

$$\mathbf{E}\check{D}_n \leq 4\phi_{max}^4(1+\sigma_Q^2)\left(1 + \|\Upsilon\|_1 + \Pi(x^4)\right)\, n^2\,. \tag{2.73}$$

Moreover, taking into account that the random variable Y_k is independent of $A_{T_{k^-}}(x)$ and of the field $\mathcal{G} = \sigma\{T_j\,,\, j \geq 1\}$ and that $\mathbf{E}\left(A_{T_{k^-}}(x)\,|\mathcal{G}\right) = 0$, we get

$$\mathbf{E}\sum_{k=1}^{+\infty} B_{T_k-}(x)\, A_{T_{k^-}}(x) Y_k^3 1_{\{T_k\leq n\}} = \sum_{k=1}^{+\infty} \mathbf{E}\mathbf{E}\left(B_{T_k-}(x)\, A_{T_{k^-}}(x) Y_k^3 1_{\{T_k\leq n\}}|\mathcal{G}\right)$$
$$= \mathbf{E}\, Y_1^3\, \mathbf{E}\sum_{k=1}^{+\infty} B_{T_k-}(x) 1_{\{T_k\leq n\}}\, \mathbf{E}(A_{T_{k^-}}(x)|\mathcal{G}) = 0\,.$$

Therefore,

$$\mathbf{E}\, D_n(x) = \varrho_2^2 \mathbf{E} Y_1^4 D_{1,n}(x) + 4D_{2,n}(x), \tag{2.74}$$

where

$$D_{1,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E}\, B_{T_k-}^2(x) 1_{\{T_k\leq n\}} \quad \text{and} \quad D_{2,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E}\, A_{T_{k^-}}^2(x) \mathbf{1}_{\{T_k\leq n\}}\,.$$

Using the bound (2.72) we can estimate the term $D_{1,n}$ as $D_{1,n}(x) \leq \phi_{max}^4 n \mathbf{E}\, N_n$. Using here Corollary 5.1, we obtain

$$D_{1,n}(x) \leq\; |\rho|_* \phi_{max}^4\, n^2\,. \tag{2.75}$$

Now, to estimate the last term in (2.74), note that the process $A_t(x)$ can be rewritten as

$$A_t(x) = \int_0^t Q_x(t,s) d\xi_s, \quad \text{with} \quad Q_x(t,s) = \sum_{j=1}^n x_j \phi_j(s)\phi_j(t). \tag{2.76}$$

Applying Lemma 2.3 again, we obtain for any $k \geq 1$

$$\mathbf{E}\left(A_{T_{k^-}}^2(x)|\mathcal{G}\right) = \varrho_1^2 \int_0^{T_k} Q_x^2(T_k,s)\mathrm{d}s + \varrho_2^2 \sum_{j=1}^{k-1} Q_x^2(T_k,T_j)\,.$$

So, we can represent the last term in (2.74) as

$$D_{2,n} = \varrho_1^2\, D_{2,n}^{(1)} + \varrho_2^2\, D_{2,n}^{(2)}\,, \tag{2.77}$$

where

$$D_{2,n}^{(1)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \int_0^{T_k} Q_x^2(T_k,s)\mathrm{d}s$$

and

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \sum_{j=1}^{k-1} Q_x^2(T_k,T_j)\,.$$

Thanks to Proposition 5.1 we obtain

$$D_{2,n}^{(1)} = \int_0^n \left(\int_0^t Q_x^2(t,s)\mathrm{d}s\right) \rho(t)\,\mathrm{d}t \;\le\; |\rho|_* \int_0^n \int_0^n Q_x^2(t,s)\mathrm{d}s\,\mathrm{d}t\,.$$

In view of the definition of Q_x in (2.76), we can rewrite the last integral as

$$\begin{aligned}\int_0^n Q_x^2(t,s)\mathrm{d}s &= \sum_{1\le i,j\le n} x_i\, x_j\, \phi_i(t)\,\phi_j(t) \int_0^n \phi_i(s)\,\phi_j(s)\,\mathrm{d}s \\ &= \sum_{i=1}^n x_i^2\,\phi_i^2(t) \int_0^n \phi_i^2(s)\,\mathrm{d}s = n\sum_{i=1}^n x_i^2\,\phi_i^2(t)\,.\end{aligned}$$

Since $\sum_{j=1}^n x_j^2 \le 1$, we obtain that,

$$\int_0^n Q_x^2(t,s)\mathrm{d}s \;\le\; \phi_{max}^2\, n \quad\text{and}\quad D_{2,n}^{(1)} \le \phi_{max}^2\,|\rho|_*\, n^2\,. \tag{2.78}$$

Let us estimate now the last term in (2.77). First, note that we can represent this term as

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \sum_{j=1}^{k-1} Q_x^2(T_k,T_j) = \sum_{j=1}^{\infty} \mathbf{1}_{\{T_j\le n\}}\, G(T_j) = \int_0^n G(t)\,\rho(t)\mathrm{d}t\,,$$

where

$$\begin{aligned}G(t) &= \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}}\, Q_x^2((t+T_k),t) = \int_0^n Q_x^2(t+v,t)\,\rho(v)\mathrm{d}v \\ &= \int_t^{n+t} Q_x^2(u,t)\,\rho(u-t)\mathrm{d}u\,.\end{aligned}$$

It is clear that, for any $0\le t\le n$,

$$\int_t^{n+t} Q_x^2(u,u-t)\,\rho(u)\,\mathrm{d}u \;\le |\rho|_* \int_0^{2n} Q_x^2(v,t)\,\mathrm{d}v\,.$$

In view of the inequality (2.78) we obtain

$$\int_0^{2n} Q_x^2(u,t)\,\mathrm{d}u = \int_0^{2n} Q_x^2(t,u)\,\mathrm{d}u \,\le 2\phi_{max}^2\, n\,.$$

Therefore,

$$\max_{0\le t\le n} G(t) \,\le\, 2|\rho|_*\,\phi_{max}^2\, n \quad \text{and} \quad D_{2,n}^{(2)} \le\, 2|\rho|_*^2\,\phi_{max}^2\, n^2\,.$$

So, estimating ϱ_2^2 by $\check{\tau}\sigma_Q$ and taking into account that $\mathbf{E}Y_1^4 \ge 1$, we obtain that we obtain that

$$\mathbf{E}\, D_n(x) \le 13\,(1+\check{\tau})\phi_{max}^4\ \mathbf{E}Y_1^4(1+|\rho|_*^2)\, n^2\sigma_Q\,.$$

Using all these bounds in (2.68) we obtain (2.62) and thus the conclusion follows. □

Remark 2.5. The properties (2.59) and (2.62) are used to obtain the oracle inequalities given in Section 2.3 (see, for example, [17]).

2.4 Efficiency

Now we study the asymptotic efficiency for the procedure (2.23) with the coefficients (2.28), with respect to the robust risk (2.5) defined by the distribution family (2.8)–(2.9). To this end, we assume that the unknown function S in the model (2.1) belongs to the Sobolev ball

$$W_{\mathbf{r}}^k = \{f \in \mathcal{C}_{per}^k[0,1] \,:\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le \mathbf{r}\}\,, \tag{2.79}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some unknown parameters, $\mathcal{C}_{per}^k[0,1]$ 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 $W_{\mathbf{r}}^k$ can be written as an ellipsoid in $\mathbf{L}_2[0,1]$, i.e.,

$$W_{\mathbf{r}}^k = \{f \in \mathcal{C}_{per}^k[0,1] \,:\, \sum_{j=1}^{\infty} a_j\,\theta_j^2 \,\le \mathbf{r}\}, \tag{2.80}$$

where $a_j = \sum_{i=0}^{k} (2\pi[j/2])^{2i}$ and $\theta_j = \int_0^1 f(v)\mathrm{Tr}_j(v)\mathrm{d}v$. We recall that the trigonometric basis $(\mathrm{Tr}_j)_{j\ge1}$ is defined in (1.15).

Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (2.5) is given by

$$\mathbf{r}_k^* = \,((2k+1)\mathbf{r})^{1/(2k+1)}\,\left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)}\,. \tag{2.81}$$

Note that this is the well-known Pinsker constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]). Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the σ-field $\sigma\{y_t\,,\, 0 \le t \le n\}$ generated by the process (2.1).

The following two results give the lower and upper bound for the robust risk in our case.

Theorem 2.4. *Under Conditions* (2.8) *and* (2.9),

$$\liminf_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) \geq \mathbf{r}_k^*\,, \tag{2.82}$$

where $\upsilon_n = n/\varsigma^*$.

Proof. First, we denote by Q_0 the distribution of the noise (2.2) with the parameter $\varrho_1 = \varsigma^*$, $\breve{\varrho} = 1$ and $\varrho_2 = 0$, i.e. the distribution for the "signal + white noise" model. So, we can estimate as below the robust risk

$$\mathcal{R}_n^*(\widetilde{S}_n, S) \geq \mathcal{R}_{Q_0}(\widetilde{S}_n, S)\,.$$

Now Theorem 6.1 from [15] yields the lower bound (2.82). Hence this finishes the proof. □

Note that if the parameters $\mathbf{r}$ and k are known, i.e. for the non-adaptive estimation case, then to obtain the efficient estimation for the "signal+white noise" model. Pinsker in [47] proposed to use the estimate $\widehat{S}_{\lambda_0}$ defined in (2.15) with the weights (2.28) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, l_0)\,, \tag{2.83}$$

where $l_0 = [\mathbf{r}/\varepsilon]\varepsilon$. For the model (2.1) – (2.2) we show the same result.

Proposition 2.4. *The estimator* $\widehat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^*\,.$$

Proof. Putting $\lambda_0(j) = 0$ for $j \geq n$ we can represent the quadratic risk for the estimator (2.15) as

$$\| \widehat{S}_{\lambda_0} - S \|^2 = \sum_{j=1}^{\infty} (1-\lambda_0(j))^2\theta_j^2 - 2H_n + \frac{1}{n}\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2\,,$$

where $H_n = n^{-1/2}\sum_{j=1}^{n}(1-\lambda_0(j))\lambda_0(j)\theta_j\xi_{j,n}$. Note that $\mathbf{E}_Q H_n = 0$ for any $Q \in Q_n$, therefore,

$$\mathbf{E}_Q \| \widehat{S}_{\lambda_0} - S \|^2 = \sum_{j=1}^{\infty} (1-\lambda_0(j))^2\theta_j^2 + \frac{1}{n}\mathbf{E}_Q\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2\,.$$

Proposition 2.2 and the last inequality in (2.8) imply that for any $Q \in \mathcal{Q}_n$

$$\mathbf{E}_Q\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2 \leq \varsigma^*\sum_{j=1}^{n}\lambda_0^2(j) + \frac{\phi_{max}^2\varsigma^*\|\Upsilon\|_1}{\check{\tau}} := \varsigma^*\sum_{j=1}^{n}\lambda_0^2(j) + \mathbf{C}_{1,n}^*\,.$$

Therefore,

$$\mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \sum_{j=j_*}^{\infty}(1-\lambda_0(j))^2\theta_j^2 + \frac{1}{\upsilon_n}\sum_{j=1}^{n}\lambda_0^2(j) + \frac{\mathbf{C}_{1,n}^*}{n},$$

where j_* and υ_n are defined in (2.28). Setting

$$\Upsilon_{1,n}(S) = \upsilon_n^{2k/(2k+1)} \sum_{j=j_*}^{\infty} (1-\lambda_0(j))^2 \theta_j^2 \quad \text{and} \quad \Upsilon_{2,n} = \frac{1}{\upsilon_n^{1/(2k+1)}} \sum_{j=1}^{n} \lambda_0^2(j)\,,$$

we rewrite the last inequality as

$$\upsilon_n^{2k/(2k+1)}\, R_n^*(\widehat{S}_{\lambda_0}, S) \le \Upsilon_{1,n}(S) + \Upsilon_{2,n} + \check{\mathbf{C}}_n\,, \tag{2.84}$$

where $\check{\mathbf{C}}_n = \upsilon_n^{2k/(2k+1)} \mathbf{C}_{1,n}^*/n$. Note that Conditions (2.9) and (2.56) imply that $\mathbf{C}_{1,n}^* = (n^{\check{\delta}})$ as $n \to \infty$ for any $\check{\delta} > 0$; therefore, $\check{\mathbf{C}}_n \to 0$ as $n \to \infty$. Putting

$$u_n = \upsilon_n^{2k/(2k+1)} \sup_{j \ge j_*} (1-\lambda_0(j))^2/a_j\,,$$

with a_j defined in (2.80), we estimate the first term in (2.84) as

$$\sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \le \sup_{S \in W_{\mathbf{r}}^k} u_n \sum_{j \ge 1} a_j \theta_j \le u_n \mathbf{r}\,.$$

Taking into account that $a_j/(\pi^{2k} j^{2k}) \to 1$ as $j \to \infty$ and $l_0 \to \mathbf{r}$ as $\varepsilon \to 0$ and using the definition of ω_{α_0} in (2.28), we obtain that

$$\begin{aligned}\limsup_{n\to\infty} u_n &\le \lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{j \ge j_*} \frac{(1-\lambda_0(j))^2}{(\pi\, j)^{2k}} \\ &= \lim_{n\to\infty} \frac{\upsilon_n^{2k/(2k+1)}}{\pi^{2k} \omega_{\alpha_0}^{2k}} = \frac{1}{\pi^{2k}\,(\mathrm{d}_k \mathbf{r})^{2k/(2k+1)}}\,.\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty} \sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \le \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k} (\mathrm{d}_k)^{2k/(2k+1)}} =: \Upsilon_1^*\,. \tag{2.85}$$

As to the second term in (2.84), note that

$$\lim_{n\to\infty} \frac{1}{\omega_{\alpha_0}} \sum_{j=1}^{n} \lambda_0^2(j) = \int_0^1 (1-t^k)^2 \mathrm{d}t = \frac{2k^2}{(k+1)(2k+1)}\,.$$

So, taking into account that $\omega_{\alpha_0}/\upsilon_n^{1/(2k+1)} \to (\mathrm{d}_k \mathbf{r})^{1/(2k+1)}$ as $n \to \infty$, the limit of $\Upsilon_{2,n}$ can be calculated as

$$\lim_{n\to\infty} \Upsilon_{2,n} = \frac{2(\mathrm{d}_k \mathbf{r})^{1/(2k+1)}\, k^2}{(k+1)(2k+1)} =: \Upsilon_2^*\,.$$

Moreover, since $\Upsilon_1^* + \Upsilon_2^* =: \mathbf{r}_k^*$, we obtain

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \le \mathbf{r}_k^*$$

and get the desired result. □

For the adaptive estimation we use the model selection procedure (2.23) with the parameter δ defined as a function of n satisfying

$$\lim_n \delta_n = 0 \quad \text{and} \quad \lim_n n^{\check{\delta}}\, \delta_n = 0 \tag{2.86}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$.

Theorem 2.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then the robust risk defined in* (2.5) *through the distribution family* (2.8)–(2.9) *for the procedure* (2.23) *based on the trigonometric basis* (1.15) *with the coefficients* (2.28) *and the parameter* $\delta = \delta_n$ *satisfying* (2.86) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le \mathbf{r}_k^* \,. \tag{2.87}$$

Theorem 2.4 and Theorem 2.5 allow us to compute the optimal convergence rate.

Corollary 2.1. *Under the assumptions of Theorem 2.5, we have*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^* \,. \tag{2.88}$$

Remark 2.6. 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, [47], [46]). We see here that the efficient robust rate is $\upsilon_n^{2k/(2k+1)}$, i.e., if the distribution upper bound $\varsigma^* \to 0$ as $n \to \infty$, we obtain a faster rate with respect to $n^{2k/(2k+1)}$, and, if $\varsigma^* \to \infty$ as $n \to \infty$, we obtain a slower rate. In the case when ς^* is constant, than the robust rate is the same as the classical non robust convergence rate.

2.5 Simulations

In this section we report the results of a Monte Carlo experiment in order to assess the performance of the proposed model selection procedure (2.23). In (2.1) we chose a 1-periodic function which is defined, for $0 \le t \le 1$, as

$$S(t) = t\sin(2\pi t) + t^2(1-t)\cos(4\pi t)\,. \tag{2.89}$$

We simulate the model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,,$$

where $\xi_t = 0.5\mathrm{d}wt + 0.5\mathrm{d}z_t$.

Here z_t is the semi-Markov process defined in (1.6) with a Gaussian $\mathcal{N}(0,1)$ sequence $(Y_j)_{j\ge1}$ and $(\tau_k)_{k\ge1}$ used in (1.7) taken as $\tau_k \sim \chi_3^2$.

n	$\overline{\mathbf{R}}$	$\overline{\mathbf{R}_*}$
20	0.04430	0.235
100	0.01290	0.068
200	0.00812	0.043
1000	0.00196	0.010

Table 2.1: Empirical risks

We use the model selection procedure (2.23) with the weights (2.28) in which $k^* = 100 + \sqrt{\ln(n)}$, $t_i = i/\ln(n)$, $m = [\ln^2(n)]$ and $\delta = (3 + \ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathbf{R}} = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\left(\widehat{S}_n(t_j) - S(t_j)\right)^2, \tag{2.90}$$

where the observation frequency $p = 100001$ and the expectation was taken as an average over $N = 10000$ replications, i.e.,

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

We set the relative quadratic risk as

$$\overline{\mathbf{R}_*} = \overline{\mathbf{R}}/||S||_p^2, \quad \text{with} \quad ||S||_p^2 = \frac{1}{p}\sum_{j=0}^{p} S^2(t_j)\,. \tag{2.91}$$

In our case $||S||_p^2 = 0.1883601$.

Table 2.1 gives the values for the sample risks (2.90) and (2.91) for different numbers of observations n.

Figures 2.1–2.4 show the behaviour of the regression function and its estimates by the model selection procedure (2.23) depending on the values of observation periods n. The black full line is the regression function (2.89) and the red dotted line is the associated estimator.

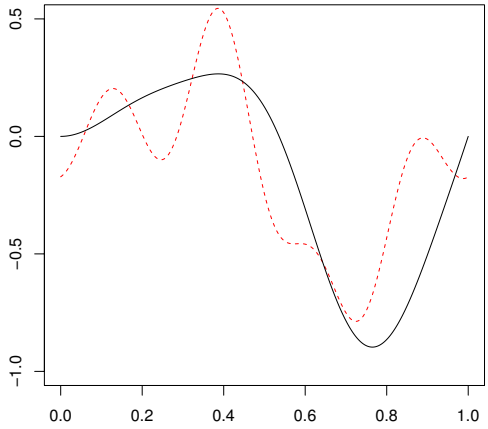

Figure 2.1: Estimator of S for $n = 20$

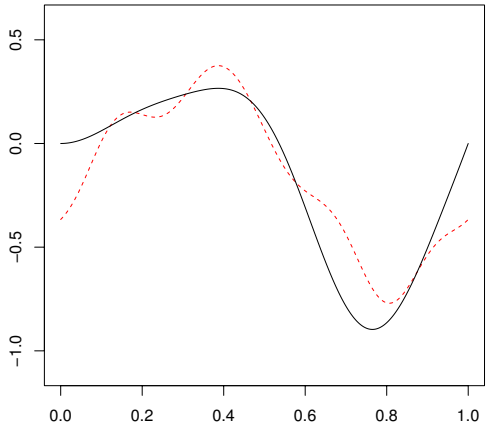

Figure 2.2: Estimator of S for $n = 100$

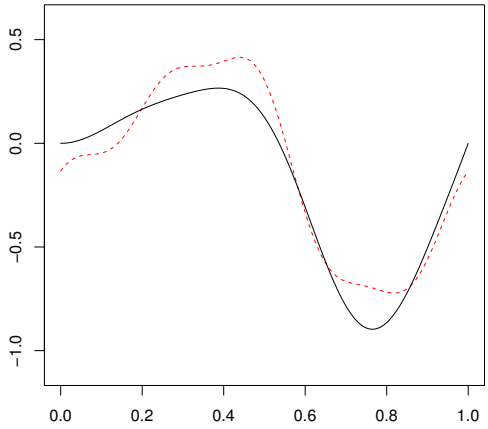

Figure 2.3: Estimator of S for $n = 200$

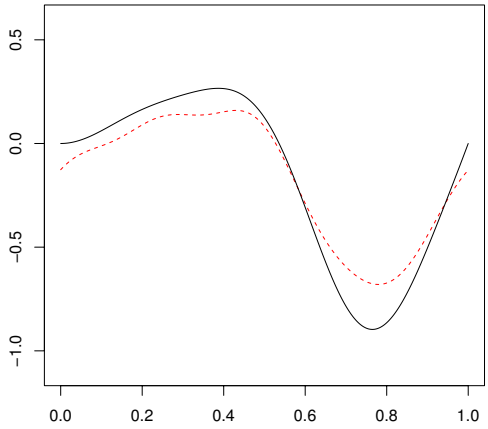

Figure 2.4: Estimator of S for $n = 1000$

Remark 2.7. From numerical simulations of the procedure (2.23) with various observation numbers n we may conclude that the quality of the proposed procedure: (i) is good for practical

needs, i.e. for reasonable (non large) number of observations; (ii) is improving as the number of observations increases.

Now we give the algorithm of the model selection procedure given in Section 2.2

Algorithm 1 Model selection procedure

Require: n, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

ρ_1, ρ_2, ς^*: satisfying Conditions (2.8) and (2.9)

$k^* \geq 1$, ε: satisfying Condition (2.26)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$m = [1/\varepsilon^2]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

    for $k \longleftarrow 1$ *to* n **do**

      Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (2.28)

    end for

  end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varepsilon, \ldots, m\varepsilon\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,n} = \frac{1}{n} \int_0^n \phi_k(t) \mathrm{d}\, y_t\,.$

$\widetilde{\theta}_{k,n} \longleftarrow \widehat{\theta}^2_{k,n} - \frac{1}{n}\,.$

The observation $(y_t)_{0 \leq t \leq n}$ are given in (2.1) with the noise process (2.2) and $(\phi_k)_{k \geq 1}$ is the basis given in (2.10)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,n}, ..., \widehat{\theta}_{n,n})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,n}, ..., \widetilde{\theta}_{n,n})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

    $J_n(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^n \lambda_{i,j}(l)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda)$.

    where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_n is the penalty term given in (2.22)

  end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_n(\lambda), \Lambda = \{\lambda_\alpha\,,\, \alpha \in \mathcal{A}\}$.

2.6 Stochastic calculus for semi-Markov processes

In this section we give some results of stochastic calculus for the process $(\xi_t)_{t\geq 0}$ given in (2.2), needed all along this paper. As the process ξ_t is the combination of a Lévy process and a semi-Markov process, these results are not standard and need to be provided.

Lemma 2.1. *Let f and g be any non-random functions from $\mathbf{L}_2[0,n]$ and $(I_t(f))_{t\geq 0}$ be the process defined in* (2.6)*. Then, for any $0 \leq t \leq n$,*

$$\mathbf{E}\, I_t(f) I_t(g) = \varrho_1^2\,(f,g)_t \,+ \varrho_2^2\,(f, g\rho)_t\,, \tag{2.92}$$

where $(f,g)_t = \int_0^t f(s)\, g(s)\mathrm{d}s$ and ρ is the density defined in (1.8)*.*

Proof. First, note that we can represent the stochastic integral $I_t(f)$ as

$$I_t(f) = \varrho_1 I_t^L(f) + \varrho_2 I_t^z(f)\,, \tag{2.93}$$

where

$$I_t^L(f) = \int_0^t f(s)\mathrm{d}L_s \quad \text{and} \quad I_t^z(f) = \int_0^t f(s)\mathrm{d}z_s\,.$$

Note that the mutual covariation for the martingales $I_t^L(f)$ and $I_t^L(g)$ (see, for example, [21]) may be calculated as

$$[I^L(f), I^L(g)]_t = \check{\varrho}^2 \int_0^t f(s)g(s)\mathrm{d}s + (1-\check{\varrho}^2) \sum_{0\leq s\leq t} f(s)g(s)\left(\Delta \check{L}_s\right)^2\,, \tag{2.94}$$

where $\Delta\check{L}_s = \check{L}_s - \check{L}_{s-}$. Taking into account that $\mathbf{E}\, I_t^L(f)\, I_t^L(g) = \mathbf{E}\,[I^L(f), I^L(g)]_t$ and that in view of the first condition in (2.3) $\Pi(x^2) = 1$, we obtain that

$$\begin{aligned}\mathbf{E}\, I_t^L(f)\, I_t^L(g) &= \check{\varrho}^2 \int_0^t f(s)g(s)\mathrm{d}s + (1-\check{\varrho}^2)\,\Pi(x^2) \int_0^t f(s)\, g(s)\mathrm{d}s \\ &= \int_0^t f(s)\, g(s)\mathrm{d}s\,.\end{aligned} \tag{2.95}$$

Moreover, note that

$$\begin{aligned}\mathbf{E} I_t^z(f) I_t^z(g) &= \mathbf{E}\left(\sum_{l=1}^{\infty} f(T_l)g(T_l)Y_l^2 1_{\{T_l\leq t\}}\right) \\ &= \mathbf{E}\left(\sum_{l=1}^{\infty} f(T_l)g(T_l)1_{\{T_l\leq t\}}\right) = \int_0^t f(s)g(s)\rho(s)\mathrm{d}s\,.\end{aligned}$$

Hence the conclusion follows. □

Lemma 2.2. *Assume that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then, for any $n \geq 1$ and for any non random function f from $\mathbf{L}_2[0,n]$, the stochastic integral* (2.6) *exists and satisfies the properties* (2.7) *with the coefficient $\varkappa_Q$ given in* (2.7)*.*

Proof. This lemma follows directly from Lemma 2.1 with $f = g$ and Proposition 5.1. □

Lemma 2.3. *Let f and g be bounded functions defined on $[0,\infty)\times\mathbb{R}$. Then, for any $k \geq 1$,*

$$\mathbf{E}\left(I_{T_{k-}}(f)\, I_{T_{k-}}(g) \mid \mathcal{G}\right) = \varrho_1^2(f\,,\, g)_{T_k} + \varrho_2^2 \sum_{l=1}^{k-1} f(T_l)\, g(T_l),$$

where $\mathcal{G}$ is the σ-field generated by the sequence $(T_l)_{l\geq 1}$, i.e., $\mathcal{G} = \sigma\{T_l\,,\, l \geq 1\}$.

Proof. Using (2.93), (2.95) and, taking into account that the process $(L_t)_{t\geq 0}$ is independent of $\mathcal{G}$, we obtain

$$\mathbf{E}\left(I_{T_{k-}}(f)\, I_{T_{k-}}(g) \mid \mathcal{G}\right) = \varrho_1^2(f\,,\, g)_{T_k} + \mathbf{E}\left(I^z_{T_{k-}}(f)\, I^z_{T_{k-}}(g) \mid \mathcal{G}\right).$$

Moreover,

$$\begin{aligned}\mathbf{E}\left(I^z_{T_{k-}}(f)\, I^z_{T_{k-}}(g) \mid \mathcal{G}\right) &= \mathbf{E}\left(\left(\sum_{l=1}^{k-1} f(T_l)Y_l\right)\left(\sum_{l=1}^{k-1} g(T_l)Y_l\right) \mid \mathcal{G}\right)\\ &= \sum_{l=1}^{k-1} f(T_l)\, g(T_l)\,.\end{aligned}$$

Thus we obtain the desired result. □

Lemma 2.4. *Assume that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then, for any measurable bounded non-random functions f and g, we have*

$$\left|\mathbf{E}\int_0^n I^2_{t-}(f)\, g(t)\,\mathrm{d}m_t\right| \leq 2\varrho_2^2 |g|_*|f|_*^2\, \|\Upsilon\|_1\, n.$$

Proof. Using the definition of the process $(m_t)_{t\geq 0}$ we can represent this integral as

$$\begin{aligned}\int_0^n I^2_{t-}(f)\, g(t)\,\mathrm{d}m_t &= \sum_{k\geq 1} I^2_{T_k-}(f)\, g(T_k)\, Y_k^2\, \mathbf{1}_{\{T_k\leq n\}}\\ &\quad - \int_0^n I_t^2(f)\, g(t)\, \rho(t)\,\mathrm{d}t =: V_n - U_n\,. \qquad (2.96)\end{aligned}$$

Note now that

$$\mathbf{E}\, V_n \;= \mathbf{E}\sum_{k\geq 1} g(T_k)\, \mathbf{E}\left(I^2_{T_k-}(f)\;\mid \mathcal{G}\right)\, \mathbf{1}_{\{T_k\leq n\}}\,.$$

Now, using Lemma 2.3 we can represent the last expectation as

$$\mathbf{E}\, V_n \;= \varrho_1^2\, \mathbf{E}\, V_n' + \varrho_2^2\, \mathbf{E}\, V_n'', \qquad (2.97)$$

where

$$V_n^{'} = \sum_{k\geq 1} g(T_k)\, \|f\|^2_{T_k}\, \mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad V_n^{''} = \sum_{k\geq 2} g(T_k)\, \mathbf{1}_{\{T_k\leq n\}} \sum_{l=1}^{k-1} f^2(T_l)\,.$$

The first term in (2.97) can be represented as

$$\mathbf{E}\, V_n^{'} = \int_0^n g(t)\, \|f\|_t^2 \rho(t)\mathrm{d}t\,.$$

To estimate the last expectation in (2.97), note that

$$\mathbf{E}\, V_n^{''} = \mathbf{E} \sum_{l\geq 1} f^2(T_l)\, \bar{g}(T_l) \mathbf{1}_{\{T_l\leq n\}} = \int_0^n f^2(v)\, \bar{g}(v)\, \rho(v)\mathrm{d}v\,,$$

where

$$\bar{g}(v) = \mathbf{E} \sum_{k\geq 1} g(v+T_k)\, \mathbf{1}_{\{T_k\leq n-v\}} = \int_v^n g(t)\, \rho(t-v)\mathrm{d}t\,.$$

Moreover, using now the representation (2.92), we calculate the expectation of the last term in (2.96)

$$\mathbf{E}\, U_n = \varrho_1^2 \int_0^n \|f\|_t^2\, g(t)\, \rho(t)\, \mathrm{d}t + \varrho_2^2 \int_0^n \check{f}(t)\, g(t)\, \rho(t)\, \mathrm{d}t\,,$$

where $\check{f}(t) = \int_0^t f^2(s)\, \rho(s)\, \mathrm{d}s$. This implies that

$$\mathbf{E} \int_0^n I_{t-}^2(f)\, g(t)\, \mathrm{d}m_t = \varrho_2^2 \int_0^n g(t)\, \delta(t)\mathrm{d}t\,,$$

where $\delta(t) = \int_0^t f^2(v)\, (\rho(t-v) - \rho(t))\, \rho(v)\, \mathrm{d}v$. Note that, in view of Proposition 5.1, the function δ can be estimated as

$$|\delta(t)| \leq |f|_*^2\, |\rho|_* \int_0^t |\Upsilon(t-v) - \Upsilon(t)|\, \mathrm{d}v \leq |f|_*^2\, |\rho|_*\, (\|\Upsilon\|_1 + t|\Upsilon(t)|)\,.$$

Therefore,

$$\left| \mathbf{E} \int_0^n I_{t-}^2(f)\, g(t)\, \mathrm{d}m_t \right| \leq 2\varrho_2^2 |g|_* |f|_*^2\, \|\Upsilon\|_1\, n$$

and this finishes the proof. □

Lemma 2.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any measurable bounded non-random functions* f *and* g*, one has*

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = 0.$$

Proof. First, note that

$$\int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = \varrho_1 \int_0^n I_t^2(f) I_t(g) g(t) \mathrm{d}L_t + \varrho_2 \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) \mathrm{d}z_t.$$

Second, we will show that

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) \mathrm{d}L_t = 0\,. \tag{2.98}$$

Using the notations (2.93), we set

$$J_1 = \int_0^n I_t^2(f) I_t^L(g) g(t) \mathrm{d}L_t \quad \text{and} \quad J_2 = \int_0^n I_t^2(f) I_t^z(g) g(t) \mathrm{d}L_t,$$

we obtain that

$$\int_0^n I_t^2(f) I_t(g) g(t) \mathrm{d}L_t = \varrho_1 J_1 + \varrho_2 J_2 \,. \tag{2.99}$$

Now let us recall the Novikov inequalities, [44], also referred to as the Bichteler–Jacod inequalities (see [32, 43]) providing bound moments of supremum of purely discontinuous local martingales for any predictable function h and any $p \geq 2$

$$\mathbf{E} \sup_{0 \leq t \leq n} \left| \int_{[0,t] \times \mathbb{R}} h \, \mathrm{d}(\mu - \nu) \right|^p \leq C_p^* \mathbf{E} \, \check{J}_{p,n}(h) \,, \tag{2.100}$$

where C_p^* is some positive constant and

$$\check{J}_{p,n}(h) = \left(\int_{[0,n] \times \mathbb{R}} h^2 \, \mathrm{d}\nu \right)^{p/2} + \int_{[0,n] \times \mathbb{R}} h^p \, \mathrm{d}\nu \,.$$

By applying this inequality for the non-random function $h(s,x) = g(s)x$, and, recalling that $\Pi(x^8) < \infty$, we obtain,

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^{\check{L}}(g) \right|^8 < \infty \,.$$

Taking into account that, for any non random square integrated function f, the integral $\left(\int_0^t f(s) \mathrm{d}w_s \right)$ is Gaussian with the parameters $\left(0, \int_0^t f^2(s) \mathrm{d}s \right)$, we obtain

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^L(g) \right|^8 < \infty.$$

Finally, by using the Cauchy's inequality, we can estimate for any $0 < t \leq n$ the following expectation as

$$\mathbf{E} \, (I_t^L(f))^4 (I_t^L(g))^2 \leq \sqrt{\mathbf{E} \, (I_t^L(f))^8} \sqrt{\mathbf{E} \, (I_t^L(f))^4}$$

i.e.,

$$\sup_{0 \leq t \leq n} \mathbf{E} \, (I_t^L(f))^4 (I_t^L(g))^2 < \infty \,.$$

Moreover, taking into account that the processes $(L_t)_{t \geq 0}$ and $(z_t)_{t \geq 0}$ are independent, we obtain that

$$\mathbf{E} \, (I_t^z(f))^4 (I_t^L(g))^2 = \mathbf{E} \, (I_t^z(f))^4 \mathbf{E} \, (I_t^L(g))^2 = \int_0^t g^2(s) \mathrm{d}s \, \mathbf{E} \, (I_t^z(f))^4 \,.$$

One can check directly here that, for $t > 0$,

$$\mathbf{E} \, |I_t^z(f)|^4 \leq |f|_*^4 \, \mathbf{E} \, Y_1^4 \, \mathbf{E} N_t^2 \,.$$

Note that the last bound in Corollary 5.1 yields $\sup_{0\le t\le n}\mathbf{E}\,(I_t^z(f))^4<\infty$ and, therefore,

$$\sup_{0\le t\le n}\mathbf{E}\,(I_t(f))^4(I_t^L(g))^2<\infty\,.$$

It follows directly that $\mathbf{E}J_1=0$. Now we study the last term in (2.99). To this end, first note that similarly to the previous reasoning we obtain that

$$\mathbf{E}\int_0^n(I_t^L(f))^2I_t^z(g)g(t)\mathrm{d}L_t=0\quad\text{and}\quad\mathbf{E}\int_0^n I_t^L(f)I_t^z(f)I_t^z(g)g(t)\mathrm{d}L_t=0\,.$$

Therefore, to show (2.98) one needs to show that

$$\mathbf{E}\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t=0\,. \tag{2.101}$$

To check this, note that, for any $0<t\le n$ and for any bounded function f,

$$I_t^z(f)=\sum_{k=1}^{\infty}f(T_k)\,Y_k\,\mathbf{1}_{\{T_k\le t\}}=\sum_{k=1}^{N_n}f(T_k)\,Y_k\,\mathbf{1}_{\{T_k\le t\}}\,,$$

i.e.,

$$\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t=\sum_{k=1}^{N_n}\sum_{l=1}^{N_n}\sum_{j=1}^{N_n}f(T_k)\,f(T_l)\,g(T_j)\,Y_jY_l\,Y_k\,I_{klj}\,,$$

where

$$I_{klj}=\int_0^n\mathbf{1}_{\{T_k\le t\}}\mathbf{1}_{\{T_l\le t\}}\mathbf{1}_{\{T_j\le t\}}\mathrm{d}L_t\,.$$

Taking into account that the $(L_t)_{t\ge0}$ is independent of the field $\mathcal{G}_z=\sigma\{z_t\,,t\ge0\}$, we obtain that $\mathbf{E}\left(I_{klj}|\mathcal{G}_z\right)=0$. Therefore,

$$\begin{aligned}&\mathbf{E}\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t\\ =\;&\mathbf{E}\sum_{k=1}^{N_n}\sum_{l=1}^{N_n}\sum_{j=1}^{N_n}f(T_k)\,f(T_l)\,g(T_j)\,Y_jY_l\,Y_k\,\mathbf{E}\left(I_{klj}|\mathcal{G}_z\right)=0.\end{aligned}$$

So, we obtain (2.101) and hence the proof is achieved. □

Appendix

Property of the penalty term

Lemma 2.6. *For any $n\ge1$ and $\lambda\in\Lambda$,*

$$P_n^0(\lambda)\le\mathbf{E}_QErr_n(\lambda)+\frac{\mathbf{C}_{1,Q,n}}{n},$$

where the coefficient $P_n^0(\lambda)$ was defined in (2.38).

Proof. By the definition of $\mathrm{Err}_n(\lambda)$ one has

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n} \left((\lambda(j) - 1)\theta_j + \frac{\lambda(j)}{n} \xi_{j,n} \right)^2 .$$

In view of Proposition 2.2, this leads to the desired result

$$\mathbf{E}_Q \, \mathrm{Err}_n(\lambda) \geq \frac{1}{n} \sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q \, \xi_{j,n}^2 \geq P_n^0(\gamma) - \frac{\mathbf{C}_{1,Q,n}}{n}.$$

Chapter 3

Non-parametric estimation for semi-Markov regression models based on discrete data

3.1 Introduction

Let us consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{3.1}$$

where S is an unknown 1-periodic function defined on $\mathbb{R}$ with values on $\mathbb{R}$, $(\xi_t)_{t\ge 0}$ is the unobserved noise process (2.2) . The problem is to estimate the unknown function S in model (3.1) on the basis of observations

$$(y_{t_j})_{0\le j\le np}, \quad t_j = j\Delta, \quad \Delta = \frac{1}{p}, \tag{3.2}$$

where integer $p \ge 1$ is the observation frequency. In this chapter we use the risks defined in (2.4) and (2.5) for the distribution family $\mathcal{Q}_n$.

The goal of this chapter is to develop a robust efficient model selection method for the model (3.1) with the semi-Markov dependence having unknown distribution. We use the approach proposed by Konev and Pergamenshchikov in [17] and [18] for continuos time regression models with non martingale noises. Unfortunately, we cannot use directly their method for the semi-Markov regression models, since their tool essentially uses the fact that the Ornstein - Uhlenbeck dependence decreases with geometrical rate and obtain sufficiently quickly the"white noise" case. In this chapter we propose new analytical tools based on renewal methods, to obtain the sharp non-asymptotic oracle inequalities. And, as a consequence, we obtain robust efficiency for proposed model selection procedures.

3.2 Model selection

In this chapter we will use the trigonometric basis $(\mathrm{Tr}_j)_{j\ge 1}$ in $\mathbf{L}_2[0,1]$ defined in (1.15). By making use of this basis, we consider the discrete Fourier transform of S

$$S(t) = \sum_{j=1}^{p} \theta_{j,p}\, \mathrm{Tr}_j(t), \quad t \in \{t_1, ..., t_p\}, \tag{3.3}$$

where the Fourier coefficients are defined by

$$\theta_{j,p} = (S, \mathrm{Tr}_j)_p = \frac{1}{p}\sum_{i=1}^{p} S(t_i)\mathrm{Tr}_j(t_i). \tag{3.4}$$

In the sequel the corresponding norm will be denoted by $\|x\|_p^2 = (x,x)_p$. These Fourier coefficients $\theta_{j,p}$ can be estimated by

$$\widehat{\theta}_{j,p} = \frac{1}{n}\int_0^n \Psi_{j,p}(t)\mathrm{d}\, y_t \quad \text{and} \quad \Psi_{j,p}(t) = \sum_{l=1}^{np} \mathrm{Tr}_j(t_l) 1_{\{t_{l-1}<t\le t_l\}}\,. \tag{3.5}$$

We note that the system of functions $(\Psi_{j,p})_{1\le j\le p}$ is orthonormal in $\mathbf{L}_2[0,1]$ because

$$\int_0^1 \Psi_{j,p}(t)\Psi_{i,p}(t)\mathrm{d}\, t \ = (\mathrm{Tr}_j, \mathrm{Tr}_i)_p = 1_{\{i=j\}}\,.$$

In the sequel we need the Fourier coefficients for the function S with respect to the new basis $(\Psi_{j,p})_{1\le j\le p}$. These coefficiens can be writen as

$$\overline{\theta}_{j,p} = \int_0^1 S(t)\Psi_{i,p}(t)\mathrm{d}\, t \ = \theta_{j,p} + h_{j,p}, \tag{3.6}$$

where

$$h_{j,p}(S) = \sum_{l=1}^{p}\int_{t_{l-1}}^{t_l} \mathrm{Tr}_j(t_l)(S(t) - S(t_l))\mathrm{d}\, t\,.$$

From (3.1) it follows directly that these Fourier coefficients satisfy the equation

$$\widehat{\theta}_{j,p} = \overline{\theta}_{j,p} + \frac{1}{\sqrt{n}}\xi_{j,p}, \quad \text{where} \quad \xi_{j,p} = \frac{1}{\sqrt{n}} I_n(\Psi_{j,p})\,. \tag{3.7}$$

For any $0 \le t \le 1$ we estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\Psi_{j,p}(t)\,, \tag{3.8}$$

where the weight vector $\lambda = (\lambda(1), ..., \lambda(n))$ belongs to some finite set Λ from $[0,1]^n$, $\widehat{\theta}_{j,p}$ was defined in (3.5). Now let us consider

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = \max_{\lambda\in\Lambda} L(\lambda)\,, \tag{3.9}$$

where $\#(\Lambda)$ is the cardinal number of Λ and $L(\lambda) = \sum_{j=1}^{n}\lambda(j)$. In the sequel we assume that $|\Lambda|_* \ge 1$ and $\lambda(j) = 0$ for $j \ge p$.

In order to find a proper weight sequence λ in the set Λ one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. Let as consider the empirical squared error

$$\mathrm{Err}(\lambda) = \|\widehat{S}_\lambda - S\|^2\,, \tag{3.10}$$

which in our case is equal to

$$\mathrm{Err}(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,p} - 2\sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\overline{\theta}_{j,p} + \|S\|^2 . \tag{3.11}$$

Since the Fourier coefficients $(\theta_j)_{j\geq 1}$ are unknown, the weight coefficients $(\lambda(j))_{1\leq j\leq p}$ cannot be determined by minimizing this quality. To circumvent this difficulty, one needs to replace the terms $\widehat{\theta}_{j,p}\overline{\theta}_{j,p}$ by their estimators $\widetilde{\theta}_{j,p}$. Let us set

$$\widetilde{\theta}_{j,p} = \widehat{\theta}^2_{j,p} - \frac{\widehat{\sigma}_n}{n}, \tag{3.12}$$

where $\widehat{\sigma}_n$ is an estimate of the proxy variance σ_Q defined in (2.8). For, example, we can take it as

$$\widehat{\sigma}_n = \frac{n}{\check{p}} \sum_{j=l}^{\check{p}} \widehat{\theta}^2_{j,p} \quad \text{and} \quad \check{p} = \min(p, n) , \tag{3.13}$$

where $l = [\sqrt{n}]$, and we set $\widehat{\sigma}_n = 0$ *for* $l > p$. For this change in the empirical squared error, one has to pay some penalty. Thus, we obtain the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda), \tag{3.14}$$

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

$$P_n(\lambda) = \frac{\widehat{\sigma}_n|\lambda|^2}{n}. \tag{3.15}$$

Minimizing the cost function, that is

$$\hat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_n(\lambda), \tag{3.16}$$

and substituting the obtained weight coefficients $\hat{\lambda}$ in (3.8), lead to the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}. \tag{3.17}$$

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

3.3 Oracle inequality

In order to prove the oracle inequality, the following conditions will be needed for the noise $(\xi_t)_{t\geq 0}$. Here we use the conditions introduced in [17] for the general semi-martingale model (2.1).

$\mathbf{L}_1$) *For all* $n \geq 1$ *and* Q *there exist a variance proxy* $\sigma_Q > 0$ *and a constant* $\mathbf{L}_{1,Q} \geq 0$ *such that*

$$\sup_{p\geq 3} \sup_{x\in[-1,1]^n} \left|B_{1,Q}(x)\right| \leq \mathbf{L}_{1,Q} < \infty\,,$$

where $B_{1,Q}(x) = \sum_{j=1}^n x_j \left(\mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q\right)$.

$\mathbf{L}_2$) *For all* $n \geq 1$ *and* Q *there exists a constant* $\mathbf{L}_{2,Q} \geq 1$ *such that*

$$\sup_{p\geq 3} \sup_{|x|\leq 1} \mathbf{E}\, B_{2,Q}^2(x) \leq \mathbf{L}_{2,Q} < \infty,$$

where $B_{2,Q}(x) = \sum_{j=1}^n x_j\, \widetilde{\xi}_{j,p}$ *and* $\widetilde{\xi}_{j,p} = \xi_{j,p}^2 - \mathbf{E}_Q \xi_{j,p}^2$.

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

$$\mathbf{g}_{n,p} = 1 + |\Lambda|_* \left(\frac{\sqrt{n}}{\check{p}} + \frac{1}{\sqrt{\check{p}}}\right). \tag{3.18}$$

Firstly, we obtain the non asymptotic oracle inequality for the model selection procedure (3.17).

Theorem 3.1. *Assume that Conditions* $\mathbf{L}_1$) *and* $\mathbf{L}_2$) *hold. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, the weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17)*, satisfies the following oracle inequality*

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + l^* \frac{\check{\iota}}{\delta n} \left(\sigma_Q + |\Lambda|_* \mathbf{E}_Q |\widehat{\sigma}_n - \sigma_Q|\right).\end{aligned} \tag{3.19}$$

Proof. Using the cost function given in (3.14), we can rewrite the empirical squared error in (3.11) as follows

$$\mathrm{Err}(\lambda) = J_n(\lambda) + 2\sum_{j=1}^n \lambda(j)\check{\theta}_{j,p} + \|S\|^2 - \rho\hat{P}_n(\lambda), \tag{3.20}$$

where

$$\check{\theta}_{j,p} = \widetilde{\theta}_{j,p} - \overline{\theta}_{j,p}\widehat{\theta}_{j,p} = \frac{1}{\sqrt{n}}\overline{\theta}_{j,p}\xi_{j,p} + \frac{1}{n}\widetilde{\xi}_{j,p} + \frac{1}{n}\varsigma_{j,n} + \frac{\sigma_Q - \widehat{\sigma}_n}{n},$$

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

$$M(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \lambda(j)\theta_j\xi_{j,p} \quad \text{and} \quad P_n^0 = \frac{\sigma_Q|\lambda|^2}{n}\,, \tag{3.21}$$

we can rewrite (3.20) as

$$\begin{aligned}\mathrm{Err}(\lambda) = J_n(\lambda) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n} L(\lambda) + 2M(\lambda) + \frac{2}{n} B_{1,Q}(\lambda) \\ + 2\sqrt{P_n^0(\lambda)}\frac{B_{2,Q}(e(\lambda))}{\sqrt{\sigma_Q n}} + \|S\|^2 - \rho P_n(\lambda),\end{aligned} \tag{3.22}$$

where $e(\lambda) = \lambda/|\lambda|$ and the function L was defined in (3.9). Let $\lambda_0 = (\lambda_0(j))_{1\le j\le p}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be defined as in (3.16). Substituting λ_0 and $\widehat{\lambda}$ in Equation (3.22), we obtain

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) - \mathrm{Err}(\lambda_0) =& J(\widehat{\lambda}) - J(\lambda_0) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n} L(\varpi) + \frac{2}{n} B_{1,Q}(\varpi) + 2M(\varpi) \\ &+ 2\sqrt{P_n^0(\widehat{\lambda})}\frac{B_{2,Q}(\widehat{e})}{\sqrt{\sigma_Q n}} - 2\sqrt{P_n^0(\lambda_0)}\frac{B_{2,Q}(e_0)}{\sqrt{\sigma_Q n}} \\ &- \delta P_n(\widehat{\lambda}) + \delta P_n(\lambda_0),\end{aligned} \tag{3.23}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{e} = e(\widehat{\lambda})$ and $e_0 = e(\lambda_0)$. Note that, by (3.9),

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

The inequality

$$2|ab| \le \delta a^2 + \delta^{-1} b^2 \tag{3.24}$$

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

$$2\sqrt{P_n^0(\lambda)}\frac{|B_{2,Q}(e(\lambda))|}{\sqrt{\sigma_Q n}} \le \delta P_n^0(\lambda) + \frac{B_{2,Q}^2(e(\lambda))}{\delta \sigma_Q\, n}.$$

Taking into account that $0 < \delta < 1$, we get

$$\begin{aligned}\mathrm{Err}(\hat{\lambda}) \le \mathrm{Err}(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{L}_{1,Q}}{n} + \frac{2B_{2,Q}^*}{\delta\sigma_Q\, n} \\ + \frac{1}{n}|\widehat{\sigma}_n - \sigma_Q|(|\widehat{\lambda}|^2 + |\lambda_0|^2) + 2\delta P_n(\lambda_0)\,,\end{aligned}$$

where $B_{2,Q}^* = \sup_{\lambda\in\Lambda} B_{2,Q}^2((e(\lambda))$. Moreover, noting that in view of (3.9) $\sup_{\lambda\in\Lambda} |\lambda|^2 \le |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \mathrm{Err}(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{L}_{1,Q}}{n} + \frac{2B_{2,Q}^*}{\delta\sigma_Q n} \\ &+ \frac{4|\Lambda|_*}{n}|\widehat{\sigma} - \sigma_Q| + 2\delta P_n(\lambda_0).\end{aligned} \tag{3.25}$$

To estimate the second term in the right side of this inequality we set

$$S_x = \sum_{j=1}^{n} x(j)\overline{\theta}_{j,p}\mathrm{Tr}_j\,, \quad x = (x(j))_{1\le j\le n} \in \mathbb{R}^n\,.$$

Thanks to (2.7) we estimate the term $M(x)$ for any $x \in \mathbb{R}^n$ as

$$\mathbf{E}_Q M^2(x) \le \varkappa_Q \frac{1}{n}\sum_{j=1}^{n} x^2(j)\overline{\theta}^2_{j,p} = \varkappa_Q \frac{1}{n}\|S_x\|^2. \tag{3.26}$$

To estimate this function for a random vector $x \in \mathbb{R}^n$, we set

$$Z^* = \sup_{x\varepsilon\Lambda_1} \frac{nM^2(x)}{\|S_x\|^2}\,, \quad \Lambda_1 = \Lambda - \lambda_0\,.$$

So, through Inequality (3.24), we get

$$2|M(x)| \le \delta\|S_x\|^2 + \frac{Z^*}{n\delta}. \tag{3.27}$$

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

$$\mathbf{E}_Q Z^* \le \sum_{x\in\Lambda_1} \frac{n\mathbf{E}_Q M^2(x)}{\|S_x\|^2} \le \sum_{x\in\Lambda_1} \varkappa_Q = \varkappa_Q \check{\iota}\,, \tag{3.28}$$

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

$$\|S_x\|^2 - \|\widehat{S}_x\|^2 = \sum_{j=1}^{n} x^2(j)(\overline{\theta}^2_{j,p} - \widehat{\theta}^2_j) \le -2M_1(x), \tag{3.29}$$

where $M_1(x) = n^{-1/2}\sum_{j=1}^{n} x^2(j)\overline{\theta}^2_{j,p}\xi_{j,n}$. Taking into account now that, for any $x \in \Lambda_1$, the components $|x(j)| \le 1$, we can estimate this term as in (3.26), i.e.

$$\mathbf{E}_Q\, M_1^2(x) \le \varkappa_Q\, \frac{\|S_x\|^2}{n}\,.$$

Similarly to the previous reasoning we set

$$Z_1^* = \sup_{x\varepsilon\Lambda_1} \frac{nM_1^2(x)}{\|S_x\|^2}$$

and we get

$$\mathbf{E}_Q\, Z_1^* \le \varkappa_Q\, \check{\iota}\,. \tag{3.30}$$

Using the same type of arguments as in (3.27), we can derive

$$2|M_1(x)| \le \delta\|S_x\|^2 + \frac{Z_1^*}{n\delta}. \tag{3.31}$$

From here and (3.29), we get

$$\|S_x\|^2 \le \frac{\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z_1^*}{n\delta(1-\delta)} \tag{3.32}$$

for any $0 < \delta < 1$. Using this bound in (3.27) yields

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

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

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

Using this bound in (3.25) we obtain

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{Z^* + Z_1^*}{n\delta(1-3\delta)} + \frac{2\mathbf{L}_{1,Q}}{n(1-3\delta)} + \frac{2B_{2,Q}^*}{\delta(1-3\delta)\sigma_Q n}\\ &+ \frac{(4|\Lambda|_* + 2)}{n(1-3\delta)}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).\end{aligned}$$

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

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{2(Z^* + Z_1^*)}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4B_{2,Q}^*}{\delta\sigma_Q n}\\ &+ \frac{(8|\Lambda|_* + 2)}{n}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}\,P_n^0(\lambda_0).\end{aligned}$$

Now, in view of Condition $\mathbf{L}_2$), we estimate the expectation of the term $B_{2,Q}^*$ in (3.25) as

$$\mathbf{E}_Q\, B_{2,Q}^* \le \sum_{\lambda\in\Lambda} \mathbf{E}_Q B_{2,Q}^2(e(\lambda)) \le \imath\mathbf{L}_{2,Q}\,.$$

Now, taking into account that $|\Lambda|_* \ge 1$, we get

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \le& \frac{1+\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\imath}}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4\imath\mathbf{L}_{2,Q}}{\delta\sigma_Q n}\\ &+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).\end{aligned}$$

By using the upper bound for $P_n(\lambda_0)$ in Lemma 3.1, we obtain that

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \le& \frac{1+3\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\imath}}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4\imath\mathbf{L}_{2,Q}}{\delta\sigma_Q n}\\ &+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta\mathbf{L}_{1,Q}}{(1-3\delta)n}\,.\end{aligned}$$

Taking into account that $1 - 3\delta \ge 1/2$ for $0 < \delta < 1/3$ and that $\kappa_Q \le (1 + \check{\tau}|\rho|_*)\sigma_Q$ and using the bounds (3.38) and (3.60) we obtain the inequality (3.19). Hence we get the desired result. □

Corollary 3.1. *Assume that Conditions* $\mathbf{L}_1$*) and* $\mathbf{L}_2$*) hold and the proxy variance* σ_Q *is known. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, for any weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17) *with* $\widehat{\sigma}_n = \sigma_Q$*, satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + l^* \frac{\sigma_Q \check{\iota}}{\delta n} \,. \tag{3.33}$$

Now we study the model selection procedure (3.17) using the proxy estimate (3.13).

Theorem 3.2. *Assume that the function* S *is continuously differentiable and the conditions* $\mathbf{L}_1$*) and* $\mathbf{L}_2$*) hold true. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, for any weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17)*, satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + l^* \frac{\nu}{\delta n} (1+\sigma_Q)^3 \left(1 + \|\dot{S}\|^2\right) \mathbf{g}_{n,p} \,. \end{aligned} \tag{3.34}$$

Let us study the robust risks (3.4) for the procedure (3.17). In this case this family consists of all distributions on the Skorokhod space $\mathcal{D}[0,n]$ with the parameters satisfying the conditions (2.8) – (2.9) . Now, to obtain the efficiency property we use the weight coefficients $(\lambda(j))_{1\leq j\leq n}$ specified in (2.28).

Our goal is to bound asymptotically the term (3.18) by any power of n. To this end we assume the following condition.

$\mathbf{H}_5$) *Assume that there exists* $\check{\delta} > 0$ *such that for any* $n \geq 3$

$$p \geq n^{5/6} \,. \tag{3.35}$$

Now Theorem 3.2 implies the following oracle inequality.

Theorem 3.3. *Assume that the unknown function* S *is continuously differentiable. Moreover, assume that Conditions* $\mathbf{H}_1$*)–*$\mathbf{H}_5$*) hold. Then for the robust risks defined in* (3.4) *through the distribution family* (2.8) – (2.9)*, the procedure* (3.17) *with the coefficients* (2.28)*, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{U}_n^*(S)}{n\delta} \,, \tag{3.36}$$

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that under condition* (2.26)*, for any* $r > 0$ *and* $\check{\delta} > 0$*,*

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\leq r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0 \,. \tag{3.37}$$

Now we need to check the conditions $\mathbf{L}_1$) and $\mathbf{L}_2$) for the process (2.2).

Proposition 3.1. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then*

$$\mathbf{L}_{1,Q} = 2\check{\tau}\,\|\Upsilon\|_1\,\sigma_Q\,. \tag{3.38}$$

Proof. Firstly, we set

$$I_n^L(f) = \int_0^n f(t)\mathrm{d}L_t \quad \text{and} \quad I_n^z(f) = \int_0^n f(t)\mathrm{d}z_t\,. \tag{3.39}$$

In view of (1.6) the last integral can be represented as

$$I_n^z(f) = \sum_{l=1}^{\infty} f(T_l)Y_l\mathbf{1}_{\{T_l\le n\}}\,. \tag{3.40}$$

Therefore,

$$\xi_{j,n} = \frac{\varrho_1}{\sqrt{n}}I_n^L(\Psi_{j,p}) + \frac{\varrho_2}{\sqrt{n}}I_n^z(\Psi_{j,p})$$

and

$$\mathbf{E}\xi_{j,n}^2 = \frac{\varrho_1^2}{n}\int_0^n \Psi_{j,p}^2(t)\mathrm{d}\,t + \frac{\varrho_2^2}{n}\mathbf{E}\sum_{l=1}^{\infty}\Psi_{j,p}^2(T_l)1_{\{T_l\le n\}}\,. \tag{3.41}$$

Using Proposition 5.1 we get

$$\begin{aligned}\mathbf{E}\sum_{l=1}^{\infty}\Psi_{j,p}^2(T_l)1_{\{T_l\le n\}} &= \int_0^n \Psi_{j,p}^2(x)\,\rho(x)\mathrm{d}x\\ &= \frac{1}{\check{\tau}}\int_0^n \Psi_{j,p}^2(x)\mathrm{d}x + \int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x\,,\end{aligned}$$

where ρ is the renewal density introduced in (1.8). Then we obtain,

$$\mathbf{E}\xi_{j,n}^2 = \sigma_Q + \frac{\varrho_2^2}{n}\int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x$$

and

$$\sup_{j\ge1}\left|\int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x\right| \le 2\|\Upsilon\|_1\,, \tag{3.42}$$

where $\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau}$. This directly implies the desired result. □

To study the function $B_{2,Q}(x)$, we have to analyze the correlation properties for the following stochastic integrals

$$\widetilde{I}_n(f) = I_n^2(f) - \mathbf{E}I_n^2(f)\,. \tag{3.43}$$

To do this we set

$$\check{c}_1 = 1 + \Pi(x^4) + \|\Upsilon\|_1^2 + |\rho|_* \quad \text{and} \quad \check{c}_2 = 12(1+\check{\tau})^2\,(1+\check{c}_1)\,. \tag{3.44}$$

Now we investigate the behavior of the integrals defined in (3.43) as functions of f.

Proposition 3.2. *For any left continuous functions $f, g : (0, \infty) \longrightarrow R$ such that $\|f\|_* \leq 1$, $\|g\|_* \leq 1$, we have*

$$|\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g)| \leq 12\sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\check{c}_1\right). \tag{3.45}$$

Proof. By Ito's formula one gets

$$\mathrm{d}I_t^2(f) = 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\check{\varrho}^2 f^2(t)\mathrm{d}\,t + \sum_{0\leq s\leq t} f^2(s)(\Delta\xi_s^d)^2\,, \tag{3.46}$$

where $\xi_t^d = \varrho_3\check{L}_t + \varrho_2 z_t$ and $\varrho_3 = \varrho_1\sqrt{1-\check{\varrho}^2}$. Taking into account that the processes $(\check{L}_t)_{t\geq 0}$ and $(z_t)_{t\geq 0}$ are independent and the time of jumps T_k defined in (1.7) has a density, we have $\Delta z_s \Delta\check{L}_s = 0$ a.s. for any $s \geq 0$. Therefore, we can rewrite the differential (3.46) as

$$\begin{aligned}\mathrm{d}I_t^2(f) =& 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\check{\varrho}^2 f^2(t)\mathrm{d}\,t \\ &+ \varrho_3^2\mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta\check{L}_s)^2 + \varrho_2^2\mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta z_s)^2\,.\end{aligned} \tag{3.47}$$

Therefore, using Lemma 2.1 we obtain

$$\mathbf{E}I_t^2(f) = \varrho_1^2\,\|f\|_t^2 + \varrho_2^2\|f\sqrt{\rho}\|_t^2\,,$$

where $\|f\|_t^2 = \int_0^t f^2(u)\mathrm{d}u$, ρ is the density of the renewal measure $\sum_{j=1}^{\infty}\eta^{(j)}$ and with η the distribution of τ_1. Therefore,

$$\mathrm{d}\widetilde{I}_t(f) = 2I_{t-}(f)f(t)d\xi_t + f^2(t)\mathrm{d}\widetilde{m}_t\,, \quad \widetilde{m}_t = \varrho_3^2\check{m}_t + \varrho_2^2 m_t\,, \tag{3.48}$$

where $\check{m}_t = \sum_{0\leq s\leq t}(\Delta\check{L}_s)^2 - t$ and $m_t = \sum_{0\leq s\leq t}(\Delta z_s)^2 - \int_0^t\rho(s)\mathrm{d}s$. By the Ito's formula we get

$$\begin{aligned}\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g) =& \mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g) \\ &+ \mathbf{E}\int_0^n \widetilde{I}_{t-}(g)\mathrm{d}\widetilde{I}_t(f) + \mathbf{E}\left[\widetilde{I}(f),\widetilde{I}(g)\right]_n.\end{aligned} \tag{3.49}$$

First, note that the process $(\check{m}_t)_{t\geq 0}$ is a martingale and, using Lemma 2.5, we get

$$\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g) = \rho_2^2\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)g^2(t)\mathrm{d}m_t = \rho_2^2\mathbf{E}\int_0^n I_{t-}^2(f)g^2(t)\mathrm{d}m_t\,.$$

The last integral can be represented as

$$\mathbf{E}\int_0^n I_{t-}^2(f)g^2(t)\mathrm{d}m_t = J_1 - J_2\,,$$

where

$$J_1 = \mathbf{E}\sum_{k\geq 1} I_{T_k-}^2(f)g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad J_2 = \int_0^n \mathbf{E}\,I_t^2(f)g^2(t)\rho(t)\mathrm{d}t\,.$$

By Lemma 2.3 we get

$$J_1 = \mathbf{E}\sum_{k\geq 1} \mathbf{E}\left(I^2_{T_k-}(f)|\mathcal{G}\right) g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} = \varrho_1^2 J_{1,1} + \varrho_2^2 J_{1,2}\,,$$

where

$$J_{1,1} = \mathbf{E}\sum_{k\geq 1} \|f\|^2_{T_k} g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad J_{1,2} = \mathbf{E}\sum_{k\geq 1}\sum_{l=1}^{k-1} f^2(T_l)\, g^2(T_k)\mathbf{1}_{\{T_k\leq n\}}\,.$$

We obtain directly that

$$J_{1,1} = \int_0^n \|f\|^2_t g^2(t)\rho(t)\mathrm{d}t$$

and

$$J_{1,2} = \mathbf{E}\sum_{l\geq 1} f^2(T_l)\sum_{k\geq l+1} g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} = \int_0^n f^2(x)\left(\int_0^{n-x} g^2(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x.$$

From Lemma 2.1 we obtain that

$$J_2 = \varrho_1^2\int_0^n \|f\|^2_t g^2(t)\rho(t)\mathrm{d}t + \varrho_2^2\int_0^n \|f\sqrt{\rho}\|^2_t g^2(t)\rho(t)\mathrm{d}t\,.$$

Therefore,

$$\mathbf{E}\int_0^n I^2_{t-}(f)g^2(t)\mathrm{d}m_t = \varrho_2^2\int_0^n f^2(x)\left(\int_x^n g^2(t)(\rho(t-x)-\rho(t))\mathrm{d}t\right)\rho(x)\mathrm{d}x\,.$$

Taking into account that $\rho(t-x)-\rho(t) = \Upsilon(t-x)-\Upsilon(t)$ we can estimate the last integral as

$$|\mathbf{E}\int_0^n I^2_{t-}(f)g^2(t)\mathrm{d}m_t| \leq 2\varrho_2^2 n\|\Upsilon\|_1\,.$$

From this and by the symmetry arguments we obtain that

$$|\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g)| + |\mathbf{E}\int_0^n \widetilde{I}_{t-}(g)\mathrm{d}\widetilde{I}_t(f)| \leq 4\varrho_2^4 n\|\Upsilon\|_1\,. \tag{3.50}$$

Note now that

$$\left[\widetilde{I}(f),\widetilde{I}(g)\right]_n = \left\langle \widetilde{I}^c(f),\widetilde{I}^c(g)\right\rangle_n + \mathbf{D}_n(f,g)\,, \tag{3.51}$$

where

$$\mathbf{D}_n(f,g) = \sum_{0\leq t\leq n}\Delta\widetilde{I}^d_t(f)\Delta\widetilde{I}^d_t(g)\,.$$

It should be noted that the continuous and the discrete parts of the processes (3.48) can be represented as

$$\widetilde{I}^c_t(f) = 2\varrho_1\check{\varrho}\int_0^t I_s(f)f(s)\mathrm{d}w_s \quad \text{and} \quad \widetilde{I}^d_t(g) = 2\int_0^t I_{s-}(f)f(s)d\xi^d_s + \int_0^t f^2(s)\mathrm{d}\widetilde{m}_s\,.$$

So, in view of Lemma 2.1,

$$\begin{aligned}\mathbf{E} < \widetilde{I}^c(f), \widetilde{I}^c(g) >_n &= 4\rho_1^2 \breve{\varrho}^2 \int_0^n \mathbf{E}(I_t(f) I_t(g)) f(t) g(t) \mathrm{d}t \\ &= 4\rho_1^4 \breve{\varrho}^2 \int_0^n (f,g)_t \, f(t) g(t) \mathrm{d}t + 4\rho_1^2 \rho_2^2 \breve{\varrho}^2 \int_0^n (f, g\rho)_t f(t) g(t) \mathrm{d}t \\ &= 4\rho_1^2 \breve{\varrho}^2 \sigma_Q \, (f,g)_n^2 + 4\rho_1^2 \rho_2^2 \breve{\varrho}^2 \int_0^n (f, g\Upsilon)_t f(t) g(t) \mathrm{d}t \,, \end{aligned} \tag{3.52}$$

with $(f,g)_t = \int_0^t f(s) g(s) \mathrm{d}s$. Taking into account that $\|f\|_* \le 1$ and $\|g\|_* \le 1$, we can estimate the last integral as

$$\int_0^n (f, g\Upsilon)_t f(t) g(t) \mathrm{d}t \le n \|\Upsilon\|_1 \,.$$

Therefore,

$$\left| \mathbf{E} \left\langle \widetilde{I}^c(f), \widetilde{I}^c(g) \right\rangle_n \right| \le 4\sigma_Q^2 \left((f,g)_n^2 + n\check{\tau} \|\Upsilon\|_1 \right) . \tag{3.53}$$

To study the last term in (3.51) note that

$$\mathbf{D}_n(f,g) = \sum_{0 \le t \le n} \left(2 I_{t-}(f) f(t) \Delta \xi_t^d + f^2(t) \Delta \widetilde{m}_t \right) \left(2 I_{t-}(g) g(t) \Delta \xi_t^d + g^2(t) \Delta \widetilde{m}_t \right) .$$

Taking into account that for any $t > 0$

$$\Delta \xi_t^d \Delta \widetilde{m}_t = \varrho_3^3 (\Delta \check{L}_t)^3 + \varrho_2^3 (\Delta z_t)^3 \,,$$

we obtain that

$$\mathbf{E} \sum_{0 \le t \le n} I_{t-}(f) f(t) g^2(t) \Delta \xi_t^d \, \Delta \widetilde{m}_t = \left(\varrho_3^3 \Pi(x^3) \, + \varrho_2^3 \mathbf{E} \, Y_1^3 \right) \int_0^n \mathbf{E} \, I_t(f) f(t) g^2(t) \mathrm{d}t = 0 \,.$$

So, using the symmetry arguments, we find that

$$\mathbf{E} \mathbf{D}_n(f,g) = 4 \mathbf{E} \, \mathbf{D}_{1,n}(f,g) + \mathbf{E} \, \mathbf{D}_{2,n}(f,g) \,, \tag{3.54}$$

where

$$\mathbf{D}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta \xi_t^d)^2 \quad \text{and} \quad \mathbf{D}_{2,n}(f,g) = \sum_{0 \le t \le n} f^2(t) \, g^2(t) (\Delta \widetilde{m}_t)^2 \,.$$

Note that

$$\mathbf{D}_{1,n}(f,g) = \varrho_3^2 \check{\mathbf{D}}_{1,n}(f,g) + \varrho_2^2 \widetilde{\mathbf{D}}_{1,n}(f,g) \,,$$

where

$$\check{\mathbf{D}}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta \check{L}_t)^2$$

and

$$\widetilde{\mathbf{D}}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta z_t)^2 \,.$$

Now, similarly to (3.52) and taking into account that $\Pi(x^2)=1$, we get

$$\begin{aligned}\mathbf{E}\,\check{\mathbf{D}}_{1,n}(f,g) &= \int_0^n f(t)g(t)\mathbf{E}\,I_t(f)I_t(g)\,\mathrm{d}t = \varrho_1^2\int_0^n f(t)g(t)\,(f,g)_t\,\mathrm{d}t \\ &\quad + \varrho_2^2\int_0^n f(t)g(t)\,(f,g\rho)_t\,\mathrm{d}t \\ &= \sigma_Q(f,g)_n^2 + \varrho_2^2\int_0^n f(t)g(t)\,(f,g\Upsilon)_t\,\mathrm{d}t\,.\end{aligned}$$

So,

$$|\mathbf{E}\,\check{\mathbf{D}}_{1,n}(f,g)| \le \sigma_Q\left((f,g)_n^2 + n\check{\tau}\|\Upsilon\|_1\right)\,. \tag{3.55}$$

Moreover, taking into account that $\mathbf{E}Y_1^2=1$ we get

$$\mathbf{E}\,\widetilde{\mathbf{D}}_{1,n}(f,g) = \mathbf{E}\sum_{k\ge1} I_{T_k-}(f)I_{T_k-}(g)f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}}\,.$$

So, in view of Lemma 2.3

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

where

$$\mathbf{D}_{1,n}^{'}(f,g) = \sum_{k\ge1}\sum_{l=1}^{k-1} f(T_l)\,g(T_l)f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}}\,.$$

Noting now that

$$\int_0^n (f,g)_t\,f(t)g(t)\rho(t)\mathrm{d}t = \frac{1}{2\check{\tau}}(f,g)_n^2 + \int_0^n (f,g)_t\,f(t)g(t)\Upsilon(t)\mathrm{d}t\,,$$

we obtain

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

Furthermore, the expectation of $\mathbf{D}_{1,n}^{'}(f,g)$ can be represented as

$$\begin{aligned}\mathbf{E}\,\mathbf{D}_{1,n}^{'}(f,g) &= \mathbf{E}\sum_{l\ge1} f(T_l)\,g(T_l)\sum_{k\ge l+1} f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}} \\ &= \int_0^n f(x)g(x)\left(\int_0^{n-x} f(x+t)g(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x \\ &= \frac{1}{2\check{\tau}}(f,g)_n^2 + \mathbf{D}_{1,n}^{''}(f,g)\,,\end{aligned}$$

where the last term in this equality can be represented as

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

This implies

$$|\mathbf{D}^{''}_{1,n}(f,g)| \le n(1+\frac{1}{\check{\tau}})(1+\|\Upsilon\|_1^2)\,.$$

Therefore,

$$|\mathbf{E}\,\widetilde{\mathbf{D}}_{1,n}(f,g)| \le \sigma_Q\left((f,g)_n^2 + n(1+\check{\tau})\|\Upsilon\|_1^2\right)\,. \tag{3.56}$$

Finally we obtain that

$$|\mathbf{E}\,\mathbf{D}_{1,n}(f,g)| \le \sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\|\Upsilon\|_1^2\right)\,. \tag{3.57}$$

As to the last term in (3.54) we can calculate directly

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

i.e.

$$\mathbf{E}\,\mathbf{D}_{2,n}(f,g) \le n\sigma_Q^2\left(\Pi(x^4)+|\rho|_*\right)(1+\check{\tau})^2\,.$$

From here we obtain that

$$|\mathbf{E}\,\mathbf{D}_n(f,g)| \le \sigma_Q^2(1+\check{\tau})^2\left(4(f,g)_n^2 + n\check{c}_1\right)\,, \tag{3.58}$$

where $\check{c}_1$ is given in (3.44). From this and (3.53) we find

$$\mathbf{E}[\widetilde{I}(f),\widetilde{I}(g)]_n \le 8\sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\check{c}_1\right)\,. \tag{3.59}$$

This bound and (3.50) implies (3.45). Hence we get the desired result. □

Using these properties we can obtain the following bound.

Proposition 3.3. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for all* $n \ge 1$,

$$\mathbf{L}_{2,Q} = \check{c}_2\,\sigma_Q^2\,, \tag{3.60}$$

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

Proof. Note that

$$\mathbf{E}\left(\sum_{j=2}^{n} x_j \widetilde{\xi}_{j,p}\right)^2 \le \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n} |x_j|\,|x_l|\,|\mathbf{E}\widetilde{I}_n(\Psi_{j,p})\widetilde{I}_n(\Psi_{l,p})|\,.$$

Using here Proposition 3.2 and taking into account that

$$(\Psi_{j,p}\,,\,\Psi_{l,p})_n = \int_0^n \Psi_{j,p}(t)\Psi_{l,p}(t)\mathrm{d}t = n\mathbf{1}_{\{j=l\}}\,,$$

we obtain the bound (3.60). Hence we obtain the desired result. □

Now we can study the estimate (3.17).

Proposition 3.4. *Assume that Conditions* $\mathbf{H}_1$*) and* $\mathbf{H}_4$*) hold true for the model* (3.1) *and that* S *is continuously differentiable. Then, for any* $n \ge 2$ *and* $p \ge 3$*,*

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \sigma_Q| \le \check{c}_3\left(\frac{\sqrt{n}}{\check{p}} + \frac{1}{\sqrt{\check{p}}}\right)(1+\|\dot{S}\|^2)(1+\sigma_Q)^2\,, \tag{3.61}$$

where $\check{c}_3 = 6\,(14 + 2|\rho|_* + 3\sqrt{1+\check{c}_1})\,(1+\check{\tau})$.

Proof. It is clear that Inequality (3.61) holds true for $l > \check{p}$. Let now $l \leqslant \check{p}$. Setting $x_j^{'} = \mathbf{1}_{\{[\sqrt{n}]\leqslant j\leqslant\check{p}\}}$ and subtituting (3.7) in (3.13) yields

$$\widehat{\sigma}_n = \frac{n}{\check{p}}\sum_{j=l}^{\check{p}}(\overline{\theta}_{j,p})^2 + \frac{2n}{\check{p}}\,M(x^{'}) + \frac{1}{\check{p}}\sum_{j=l}^{\check{p}}\xi_{j,p}^2\,, \tag{3.62}$$

where $M(x^{'})$ is defined in (3.21). Furthermore, putting $x_j^{''} = \check{p}^{-1/2}\mathbf{1}_{\{l\leqslant j\leqslant\check{p}\}}$, one can write the last term on the right hand side of (3.62) as

$$\frac{1}{\check{p}}\sum_{j=l}^{\check{p}}\xi_{j,p}^2 = \frac{1}{\sqrt{\check{p}}}\,B_{2,Q}(x^{''}) + \frac{1}{\check{p}}B_{1,Q}(x^{'}) + \frac{(\check{p}-l+1)\sigma_Q}{\check{p}}\,,$$

where the functions $B_{1,Q}$ and $B_{2,Q}$ are given in conditions $\mathbf{L}_1$) and $\mathbf{L}_2$). Using Proposition 3.1, Proposition 3.3 and Lemma 3.3 , we come to the following upper bound

$$\mathbf{E}_Q|\widehat{\sigma}_n - \sigma_Q| \le \frac{16\|\dot{S}\|^2 n}{lp} + \frac{2n}{p}\mathbf{E}_Q\,|\mathbf{M}(x^{'})| + \frac{\mathbf{L}_{1,Q}}{p} + \frac{\sqrt{\mathbf{L}_{2,Q}}}{\sqrt{p}} + \frac{\sigma_Q(l-1)}{p}.$$

In the same way as in (3.26), we obtain

$$\mathbf{E}_Q\,|M(x^{'})| \le \left(\frac{\varkappa_Q}{n}\sum_{j=l}^{p}\overline{\theta}_{j,p}^2\right)^{1/2} \le \frac{4(\varkappa_Q\|\dot{S}\|^2)^{1/2}}{l}.$$

Taking into account that $\kappa_Q \le (1+\check{\tau}|\rho|_*)\sigma_Q$ and using the bounds (3.38) and (3.60) we obtain the inequality (3.61). Hence we obtain the desired result. □

Remark 3.1. Propositions 3.1 and 3.3 are used to obtain the oracle inequalities given in Section 4.4 (see, for example, [17]).

3.4 Efficiency

Now we study the asymptotically efficiency properties for the procedure (3.17) with the coefficients (2.28), with respect to the robust risk (3.4) defined by the distribution family (2.8) – (2.9). To this end, we assume that the unknown function S in the model (3.1) belongs to the Sobolev ball

$$W_{\mathbf{r}}^{k} = \{ f \in \mathcal{C}_{per}^{k}[0,1] \, , \sum_{j=0}^{k} \| f^{(j)} \|^2 \leq \mathbf{r} \} \, , \tag{3.63}$$

where $r > 0 \, , \; k \geq 1$ are some parameters, $\mathcal{C}_{per}^{k}[0,1]$ 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 $0 \leq i \leq k$. The function class $W_{\mathbf{r}}^{k}$ can be written as an ellipsoid in $\mathbf{L}_2$, i.e.

$$W_{\mathbf{r}}^{k} = \{ f \in \mathcal{C}_{per}^{k}[0,1] \, : \, \sum_{j=1}^{\infty} a_j \, \theta_j^2 \, \leq \mathbf{r} \}, \tag{3.64}$$

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

Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (3.4) is given by

$$\mathbf{r}_k^* = \, ((2k+1)\mathbf{r})^{1/(2k+1)} \, \left(\frac{k}{(k+1)\pi} \right)^{2k/(2k+1)} \, . \tag{3.65}$$

Note that this is the well-known Pinsker's constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]).

Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the σ-algebra $\sigma\{ y_t \, , \, 0 \leq t \leq n \}$ generated by the process (3.1).

Theorem 3.4. *Under the conditions* (2.8) *and* (2.9)*,*

$$\liminf_{n \to \infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S \in W_{\mathbf{r}}^{k}} \mathcal{R}_n^*(\widehat{S}_n, S) \geq \mathbf{r}_k^* \, , \tag{3.66}$$

where $\upsilon_n = n/\varsigma^*$.

Note that if the parameters r and k are known, i.e. for the non-adaptive estimation case, then to obtain the efficient estimation for the "signal+white noise"model. Pinsker in [47] proposed to use the estimate $\widehat{S}_{\lambda_0}$ defined in (3.8) with the weights (2.28) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, l_0) \, , \tag{3.67}$$

where $l_0 = [\mathbf{r}/\varepsilon]\varepsilon$. For the model (3.1) – (2.2) we show the same result.

Proposition 3.5. *The estimator* $\widehat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n \to \infty} \upsilon_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^{k}} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^* \, .$$

Proof. First, we note that in view of (3.8) one can represent the quadratic risk for the empiric norm $\|\cdot\|_p$ as

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

where $\overline{\Theta}_p = \sum_{j=1}^{p}\left(\theta_{j,p} - \lambda_0(j)\,\overline{\theta}_{j,p}\right)^2$. We put here $\lambda_0(j) = 0$ for $j > n$ if $p > n$. The first term can be estimated by the bound (3.38) as

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

where $\mathbf{L}_{1,n}^* = \sup_{Q\in\mathcal{Q}_n} \mathbf{L}_{1,Q}$. Therefore, taking into account that $\upsilon_n = n/\sigma^*$, we get

$$\sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q \|\widehat{S}_{\lambda_0} - S\|_p^2 \le \frac{1}{\upsilon_n}\sum_{j=1}^{n} \lambda_0^2(j) + \frac{\mathbf{L}_{1,n}^*}{n} + \overline{\Theta}_p\,.$$

Note that

$$\lim_{n\to\infty} \frac{1}{\upsilon_n^{1/(2k+1)}} \sum_{j=1}^{n} \lambda_0^2(j) = \frac{2(\tau_k\,\mathbf{r})^{1/(2k+1)}\,k^2}{(k+1)(2k+1)}\,. \tag{3.68}$$

Furthermore, by Inequality (3.24) for any $0 < \widetilde{\varepsilon} < 1$ we get

$$\overline{\Theta}_p \le (1+\widetilde{\varepsilon})\,\Theta_p + (1+\widetilde{\varepsilon}^{-1}) \sum_{j=1}^{p} h_{j,p}^2\,, \tag{3.69}$$

where $\Theta_p = \sum_{j=1}^{p}(1-\lambda_0(j))^2\,\theta_{j,p}^2$. In view of Definition (2.28), we can represent this term as

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

where $\iota_0 = j_*(\alpha_0)$, $\omega_0 = \omega_{\alpha_0} = (\tau_k l_0 \upsilon_n)^{1/(2k+1)}$ and $l_0 = [\mathbf{r}/\varepsilon]\,\varepsilon$. Applying Lemma 3.5 yields

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

Similarly, through Lemma 3.4 we have

$$\Theta_{2,p} \le (1+\widetilde{\varepsilon}) \sum_{j\ge[\omega_0]+1} \theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,\mathbf{r}\,p^{-2}\,.$$

Hence,

$$\Theta_p \le (1+\widetilde{\varepsilon})\,\Theta_{\iota_0}^* + (1+\widetilde{\varepsilon}^{-1})\left(4\pi^2 r\omega_0^3 + \mathbf{r}\right) p^{-2}\,,$$

where $\Theta_l^* = \sum_{j\geq l} (1-\lambda_0(j))^2\, \theta_j^2$. Moreover, note that

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

Moreover, $W_{\mathbf{r}}^k \subseteq W_{\mathbf{r}}^2$ for any $k\geq 2$. From here and Lemma 3.6 we get

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

Moreover, in view of Condition $\mathbf{H}_5$) we have

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

So,

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \overline{\Theta}_p \leq \limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \Theta_{\iota_0}^*\,.$$

To estimate the term $\Theta_{\iota_0}^*$ we set

$$\mathbf{U}_n = \upsilon_n^{2k/(2k+1)} \sup_{j\geq \iota_0} (1-\lambda_0(j))^2/a_j\,,$$

where the sequence $(a_j)_{j\geq 1}$ is defined in (3.64). This leads to the inequality

$$\sup_{S\in W_{\mathbf{r}}^1} \upsilon_n^{2k/(2k+1)}\, \Theta_{\iota_0}^* \leq \mathbf{U}_n \sum_{j\geq 1} a_j\, \theta_j^2 \leq \mathbf{U}_n\, \mathbf{r}\,.$$

Taking into account that $\lim_{n\to\infty} t_0 = \mathbf{r}$, we get

$$\limsup_{n\to\infty} \mathbf{U}_n \leq \pi^{-2k}\, (\tau_k\, \mathbf{r})^{-2k/(2k+1)}\,,$$

where the coefficient τ_k is given in (2.28). This implies immediately that

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_r^k} \overline{\Theta}_p \leq \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2k/(2k+1)}}\,. \tag{3.70}$$

Moreover, note that

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

So, applying (3.68) and (3.70), yields

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q\, \|\widehat{S}_{\lambda_0} - S\|_p^2 \leq R_k^*\,. \tag{3.71}$$

Furthermore, Lemma 3.2 yields that for any $\widetilde{\varepsilon} > 0$

$$\sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq (1+\widetilde{\varepsilon}) \sup_{S\in W_{\mathbf{r}}^k} \sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q\, \|\widehat{S}_{\lambda_0} - S\|_p^2 + (1+\widetilde{\varepsilon}^{-1})\mathbf{r}\, p^{-2}\,.$$

So, in view of Condition $\mathbf{H}_5$), we derive the desired inequality

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \le R_k^* \,.$$

Hence the conclusion follows. □

For the adaptive estimation we use the model selection procedure (3.17) with the parameter δ defined as a function of n satisfying

$$\lim_{n\longrightarrow\infty} \delta_n = 0 \quad \text{and} \quad \lim_{n\longrightarrow\infty} n^{\check{\delta}}\, \delta_n = 0 \tag{3.72}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$.

Theorem 3.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold. Then the robust risk defined in* (3.4) *through the distribution family* (2.8) – (2.9) *for the procedure* (3.17) *with the coefficients* (2.28) *and the parameter* $\delta = \delta_n$ *satisfying* (3.72) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le \mathbf{r}_k^* \,. \tag{3.73}$$

Theorem 3.4 and Theorem 3.5 imply the following result.

Corollary 3.2. *Under the conditions of Theorem 3.5,*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^* \,. \tag{3.74}$$

Remark 3.2. It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball $W_{\mathbf{r}}^k$ is $n^{2k/(2k+1)}$ (see, for example, [47], [46]). We see here that the efficient robust rate is $\upsilon_n^{2k/(2k+1)}$, i.e. if the distribution upper bound $\varsigma^* \to 0$ as $n \to \infty$ we obtain a faster rate with respect to $n^{2k/(2k+1)}$, and if $\varsigma^* \to \infty$ as $n \to \infty$ we obtain a slower rate. In the case when ς^* is constant the robust rate is the same as the classical non robuste convergence rate.

3.5 Simulations

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (3.17). In (3.1) we chose a 1-periodic function which for $0 \le t \le 1$ is defined as

$$S(t) = \begin{cases} |t - \frac{1}{2}| & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4}\,, \\ \\ \frac{1}{4} & \text{else.} \end{cases} \tag{3.75}$$

We simulate the model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t \,,$$

n	$\overline{\mathbf{R}}$	$\overline{\mathbf{R}_*}$
20	0.0398	0.211
100	0.0091	0.0483
200	0.0067	0.0355
1000	0.0022	0.0116

Table 3.1: Empirical risks

where $\xi_t = 0.5\mathrm{d}wt + 0.5\mathrm{d}z_t$. Here z_t is the semi-Markov process defined in (1.6) with a Gaussian $\mathcal{N}(0,1)$ sequence $(Y_j)_{j\geq 1}$ and $(\tau_k)_{k\geq 1}$ used in (1.7) taken as $\tau_k \sim \chi^2_3$.

We use the model selection procedure (3.17) with the weights (2.28) in which $k^* = 100 + \sqrt{(}\ln(n))$, $t_i = i/\ln(n)$, $m = [\ln^2(n)]$ and $\delta = (3+\ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathbf{R}} = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\left(\hat{S}_n(t_j) - S(t_j)\right)^2 , \tag{3.76}$$

where the observation frequency $p = 100001$ and the expectations was taken as an average over $N = 10000$ replications, i.e.

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

We set the relative quadratic risk as

$$\overline{\mathbf{R}_*} = \overline{\mathbf{R}}/\|S\|_p^2 \quad \text{and} \quad \|S\|_p^2 = \frac{1}{p}\sum_{j=0}^{p} S^2(t_j)\,. \tag{3.77}$$

In our case $\|S\|_p^2 = 0.1883601$.

Table 3.1 gives the values for the sample risks (3.76) and (3.77) for different numbers of observations n.

The Figures 3.1–3.4 show the behavior of the regression function and its estimates by the model selection procedure (3.17) depending on the values of observation periods n. The black full line is the regression function (3.75) and the red dotted line is the associated estimator.

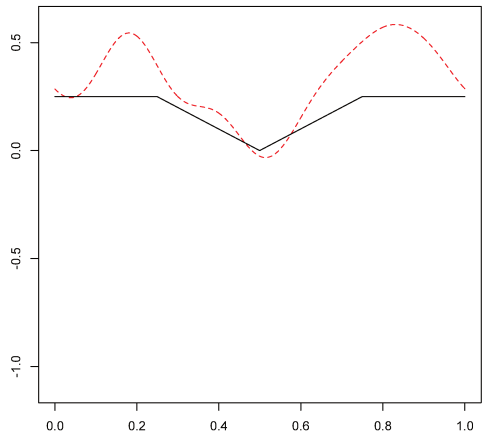

Figure 3.1: Estimator of S for n=20.

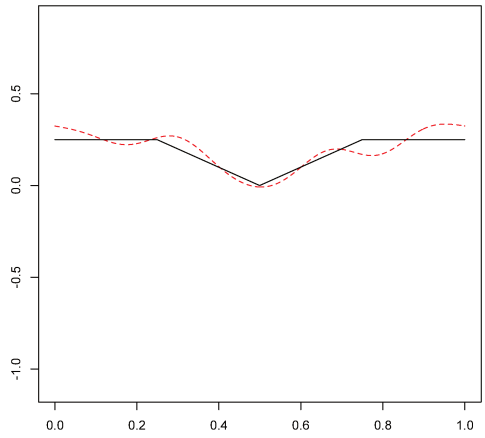

Figure 3.2: Estimator of S for n=100.

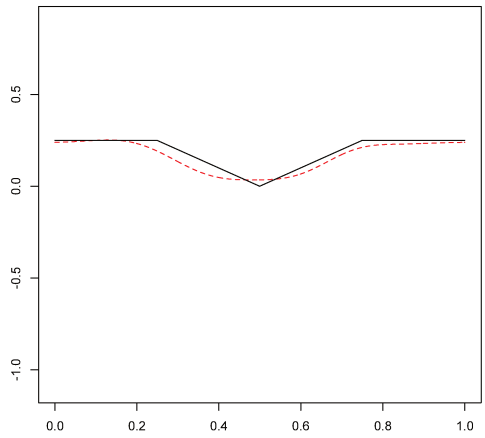

Figure 3.3: Estimator of S for n=200.

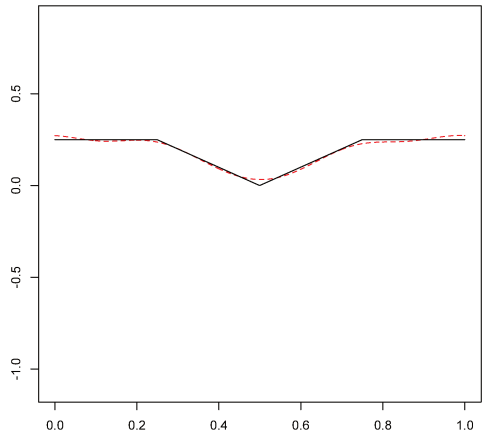

Figure 3.4: Estimator of S for n=1000.

Remark 3.3. From numerical simulations of the procedure (3.17) with various observations numbers n we may conclude that the quality of the proposed procedure is good for practical needs, i.e. for reasonable (non large) number of observations. We can also add that the quality of the estimation improves as the number of observations increases.

Now we give the algorithm of the model selection procedure given in Section 3.2

Algorithm 2 Model selection procedure

Require: n, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

p : satisfying Condition $\mathbf{H}_5$) given in (3.35)

ρ_1, ρ_2, ς^*: satisfying Conditions (2.8) and (2.9)

$k^* \geq 1$, ε: satisfying Condition (2.26)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$m = [1/\varepsilon^2]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

for $k \longleftarrow 1$ *to* n **do**

Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (2.28)

end for

end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varepsilon, \ldots, m\varepsilon\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,p} = \frac{1}{n} \int_0^n \Psi_{k,p}(t) \mathrm{d}\, y_t$.

$\widetilde{\theta}_{k,p} \longleftarrow \widehat{\theta}^2_{k,p} - \frac{1}{n}$.

The observation $(y_t)_{0 \leq t \leq n}$ are given in (2.1) with the noise process (2.2) and $(\Psi_{k,p})_{1 \leq k \leq p}$ is the basis given in (3.5)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,p}, ..., \widehat{\theta}_{n,p})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,p}, ..., \widetilde{\theta}_{n,p})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

$J_n(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l) \widehat{\theta}^2_{j,p} - 2 \sum_{j=1}^n \lambda_{i,j}(l) \widetilde{\theta}_{j,p} + \delta\, P_n(\lambda)$.

where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_n is the penalty term given in (3.15)

end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_n(\lambda), \Lambda = \{\lambda_\alpha\,, \alpha \in \mathcal{A}\}$.

Appendix

Property of the penalty term

Lemma 3.1. *For any $n \geq 1$ and $\lambda \in \Lambda$,*

$$P_n^0(\lambda) \leq \mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{\mathbf{L}_{1,Q}}{n},$$

where the coefficient $P_n^0(\lambda)$ is defined in (3.21) *and the $\mathbf{L}_{1,Q}$ is defined in* (3.38).

Proof. By the definition of $\mathrm{Err}(\lambda)$ in (3.10) one has

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

In view of Proposition 3.1 we obtain that

$$\mathcal{R}_Q(\widehat{S}_\lambda, S) = \mathbf{E}_Q \,\mathrm{Err}(\lambda) \geq \frac{1}{n}\sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q\, \xi_{j,n}^2 \geq P_n^0(\lambda) - \frac{\mathbf{L}_{1,Q}}{n}\,.$$

Hence we otain Lemma 3.1.

Properties of the Fourier coefficients

Lemma 3.2. *Let f be an absolutely continuous function, $f : [0,1] \to \mathbb{R}$, with $\|\dot{f}\| < \infty$ and g be a simple function, $g : [0,1] \to \mathbb{R}$ of the form $g(t) = \sum_{j=1}^{p} c_j\, \chi_{(t_{j-1},t_j]}(t)$, where c_j are some constants. Then for any $\varepsilon > 0$, the function $\Delta = f - g$ satisfies the following inequalities*

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

Lemma 3.3. *Let the function $S(t)$ in* (3.1) *be absolutly continuous and have an absolutely integrable derivative. Then the coefficients $(\overline{\theta}_{j,p})_{1\leqslant j\leqslant p}$ defined in* (3.6) *satisfy the inequalities*

$$|\overline{\theta}_{1,p}| \leqslant \|S\|_1 \quad \textit{and} \quad \max_{2\leqslant j\leqslant p} j|\overline{\theta}_{j,p}| \leqslant 2\sqrt{2}\|\dot{S}\|_1\,. \tag{3.78}$$

Lemma 3.4. *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 $W_{\mathbf{r}}^1$ satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality*

$$\sum_{j=N}^{p} \theta_{j,p}^2 \;\leq\; (1+\widetilde{\varepsilon}) \sum_{j\geq N} \theta_j^2 \;+ (1+\widetilde{\varepsilon}^{-1})\,\mathbf{r}\,p^{-2}\,. \tag{3.79}$$

Lemma 3.5. *For any $p \geq 2$ and $\mathbf{r} > 0$, the coefficients $(\theta_{j,p})_{1\leq j\leq p}$ of functions S from the class $W_{\mathbf{r}}^1$ satisfy the following inequality*

$$\max_{1\leq j\leq p} \sup_{S\in W_{\mathbf{r}}^1} \left(|\theta_{j,p} - \theta_j| - 2\pi\sqrt{\mathbf{r}}\, j\, p^{-1} \right) \leq 0\,. \tag{3.80}$$

Lemma 3.6. *For any $p \geq 2$ and $\mathbf{r} > 0$ the correction coefficients $\left(h_{j,p}\right)_{1\leq j\leq p}$ for the functions S from the class $W^2_{\mathbf{r}}$ satisfy the following inequality*

$$\sup_{S\in W^2_{\mathbf{r}}} \sum_{j=1}^{p} h^2_{j,p} \leq 3\mathbf{r}\, p^{-2}\,. \tag{3.81}$$

Lemmas 3.2 – 3.6 are proven in [18].

Chapter 4

Non-parametric estimation for Lévy regression models

4.1 Introduction

Let us consider a regression model in continuous time with the Levy noise

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \varepsilon\, \mathrm{d}\xi_t\,, \quad 0 \le t \le 1\,, \tag{4.1}$$

where S is an unknown function defined on $\mathbb{R}$ with values in $\mathbb{R}$, $(\xi_t)_{0\le t\le 1}$ is some unobserved noise and $\varepsilon > 0$ is the noise intensity. The problem is to estimate the function S on the basis of observations $(y_t)_{0\le t\le 1}$ when $\varepsilon \to 0$. In this chapter we consider the estimation problem in the adaptive setting, i.e. when the regularity of S is unknown and we assume that the noise $(\xi_t)_{0\le t\le 1}$ is a Lévy process with unknown distribution Q on the Skorokhod space $\mathcal{D}[0,1]$. We know only that this distribution belongs to some distribution family $\mathcal{Q}_\varepsilon$ specified below.

Note that if $(\xi_t)_{0\le t\le 1}$ is the Brownian motion, then we obtain the well known "signal+white noise" model (see, for example, [9], [47], [41]). It should be noted also that the model (4.1) is very popular in the statistical radio-physics. This is the estimation problem of the signal S, observed under the white noise, when the signal/noise ratio goes to infinity.

By making use of the robust estimation approach developed for nonparametric problems in [36, 17, 18] we set the robust risk as

$$\mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) = \sup_{Q\in\mathcal{Q}^*_\varepsilon} \mathcal{R}_Q(\widehat{S}_\varepsilon, S), \tag{4.2}$$

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

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

The goal of this chapter is to develop the sharp model selection method for estimating the unknown signal S. The interest in such statistical procedures can be explained by the fact that they provide adaptive solutions for the nonparametric estimation through the sharp non-asymptotic oracle inequalities which give non-asymptotic upper bound for the quadratic risk including the minimal risk over chosen family of estimators with some coefficient closed to

one (see, for example, [37] for discrete time and [18] for continuous time). The origin of the model selection method goes back to early seventies with the pioneering papers by Akaike [30] and Mallows [23] who suggested to use penalization in a log-likelihood type criterion. Barron, Birgé, Massart [31], Massart [45] and Kneip [40] developed a non-asymptotic model selection method which enables one to derive non-asymptotic oracle inequalities for the non-parametric regression models with Gaussian disturbances. Unfortunately, these methods cannot be applied to the non-Gaussian regression models, since the estimators of the Fourier coefficients in such cases are not independent random variables. For these reasons, in order to estimate the function in non-Gaussian regression models, we use the model selection method developed by [37, 38, 39] for non-Gaussian heteroscedastic regression models in discrete time.

When constructing the sharp model selection procedures, in this chapter, we will use the approach close to that of the papers [14], [15], [16], [18] developed for the estimation of a 1-periodic function in continuous time on a large time interval, i.e.

$$\mathrm{d}x_t = S(t)\mathrm{d}t + \mathrm{d}\eta_t\,, \quad 0 \le t \le n\,.$$

Note that, for any $0 < t < 1$, setting $y_t = n^{-1}\sum_{j=1}^{n}(x_{t+j} - x_j)$, we can represent this model as a model with small parameter of the form (4.1)

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon\,\mathrm{d}\xi_t\,,$$

where $\varepsilon = n^{-1/2}$ and $\xi_t = n^{-1/2}\sum_{j=1}^{n}(\eta_{t+j} - \eta_t)$. The main difference between this model and the original one is that the jumps are small, i.e.

$$\Delta\xi_t = \mathrm{O}(n^{-1/2}) = \mathrm{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0\,,$$

but we have not such property in the model (4.1). Therefore, unfortunately, we cannot use directly the method developed for the estimation problem on the large time interval to the model (4.1). So, the main goal of this paper is to develop a new sharp model selection method for the estimation problem of the function S as $\varepsilon \to 0$.

As an application of the sharp model selection method in this chapter we consider the problem of the detection of the number of signals for the model (4.1). In many areas of science and technology the problem arise how to select the number degrees of freedom for a statistical model that describes the phenomenons under study most adequately [30]. An important class of such problems is the detection problem of the number of signals with unknown parameters in the noise. For example, in the signal multi-path information transmission there is a detection problem of the number of rays in a multipath channel. This problem is often reduced to the detection of the number of signals. As a result, effective algorithms for the detection of the number of signals can significantly improve the noise immunity in the data transmission over a multipath channel [34, 42, 33, 48, 50, 49, 51]. In all this chapter the problem of the detection of the number of signals are considered only for observation with white noise. In this chapter we consider this problem for non-Gaussian noise with jumps given by (4.3).

4.2 Transformation of the observations

In this chapter the noise process $(\xi_t)_{0\le t\le 1}$ is defined by the following Lévy process

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \widetilde{\mu})_t \,. \tag{4.3}$$

Here, ϱ_1 and ϱ_2 are some constants, $(w_t)_{t\ge 0}$ is a standard Brownian motion, $\mu(\mathrm{d}s\,\mathrm{d}x)$ is the jump measure with the deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is some positive measure on $\mathbb{R}$ (see, for example [10, 6] for details).

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^4) < \infty \,. \tag{4.4}$$

Note that $\Pi(\mathbb{R})$ may be equal to $+\infty$. In the sequel we will denote by Q the distribution of the process $(\xi_t)_{0\le t\le 1}$ in the Skorokhod space $\mathbf{D}[0,1]$ and by $\mathcal{Q}^*_\varepsilon$ we denote all these distributions for which the parameters ϱ_1 and ϱ_2 satisfy the condition

$$\varkappa_Q = \varrho_1^2 + \varrho_2^2 \le \varsigma^*_\varepsilon \,, \tag{4.5}$$

where the bound ς^*_ε is such that for any $\check{\delta} > 0$

$$\liminf_{\varepsilon\to 0} \varepsilon^{-\check{\delta}} \varsigma^*_\varepsilon > 0 \quad \text{and} \quad \lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \varsigma^*_\varepsilon = 0 \,. \tag{4.6}$$

First of all, we need to eliminate the large jumps in the observations (4.1), i.e. we transform this model as

$$\check{y}_t = y_t - \sum_{0\le s\le t} \Delta y_s \, \mathbf{1}_{\{|\Delta y_s| > \overline{a}\}} \,. \tag{4.7}$$

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

$$\mathrm{d}\check{y}_t = S(t)\mathrm{d}t + \varepsilon \mathrm{d}\check{\xi}_t - \varepsilon\, \varrho_2\, \Pi(\overline{h}_\varepsilon)\, \mathrm{d}t \,, \tag{4.8}$$

where

$$\check{\xi}_t = \varrho_1 w_t + \varrho_2\, \check{z}_t \quad \text{and} \quad \check{z}_t = h_\varepsilon * (\mu - \widetilde{\mu})_t \,.$$

The functions $h_\varepsilon(x) = x\mathbf{1}_{\{|x|\le \widetilde{a}_\varepsilon\}}$ and $\overline{h}_\varepsilon(x) = x\mathbf{1}_{\{|x|>\widetilde{a}_\varepsilon\}}$ where the truncated threshold is defined by $\widetilde{a}_\varepsilon = \overline{a}/\varrho_2\varepsilon$.

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

$$\sup_{0\le j\le n} \sup_{0\le t\le 1} |\phi_j(t)| \le \phi^* < \infty \,, \tag{4.9}$$

where $n = n_\varepsilon = [1/\varepsilon^2]$ and $[x]$ denotes the integer part of x. For example, we can take the trigonometric basis defined in (1.15)

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

$$I_t(f) = \int_0^t f(s)\mathrm{d}\xi_s \quad \text{and} \quad \check{I}_t(f) = \int_0^t f(s)\mathrm{d}\check{\xi}_s \tag{4.10}$$

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

$$\mathbf{E}\, I_t^2(f) = \varkappa_Q \, \|f\|_t^2 \quad \text{and} \quad \mathbf{E}\, \check{I}_t^2(f) = \check{\varkappa}_Q \, \|f\|_t^2 \,, \tag{4.11}$$

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

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

.

To estimate the function S we use the following Fourier series

$$S(t) = \sum_{j\geq 1} \theta_j \, \phi_j(t) \,, \tag{4.12}$$

where

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

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

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

and, for $j \geq 2$,

$$\widehat{\theta}_{j,\varepsilon} = \int_0^1 \phi_j(t)\mathrm{d}\,\check{y}_t \,. \tag{4.13}$$

Taking into account here that for any j the integral $\int_0^1 \phi_j(t)\mathrm{d}t = 0$ we obtain from (4.8) that these Fourier coefficients can be represented as

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

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

$$\widehat{\theta}_{j,\varepsilon} = \theta_j + \varepsilon\,\check{\xi}_j \,. \tag{4.14}$$

Now, according to the model selection approach developed in [17] - [18] we need to define for any $u \in \mathbb{R}^n$ the following functions

$$B_{1,\varepsilon}(u) = \sum_{j=1}^n u_j \, \varsigma_j \quad \text{and} \quad B_{2,\varepsilon}(u) = \sum_{j=1}^n u_j \, \widetilde{\xi}_j \,, \tag{4.15}$$

where $\varsigma_j = \mathbf{E}\,(\check{\xi}_j)^2 - \check{\varkappa}_Q$ and $\widetilde{\xi}_j = (\check{\xi}_j)^2 - \mathbf{E}\,(\check{\xi}_j)^2$.

Proposition 4.1. *The following upper bound holds.*

$$\sup_{u\in[0,1]^n} \left|B_{1,\varepsilon}(u)\right| \le \varkappa_Q \,. \tag{4.16}$$

Proof. Taking into account that $\varsigma_1 = \varkappa_Q - \check{\varkappa}_Q \le \varkappa_Q$ and $\varsigma_j = 0$ for $j \ge 2$ we immediately have the upper bound (4.16). □

Now let us recall the Novikov inequalities, [44], also referred to as the Bichteler–Jacod inequalities, see ([32, 43]), providing bounds of the moments of the supremum of purely discontinuous local martingales for $p \ge 2$,

$$\mathbf{E}\sup_{t\le 1}|g*(\mu-\nu)_t|^p \le C_p^* \left(\mathbf{E}\left(|g|^2*\nu_1\right)^{p/2} + \mathbf{E}\left(|g|^p*\nu_1\right)\right) , \tag{4.17}$$

where C_p^* is some positive constant.
Now, for any $u \in \mathbb{R}^n$ we set

$$|u|^2 = \sum_{j=1}^n u_j^2 \quad \text{and} \quad \#(u) = \sum_{j=1}^n \mathbf{1}_{\{u_j\neq 0\}} \,. \tag{4.18}$$

Proposition 4.2. *For any fixed truncated model parameter $a > 0$ and for any vector $u \in \mathbb{R}^n$ with $|u| \le 1$, we have*

$$\mathbf{E}\left|B_{2,\varepsilon}^2(u)\right| \le U_Q + 6\varkappa_Q \left(\frac{a}{\varepsilon}\right)^2 \#(u)\,(\phi^*)^4 , \tag{4.19}$$

where $U_Q = 24\varkappa_Q^2 + 6\varrho_2^4\,\Pi(x^4)$.

Proof. First note that

$$B_{2,\varepsilon}^2(u) \le 2\,\widetilde{\xi}_1^2 + 2B_{2,\varepsilon}^2(u') \,, \tag{4.20}$$

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

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

To study the last term in the right hand side of the inequality (4.20) we set for any function f from $\mathbf{L}_2[0,1]$

$$\widetilde{I}_t(f) = \check{I}_t^2(f) - \mathbf{E}\,\check{I}_t^2(f) \,.$$

Note that for $j \ge 2$ we define the random variables $\widetilde{\xi}_j = \widetilde{I}_1(\phi_j)$. So,

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

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

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

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

$$\mathrm{d}\check{I}_t(f) = \varrho_1 \,\mathrm{d}w_t + \varrho_2 \,\mathrm{d}\check{z}_t\,,$$

we obtain that

$$\mathrm{d}\widetilde{I}_t(f) = 2\varrho_1\, \check{I}_t(f) f(t) \mathrm{d}w_t + 2\varrho_2\, \check{I}_{t-}(f) f(t) \mathrm{d}\check{z}_t + \varrho_2^2\, f^2(t)\, \mathrm{d}\check{m}_t\,.$$

So, setting

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

we obtain that

$$\mathrm{d}D_t = 2\varrho_1\, V_t\, \mathrm{d}w_t + 2\varrho_2\, V_{t-}\, \mathrm{d}\check{z}_t + \varrho_2^2\, \Psi_t\, \mathrm{d}\check{m}_t\,.$$

So, we obtain that

$$D_1^2 \le 12\varrho_1^2 \left(\int_0^1 V_t \,\mathrm{d}w_t \right)^2 + 12\varrho_2^2\, \check{M}_1^2 + 3\varrho_2^4 \left(\int_0^1 \Psi_{t-}\, \mathrm{d}\check{m}_t \right)^2, \tag{4.21}$$

where

$$\check{M}_t = \int_0^t V_{s-}(u)\, \mathrm{d}\check{z}_s.$$

Moreover, taking into account that for any f, g from $\mathbf{L}_2[0,1]$

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

we get

$$2\int_0^1 \mathbf{E}\, V_t^2 \,\mathrm{d}t = 2\sum_{i,j=2}^{n} u_i\, u_j \int_0^1 \phi_i(t)\phi_j(t)\, \mathbf{E}\, \check{I}_t(\phi_i)\, \check{I}_t(\phi_j)\, \mathrm{d}t \;\; = \check{\varkappa}_Q \sum_{i=2}^{n} u_i^2 \left(\int_0^1 \phi_i^2(t)\, \mathrm{d}t \right)^2.$$

Thus,

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

Now, to estimate the second term in the inequality (4.21) note that in view of the inequality (4.17) for any bounded function f and any $0 \le t \le 1$

$$\begin{aligned}
\mathbf{E}\, \check{I}_t^4(f) &\le 8\varrho_1^4\, \mathbf{E} \left(\int_0^t f(s) \mathrm{d}w_s \right)^4 + 8\varrho_2^4\, \mathbf{E} \left(\int_0^t f(s-) \mathrm{d}\check{z}_t \right)^4 \\
&\le 24\varrho_1^4 \int_0^1 f^2(t) \mathrm{d}t + C_4^* \left(\left(\Pi(h_\varepsilon^2) \int_0^1 f^2(t) \mathrm{d}t \right)^2 + \Pi(h_\varepsilon^4) \int_0^1 f^4(t) \mathrm{d}t \right),
\end{aligned}$$

i.e.

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

Now it is easy to see that through the Hölder's inequality the term V_t can be estimated as

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

From here and the inequality (4.17) it follows that

$$\sup_{0\le t\le 1} \mathbf{E}\, \check{M}_t^4 \le C_4^* \, \left(\left(\Pi(h_\varepsilon^2) \right)^2 + \Pi(h_\varepsilon^4) \right) \, \int_0^1 \mathbf{E}\, V_t^4 \, \mathrm{d}t < \infty$$

and, therefore,

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

This implies that

$$\mathbf{E} \int_0^1 \check{M}_{t-} \, \mathrm{d}\check{M}_t = 0 \, .$$

Thus, the Ito's formula implies

$$2\mathbf{E}\, \check{M}_1^2 = \mathbf{E} \sum_{0\le t\le 1} (\Delta \check{M}_t)^2 = 2\Pi(h_\varepsilon^2) \, \int_0^1 \mathbf{E}\, V_t^2 \mathrm{d}t \le \Pi(h_\varepsilon^2) \, \check{\varkappa}_Q \, .$$

In the same way we calculate

$$\begin{aligned} \varrho_2^2 \, \mathbf{E} \, \left(\int_0^1 \Psi_{t-} \mathrm{d}\check{m}_t \right)^2 &= \varrho_2^2 \mathbf{E} \sum_{0\le t\le 1} (\Delta \check{m}_t)^2 \, \Psi_{t-}^2 \\ &= \varrho_2^2 \Pi(h_\varepsilon^4) \, \int_0^1 \Psi_t^2 \mathrm{d}t \le (a/\varepsilon)^2 \, (\phi^*)^4 \, \#(u) \, . \end{aligned}$$

So, we obtain that

$$\mathbf{E}\, D_1^2 \le 6\varrho_1^2 \, \check{\varkappa}_Q + 3\varrho_2^2 \, \Pi(x^2) \left(2\check{\varkappa}_Q + (\phi^*)^4 \right) \le 6 \, \varkappa_Q^2 + 3\varrho_2^2 (\phi^*)^4 \, .$$

Similarly we obtain that

$$\mathbf{E}\widetilde{\xi}_1^2 \le 6\varkappa_Q^2 + 3\varrho_2^4 \, \Pi(x^4) \, .$$

This implies the upper bound (4.19). □

4.3 Model selection

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

$$\widehat{S}_\lambda(x) = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,\varepsilon} \phi_j(x) \, , \tag{4.22}$$

where $n = [1/\varepsilon^2]$, the weights $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$, $\widehat{\theta}_{j,\varepsilon}$ is defined in (4.13) and ϕ_j in (4.9) . Now we set

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = \max_{\lambda\in\Lambda} \sum_{j=1}^{n} \mathbf{1}_{\{\lambda_j>0\}}\,, \tag{4.23}$$

where $\#(\Lambda)$ is the cardinal number of Λ. In the sequel we assume that $|\Lambda|_* \ge 1$. Now we chose the truncating parameter $\overline{a}_\varepsilon$ as

$$\overline{a}_\varepsilon = \frac{\varepsilon}{|\Lambda|_*}\,. \tag{4.24}$$

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

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

which in our case is equal to

$$\mathrm{Err}_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,\varepsilon}\theta_j + \sum_{j=1}^{\infty} \theta_j^2. \tag{4.25}$$

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

$$\widetilde{\theta}_{j,\varepsilon} = \widehat{\theta}^2_{j,\varepsilon} - \varepsilon^2\widehat{\varkappa}_\varepsilon\,, \tag{4.26}$$

where $\widehat{\varkappa}_\varepsilon$ is a some estimate for the variance parameter $\check{\varkappa}_Q$ from (4.11). If it is known we set $\widehat{\varkappa}_\varepsilon = \check{\varkappa}_Q$ if not this estimator will be prescribed later.
Finally, to choose the weights we will minimize the following cost function

$$J_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,\varepsilon} + \delta\,\widehat{P}_\varepsilon(\lambda)\,, \tag{4.27}$$

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

$$\widehat{P}_\varepsilon(\lambda) = \varepsilon^2\widehat{\varkappa}_\varepsilon|\lambda|^2 \quad \text{and} \quad |\lambda|^2 = \sum_{j=1}^{n} \lambda_j^2\,. \tag{4.28}$$

Note that, if the $\varkappa_Q$ is known, then the penalty is defined as

$$P_\varepsilon(\lambda) = \varepsilon^2\,\check{\varkappa}_Q|\lambda|^2\,. \tag{4.29}$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}\,, \tag{4.30}$$

where

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_\varepsilon(\lambda). \tag{4.31}$$

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

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

$$\mathcal{A} = \{1,\dots,k^*\}\times\{l_1,\dots,l_m\}\,, \tag{4.32}$$

where

$$l_i = i\,\varpi \quad \text{and} \quad m = [1/\varpi]\,.$$

We assume that both the parameters $k^*\ge 1$ and $0<\varpi<1$ are functions of ε, i.e.$k^* = k^*_\varepsilon$ and $\varpi = \varpi_\varepsilon$, such that

$$\begin{cases} \lim_{\varepsilon\to 0}\, k^*_\varepsilon = +\infty\,, \quad \lim_{\varepsilon\to 0}\, \dfrac{k^*_\varepsilon}{|\ln\varepsilon|} = 0\,, \\ \\ \lim_{\varepsilon\to 0}\varpi_\varepsilon = 0 \quad \text{and} \quad \lim_{\varepsilon\to 0}\,\varepsilon^{-\delta}\varpi_\varepsilon\, = +\infty, \end{cases} \tag{4.33}$$

for any $\delta>0$. One can take, for example, for $0<\varepsilon<1$

$$\varpi_\varepsilon = \frac{1}{|\ln\varepsilon|} \quad \text{and} \quad k^*_\varepsilon = k^*_0 + \sqrt{|\ln\varepsilon|}\,, \tag{4.34}$$

where $k^*_0\ge 0$ is some fixed constant and the threshold ς^*_ε is introduced in (4.5). For each $\alpha=(\beta,l)\in\mathcal{A}$, we introduce the weight sequence

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

where $p=[\varepsilon^{-2}]$,

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\le j<j_*\}} + \Bigl(1-(j/\omega_\alpha)^\beta\Bigr)\,\mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}\,. \tag{4.35}$$

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

$$\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon. \quad \text{and} \quad \mathrm{d}_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\,. \tag{4.36}$$

Now we define the set Λ as

$$\Lambda\ =\ \{\lambda_\alpha\,,\ \alpha\in\mathcal{A}\}\,. \tag{4.37}$$

Note, that these weight coefficients are used in [17, 18] for continuous time regression models to show the asymptotic efficiency.

In the sequel we need to estimate the variance parameter $\varkappa_Q$ from (4.11). To this end we set for any $0<\varepsilon\le 1/\sqrt{3}$

$$\widehat{\varkappa}_\varepsilon = \sum_{j=[1/\varepsilon]+1}^{n}\widehat{\mathbf{T}}^2_{j,\varepsilon}\,, \quad n=[1/\varepsilon^2]\,, \tag{4.38}$$

where $\widehat{\mathbf{T}}_{j,\varepsilon}$ are the estimators of the Fourrier coefficients with respect to the trigonometric basis (1.15) , i.e.

$$\widehat{\mathbf{T}}_{j,\varepsilon} = \int_0^1 \mathrm{Tr}_j(t)\mathrm{d}\check{y}_t\,. \tag{4.39}$$

Remark 4.1. Note that similar sharp oracle inequalities were obtained before in the papers [37] and [17] for the nonparametric regression models in the discrete and continuous time respectively. In this chapter we obtain these inequalities for the model selection procedures based on any arbitrary orthogonal basis function. We use the trigonometric function only to estimate the noise intensity $\check{\varkappa}_Q$.

4.4 Oracle inequality

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

$$\Psi_{Q,\varepsilon} = 8\check{\varkappa}_Q(1+\check{\iota}) + \frac{4U_{1,Q}\check{\iota}}{\check{\varkappa}_Q}\,, \tag{4.40}$$

where

$$U_{1,Q} = 24\check{\varkappa}_Q^2 + 6\varrho_2^4\,\Pi(x^4) + 6\check{\varkappa}_Q\,(\phi^*)^4\,.$$

We start with the sharp oracle inequalities.

Theorem 4.1. *Assume that for the model* (4.1) *the condition* (4.4) *holds. Then, for any* $0<\delta<1/6$*, the estimator of* S *given in* (4.30) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*,S) \le \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda,S) + \varepsilon^2\frac{\Psi_{Q,\varepsilon}+12|\Lambda|_*\,\mathbf{E}_S|\widehat{\varkappa}_\varepsilon-\check{\varkappa}_Q|}{\delta}\,. \tag{4.41}$$

Proof. First, note that we can rewrite the empirical squared error in (4.25) as follows

$$\mathrm{Err}_\varepsilon(\lambda) = J_\varepsilon(\lambda) + 2\sum_{j=1}^n \lambda(j)\check{\theta}_{j,\varepsilon} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda), \tag{4.42}$$

where $\check{\theta}_{j,\varepsilon} = \widetilde{\theta}_{j,\varepsilon} - \theta_j\widehat{\theta}_{j,\varepsilon}$. Now using the definition of $\widetilde{\theta}_{j,\varepsilon}$ in (4.26) we obtain that

$$\check{\theta}_{j,\varepsilon} = \varepsilon\theta_j\xi_j + \varepsilon^2\widetilde{\xi}_{j,\varepsilon} + \varepsilon^2\varsigma_{j,\varepsilon} + \varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,,$$

where $\varsigma_{j,\varepsilon} = \mathbf{E}(\check{\xi}_{j,\varepsilon})^2 - \check{\varkappa}_Q$ and $\widetilde{\xi}_j = (\check{\xi}_j)^2 - \mathbf{E}(\check{\xi}_j)^2$. Setting

$$M_\varepsilon(\lambda) = \varepsilon\sum_{j=1}^n \lambda(j)\theta_j\check{\xi}_j \quad\text{and}\quad L(\lambda) = \sum_{j=1}^n \lambda(j)\,, \tag{4.43}$$

we can rewrite (4.42) as

$$\begin{aligned}\mathrm{Err}_\varepsilon(\lambda) &= J_\varepsilon(\lambda) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,L(\lambda) + 2M_\varepsilon(\lambda) + 2\varepsilon^2 B_{1,\varepsilon}(\lambda)\\ &\quad + 2\varepsilon\sqrt{P_\varepsilon(\lambda)}\frac{B_{2,\varepsilon}(u_\lambda)}{\sqrt{\check{\varkappa}_Q}} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda),\end{aligned} \tag{4.44}$$

where $u_\lambda = \lambda/|\lambda|$, the exact penalization is defined in (4.29) and the functions $B_{1,\varepsilon}(\cdot)$ and $B_{2,\varepsilon}(\cdot)$ are defined in (4.15). It should be noted that for the truncated parameter (4.24) the bound (4.19) implies

$$\sup_{\lambda\in\Lambda} \mathbf{E}_Q \left| B^2_{2,\varepsilon}(u_\lambda) \right| \le U_Q + 6\check{\varkappa}_Q \left(\frac{a}{\varepsilon}\right)^2 |\Lambda|_* \, (\phi^*)^4 = U_{1,Q}\,, \tag{4.45}$$

where $U_{1,Q} = U_Q + 6\check{\varkappa}_Q \, (\phi^*)^4$.

Let $\lambda_0 = (\lambda_0(j))_{1\le j\le n}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be as in (4.31). Substituting λ_0 and $\widehat{\lambda}$ in Equation (4.44), we obtain

$$\begin{aligned}
\mathrm{Err}_\varepsilon(\widehat{\lambda}) - \mathrm{Err}_\varepsilon(\lambda_0) = {} & J(\widehat{\lambda}) - J(\lambda_0) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\, L(\varpi) \\
& + 2\varepsilon^2 B_{1,\varepsilon}(\varpi) + 2M_\varepsilon(\varpi) \\
& + 2\varepsilon\sqrt{P_\varepsilon(\widehat{\lambda})}\frac{B_{2,\varepsilon}(\widehat{u})}{\sqrt{\check{\varkappa}_Q}} - 2\varepsilon\sqrt{P_\varepsilon(\lambda_0)}\frac{B_{2,\varepsilon}(u_0)}{\sqrt{\check{\varkappa}_Q}} \\
& - \delta\widehat{P}_\varepsilon(\widehat{\lambda}) + \delta\widehat{P}_\varepsilon(\lambda_0),
\end{aligned} \tag{4.46}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{u} = u_{\widehat{\lambda}}$ and $u_0 = u_{\lambda_0}$. Note that by (4.23)

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

The inequality

$$2|ab| \le \delta a^2 + \delta^{-1} b^2 \tag{4.47}$$

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

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

From the bound (4.16) it follows that for $0 < \delta < 1$

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

where $B^*_{2,\varepsilon} = \sup_{\lambda\in\Lambda} B^2_{2,n}(u_\lambda)$. It should be noted that through (4.45) we can estimate this term as

$$\mathbf{E}_Q\, B^*_{2,\varepsilon} \le \sum_{\lambda\in\Lambda} \mathbf{E}_Q\, B^2_{2,\varepsilon}(u_\lambda) \le \imath U_{1,Q}\,. \tag{4.48}$$

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

$$\begin{aligned}
\mathrm{Err}_\varepsilon(\widehat{\lambda}) \le {} & \mathrm{Err}_\varepsilon(\lambda_0) + 2M_\varepsilon(\varpi) + 2\varepsilon^2\frac{B^*_{2,\varepsilon}}{\delta\check{\varkappa}_Q} + 2\varepsilon^2\,\check{\varkappa}_Q \\
& + \frac{6\varepsilon^2|\Lambda|_*}{n}|\widehat{\varkappa} - \check{\varkappa}_Q| + 2\delta P_\varepsilon(\lambda_0).
\end{aligned} \tag{4.49}$$

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

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

Moreover, note that

$$M_\varepsilon^2(\upsilon) \le 2\upsilon^2(1)\,\xi_1^2 + 2\,\check{I}_1(\Phi)\,,$$

where $\Phi(t) = \sum_{j=2}^{n} \upsilon(j)\theta_j\phi_j(t)$. Therefore, thanks to (4.11) we obtain that for any non-random $\upsilon \in \mathbb{R}^n$

$$\mathbf{E}M_\varepsilon^2(\upsilon) \le 2\check{\varkappa}_Q\varepsilon^2\sum_{j=1}^{n}\upsilon^2(j)\theta_j^2 = 2\check{\varkappa}_Q\varepsilon^2||S_\upsilon||^2\,. \tag{4.50}$$

To estimate this function for a random vector we set

$$M_\varepsilon^* = \sup_{\upsilon\in\Lambda_1}\frac{M^2(\upsilon)}{\varepsilon^2||S_\upsilon||^2} \quad \text{and} \quad \Lambda_1 = \Lambda - \lambda_0\,.$$

So, through the inequality (4.47)

$$2|M_\varepsilon(\upsilon)| \le \delta||S_\upsilon||^2 + \varepsilon^2\frac{M_\varepsilon^*}{\delta}\,. \tag{4.51}$$

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

$$\mathbf{E}\,M_\varepsilon^* \le \sum_{\upsilon\in\Lambda_1}\frac{\mathbf{E}\,M_\varepsilon^2(\upsilon)}{\varepsilon^2||S_\upsilon||^2} \le 2\sum_{\upsilon\in\Lambda_1}\check{\varkappa}_Q = 2\check{\varkappa}_Q\,\check{\iota}\,, \tag{4.52}$$

where $\nu = \#(\Lambda)$. Moreover, note that, for any $\upsilon \in \Lambda_1$,

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

where $\upsilon^2 = (\upsilon^2(j))_{1\le j\le n}$. Taking into account that, for any $x \in \Lambda_1$, the components $|\upsilon(j)| \le 1$, we can estimate the last term as in (4.50), i.e.

$$\mathbf{E}\,M_\varepsilon^2(\upsilon^2) \le 2\varepsilon^2\check{\varkappa}_Q\,||S_\upsilon||^2\,.$$

Similarly, setting

$$M_{1,\varepsilon}^* = \sup_{\upsilon\varepsilon\Lambda_1}\frac{M_\varepsilon^2(\upsilon^2)}{\varepsilon^2||S_\upsilon||^2},$$

we obtain

$$\mathbf{E}_Q\,M_{1,\varepsilon}^* \le 2\check{\varkappa}a_Q\,\check{\iota}\,. \tag{4.53}$$

In the same way we find that

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

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

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

So, from (4.51) we get

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

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

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

Using this bound in (4.49) we obtain

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

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

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

Using here the bounds (4.48), (4.52), (4.53) we obtain that

$$\mathcal{R}(\widehat{S}_*, S) \le \frac{1+\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{8\varepsilon^2\check{\varkappa}_Q\check{\iota}}{\delta} + \frac{4\varepsilon^2 U_{1,Q}\check{\iota}}{\delta\check{\varkappa}_Q}$$
$$+ 4\varepsilon^2\,\check{\varkappa}_Q + 12\varepsilon^2\,|\Lambda|_*\mathbf{E}_Q\,|\widehat{\varkappa} - \check{\varkappa}_Q| + \frac{2\delta}{1-3\delta}\,P_\varepsilon(\lambda_0)\,.$$

Now, Lemma 4.1 implies directly the inequality (4.1). Hence we get the desired result. □

Corollary 4.1. *Assume that for the model* (4.1) *the condition* (4.4) *holds. If the variance parameter* $\check{\varkappa}_Q$ *is known, then for any* $0 < \delta < 1/6$*, the estimator of* S *given in* (4.30)*, with the truncated parameter* $\overline{a} = \varepsilon/\sqrt{|\Lambda|_*}$ *satisfying the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda, S) + \varepsilon^2\frac{\Psi_{Q,\varepsilon}}{\delta}\,. \tag{4.54}$$

We need to study the estimate (4.38).

Proposition 4.3. *Assume that in the model* (4.1) *the unknown function* S *is continuously differentiable. Then, for any* $0 < \varepsilon \le 1/\sqrt{3}$

$$\mathbf{E}_Q \left| \widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q \right| \le \varepsilon \Upsilon_Q(S) + \frac{\sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} , \tag{4.55}$$

where $\Upsilon_Q(S) = 4(\|\dot{S}\| + 1)^2 \left(1 + \sqrt{\check{\varkappa}_Q} + 2\check{\varkappa}_Q + \sqrt{U_Q}\right)$.

Proof. We use here the same method as in [14]. First, note that from the definitions (4.14) and (4.39) we obtain

$$\widehat{\mathbf{T}}_{j,\varepsilon} = \mathbf{T}_j + \varepsilon \check{\eta}_j , \tag{4.56}$$

where

$$\mathbf{T}_j = \int_0^1 S(t)\, Tr_j(t) \mathrm{d}t \quad \text{and} \quad \check{\eta}_j = \int_0^1 \mathrm{Tr}_j(t)\, \mathrm{d}\check{\xi}_t .$$

So, we have

$$\widehat{\varkappa}_\varepsilon = \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j^2 + 2\check{M}_\varepsilon + \varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} (\check{\eta}_j)^2 , \tag{4.57}$$

where $\check{M}_\varepsilon = \varepsilon \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j \check{\eta}_j$. Note that for continiously differentiable functions (see, for example, Lemma A.6 in [14]) the Fourrier coefficients $(\mathbf{T}_j)$ for any $n \ge 1$ satisfy the following inequality

$$\sum_{j=[1/\varepsilon]+1}^{\infty} \mathbf{T}_j^2 \le 4\varepsilon \left(\int_0^1 |\dot{S}(t)| \mathrm{d}t \right)^2 \le 4\varepsilon \|\dot{S}\|^2 . \tag{4.58}$$

The term $\check{M}_\varepsilon$ can be estimated in the same way as in (4.50), i.e.

$$\mathbf{E}_Q \check{M}_\varepsilon^2 \le \check{\varkappa}_Q \varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j^2 \le 4\varepsilon^3 \check{\varkappa}_Q \|\dot{S}\|^2 .$$

Moreover, taking into account that for $j \ge 2$ the expectation $\mathbf{E}\,(\check{\eta}_j)^2 = \check{\varkappa}_Q$ we can represent the last term in (4.57) as

$$\varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} (\check{\eta}_j)^2 = \varepsilon^2\, \check{\varkappa}_Q (n - [1/\varepsilon]) + \varepsilon\, B_{2,\varepsilon}(x') ,$$

where the function $B_{2,\varepsilon}(x')$ is defined in (4.15) and $x'_j = \varepsilon \mathbf{1}_{\{1/\varepsilon < j \le 1/\varepsilon^2\}}$. We remind that $n = [1/\varepsilon^2]$. Therefore, in view of Proposition 4.2 we obtain

$$\mathbf{E}_Q \left| \varepsilon^2 \sum_{j=[\sqrt{1/\varepsilon}]+1}^{n} \check{\eta}_j^2 - \check{\varkappa}_Q \right| \le 2\varepsilon\, \check{\varkappa}_Q + \varepsilon \sqrt{U_Q} + \frac{\sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} .$$

So, we obtain the bound (4.55). □

It is clear that in the case when $|\Lambda|_* \le 1/\varepsilon$ we obtain that

$$\mathbf{E}_Q \, |\widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q| \le \frac{\Upsilon_Q(S) + \sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} \, . \tag{4.59}$$

Now using this proposition we can obtain the following inequality.

Theorem 4.2. *Assume that for the model* (4.1) *the condition* (4.4) *holds and the unknown function S is continuously differentiable. Then the procedure* (4.30) *with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + \varepsilon^2 \frac{\Psi_{Q,\varepsilon} + (\|\dot{S}\| + 1)^2 g_{1,Q} + g_{2,Q}}{\delta} \, , \end{aligned} \tag{4.60}$$

where

$$g_{1,Q} = 48 \left(1 + \sqrt{\check{\varkappa}_Q} + 2\check{\varkappa}_Q + \sqrt{U_Q}\right) \quad \textit{and} \quad g_{2,Q} = 12\sqrt{6\check{\varkappa}_Q} \, .$$

Now we study the robust risk defined in (4.2) for the procedure (4.30).

We assume also that the upper bound for the basis functions in (4.9) may be dependent on $n \ge 1$, i.e. $\phi_* = \phi_*(n)$, such that for any $\check{\epsilon} > 0$

$$\lim_{n\to\infty} \frac{\phi_*(n)}{n^{\check{\epsilon}}} = 0 \, . \tag{4.61}$$

Theorem 4.3. *Assume that for the model* (4.1) *Condition* (4.4) *holds and the unknown function S is continuously differentiable. Then the robust risk of the procedure* (4.30) *with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfy the following oracle inequality*

$$\mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*_\varepsilon(\widehat{S}_\lambda, S) + \varepsilon^2 \, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta} \, , \tag{4.62}$$

where the term $\mathbf{U}^*_\varepsilon(S) > 0$ is such that under the conditions (4.61) *and* (4.33) *for any $r > 0$ and $\check{\delta} > 0$*

$$\lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \sup_{\|\dot{S}\|\le r} \mathbf{U}^*_\varepsilon(S) \, = 0 \, . \tag{4.63}$$

4.5 Adaptive robust efficiency

Now we study the asymptotically efficiency properties for the procedure (4.30) with the coefficients (4.35) with respect to the robust risks (4.12) defined by the distribution family (4.5) – (4.6). To this end we assume that the unknown function (4.12) belongs to the following ellipsoid in $\mathbf{L}_2$,

$$W^k_{\mathbf{r}} = \{S \in \mathbf{L}_2[0,1] \, : \, \sum_{j=1}^{\infty} a_j \, \theta_j^2 \, \le \mathbf{r}\}, \tag{4.64}$$

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

It is easy to see that in the case when the functions $(\phi_j)_{j\ge1}$ are trigonometric (1.15), then this set coincides with the Sobolev ball

$$W^k_{\mathbf{r}} = \{f \in \mathbf{C}^k_{per}[0,1] \, : \, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le \mathbf{r}\}\,, \tag{4.65}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some parameters, $\mathbf{C}^k_{per}[0,1]$ 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 $0 \le i \le k$. Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (4.12) is given by

$$l_*(\mathbf{r}) = \, ((2k+1)\mathbf{r})^{1/(2k+1)} \, \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)} . \tag{4.66}$$

Note that this is the well-known Pinsker's constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]).

Let $\mathcal{S}_\varepsilon$ be the set of all estimators $\widehat{S}_\varepsilon$ measurable with respect to the σ-algebra $\sigma\{y_t \, , \, 0 \le t \le 1\}$ generated by the process (4.1).

Theorem 4.4. *For the distribution family* (4.5) – (4.6)*, the robust risks admit the following lower bound*

$$\liminf_{\varepsilon\to0} \, \upsilon_\varepsilon^{2k/(2k+1)} \inf_{\widehat{S}_\varepsilon\in\mathcal{S}_\varepsilon} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) \ge \, l_*(\mathbf{r})\,, \tag{4.67}$$

where $\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon$.

We set the parameter δ in (4.27) as a function of ε, i.e. $\delta = \delta_\varepsilon$ is such that

$$\lim_{\varepsilon\to0} \, \delta_\varepsilon = 0 \quad \text{and} \quad \lim_{\varepsilon\to0} \, \varepsilon^{-\check{\delta}} \, \delta_\varepsilon = 0 \tag{4.68}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_\varepsilon = (6 + |\ln\varepsilon|)^{-1}$.

Theorem 4.5. *Assume that Conditions* (4.33) *hold. Then the model selection procedure admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to0} \, \upsilon_\varepsilon^{2k/(2k+1)} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le l_*(\mathbf{r})\,. \tag{4.69}$$

Theorem 4.4 and Theorem 4.5 imply the following result

Corollary 4.2. *Under the conditions of Theorem 4.5, we have*

$$\lim_{\varepsilon\to 0} \upsilon_\varepsilon^{2k/(2k+1)} \inf_{\widehat{S}_\varepsilon\in\mathcal{S}_\varepsilon} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) = l_*(\mathbf{r})\,. \tag{4.70}$$

Remark 4.2. It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball W^k_r is $\varepsilon^{-4k/(2k+1)}$ (see, for example, [47]). We see here that the efficient robust rate is $\upsilon_\varepsilon^{2k/(2k+1)}$, i.e. if the distribution upper bound $\varsigma^*_\varepsilon \to 0$ as $n\to\infty$ we obtain a faster rate with respect to $\varepsilon^{-4k/(2k+1)}$, and if $\varsigma^*_\varepsilon \to \infty$ as $\varepsilon\to 0$ we obtain a slower rate. In the case when ς^*_ε is constant the robust rate is the same as the classical non robust convergence rate.

4.5.1 Lower bound

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

$$\sup_{S\in W^k_r} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) \geq \sup_{S\in W^k_r} \mathcal{R}_Q(\widehat{S}_\varepsilon, S)\,. \tag{4.71}$$

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

$$d = d_\varepsilon = \left[\frac{k+1}{k}\upsilon_\varepsilon^{1/(2k+1)}\, l_*(\mathbf{r}_0)\right] \quad \text{and} \quad \mathbf{r}_0 = (1-\check{\gamma})\mathbf{r}\,. \tag{4.72}$$

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

$$S_z(x) = \sum_{j=1}^{d} z_j\,\phi_j(x)\,. \tag{4.73}$$

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

$$\kappa = (\kappa_j)_{1\leq j\leq d} \quad \text{and} \quad \kappa_j = s_j\,\eta_j\,, \tag{4.74}$$

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

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

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

$$\widetilde{\mathcal{R}}_Q(\widehat{S}) = \int_{\mathbb{R}^d} \mathcal{R}_Q(\widehat{S}, S_z)\,\mu_\kappa(\mathrm{d}z)\,. \tag{4.75}$$

Furthermore, for any function f, we denote by $\mathbf{p}(f)$ its projection in $\mathbf{L}_2[0,1]$ onto $W^k_{\mathbf{r}}$, i.e.

$$\|f-\mathbf{p}(f)\| = \inf_{h\in W^k_{\mathbf{r}}} \|f-h\|\,.$$

Since $W^k_{\mathbf{r}}$ is a convex and closed set in $\mathbf{L}_2[0,1]$, this projector exists and is unique for any function $f \in \mathbf{L}_2[0,1]$ and, moreover,

$$\|f-h\|^2 \geq \|\mathbf{p}(f)-h\|^2 \quad \text{for any} \quad h \in W^k_{\mathbf{r}}\,.$$

So, setting $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{S})$, we obtain that

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

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

$$\sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}_Q(\widehat{S},S) \;\geq\; \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon \tag{4.76}$$

and

$$\Delta_\varepsilon = \int_{\{z\in\mathbb{R}^d\,:\,S_z\notin W^k_{\mathbf{r}}\}} (\mathbf{r}+\|S_z\|^2)\,\mu_\kappa(\mathrm{d}z)\,.$$

Therefore, in view of (4.71)

$$\sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon,S) \geq \sup_{Q\in\mathcal{Q}^*_\varepsilon} \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon\,. \tag{4.77}$$

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

$$\lim_{\varepsilon\to 0} \varepsilon^{-\check{\delta}}\,\Delta_\varepsilon = 0\,. \tag{4.78}$$

Now it is easy to see that

$$\|\widehat{\mathbf{p}}-S_z\|^2 \geq \sum_{j=1}^d (\widehat{z}_j - z_j)^2\,,$$

where $\widehat{z}_j = \int_0^1 \widehat{\mathbf{p}}(t)\,\phi_j(t)\mathrm{d}t$. So, in view of Lemma 4.2 and reminding that $\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon$ we obtain

$$\begin{aligned}\sup_{Q\in\mathcal{Q}^*_\varepsilon} \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) \;&\geq\; \sup_{0<\varrho_1^2\leq\varsigma^*_\varepsilon} \sum_{j=1}^d \frac{1}{\varepsilon^{-2}\,\varrho_1^{-2}+\upsilon_\varepsilon\,(s^*_j)^{-1}} \\ &= \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d \frac{s^*_j}{s^*_j+1} = \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d\left(1-\frac{j^k}{d^k_\varepsilon}\right).\end{aligned}$$

Therefore, using now Definition (4.72), Inequality (4.77) and the limit (4.78), we obtain that

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

Taking here the limit as $\check{\gamma}\to 0$ implies the desired result . □

4.5.2 Upper bound

Known smoothness

First we suppose that the parameters $k \geq 1$, $\mathbf{r} > 0$ in (4.65) and ς_ε^* in (4.5) are known. Let the family of admissible weighted least square estimates $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ be given by (4.37). Consider the pair

$$\check{\alpha} = (k, \check{r}) \quad \text{and} \quad \check{r} = \varpi\,[\mathbf{r}/\varpi]\,,$$

where $\check{\varsigma}_\varepsilon = \varpi\varsigma_\varepsilon^*$ and ϖ satisfy the conditions in (4.33). Denote the corresponding estimate as

$$\check{S} = \widehat{S}_{\check{\lambda}} \quad \text{and} \quad \check{\lambda} = \lambda_{\check{\alpha}}\,. \tag{4.79}$$

Note that for sufficiently small ε the pair $\check{\alpha}$ belongs to the set (4.32).

Theorem 4.6. *The estimator $\check{S}$ admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to0}\, \upsilon_\varepsilon^{2k/(2k+1)} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon\left(\check{S}, S\right) \leq l_*(\mathbf{r})\,. \tag{4.80}$$

Proof. Substituting (4.14) and taking into account the definition (4.79) one gets

$$\|\check{S} - S\|^2 = \sum_{j=1}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 - 2\check{M}_\varepsilon + \varepsilon^2 \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,\check{\xi}_j^2\,,$$

where $\check{M}_\varepsilon = \varepsilon \sum_{j=1}^{\infty} (1-\check{\lambda}(j))\,\check{\lambda}(j)\,\theta_j\,\check{\xi}_j$. Note now that for any $Q \in \mathcal{Q}^*_\varepsilon$ the expectation $\mathbf{E}_{Q,S}\,\check{M}_\varepsilon = 0$ and, in view of the upper bound (4.12),

$$\sup_{Q\in\mathcal{Q}^*_\varepsilon} \mathbf{E}_{Q,S} \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,\check{\xi}_j^2 \leq \varsigma_\varepsilon^* \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,.$$

Therefore,

$$\mathcal{R}^*_\varepsilon(\check{S}, S) \leq \sum_{j=\check{\iota}}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 + \frac{1}{\upsilon_\varepsilon} \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,, \tag{4.81}$$

where $\check{j}_* = j_*(\check{\alpha})$. Setting

$$u_\varepsilon = (\upsilon_\varepsilon)^{2k/(2k+1)} \sup_{j\geq\check{j}_*} (1-\check{\lambda}(j))^2/a_j\,,$$

we obtain that for each $S \in W^k_{\mathbf{r}}$

$$\Upsilon_{1,\varepsilon}(S) = (\upsilon_\varepsilon)^{2k/(2k+1)} \sum_{j=\check{j}_*}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 \leq u_\varepsilon \sum_{j=\check{j}_*}^{\infty} a_j\,\theta_j^2 \leq u_\varepsilon\,\mathbf{r}\,.$$

Tazking into account that $\check{r} \to \mathbf{r}$, we obtain that

$$\limsup_{\varepsilon\to0} \sup_{S\in W^k_{\mathbf{r}}} \Upsilon_{1,\varepsilon}(S) \leq \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k}(\mathrm{d}_k)^{2k/(2k+1)}} := \Upsilon_1^*\,,$$

where the coefficient τ_k is given in (4.35). To estimate the last term in the right hand of (4.81), we set

$$\Upsilon_{2,\varepsilon} = \frac{1}{(\upsilon_\varepsilon)^{1/(2k+1)}} \sum_{j=1}^{+\infty} \check{\lambda}^2(j)\,.$$

It is easy to check that

$$\limsup_{\varepsilon\to 0}\, \Upsilon_{2,\varepsilon} \le \frac{2(\mathbf{r}\mathrm{d}_k)^{1/(2k+1)}\, k^2}{(k+1)(2k+1)} := \Upsilon_2^*\,.$$

Therefore, taking into account that, by the definition of the Pinsker's constant in (4.66), $\Upsilon_1^* + \Upsilon_2^* = l_*(\mathbf{r})$, we arrive at the inequality

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

Hence we obtain the desired result. □

Unknown smoothness

Combining Theorem 4.6 and Theorem 4.3 yields Theorem 4.5. □

4.6 Detection of the number of signals

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

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

where $(n_t)_{t\ge 0}$ is some noise, usually this "white noise". The energetic parameters $(\theta_j)_{j\ge 1}$ and the number of the signals q are unknown and the signals $(\phi_j)_{j\ge 1}$ are known orthonormal functions, i.e.

$$\int_0^1 \phi_i(t)\,\phi_j(t)\,\mathrm{d}t = \mathbf{1}_{\{i\ne j\}}\,.$$

The problem is to estimate q, when the signal/noise ratio goes to infinity. To describe this problem in the framework of the mathematical model we use the following stochastic differential equation

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \phi_j(t)\right) \mathrm{d}t + \varepsilon \mathrm{d}w_t\,, \tag{4.82}$$

where $(w_t)_{t\geq 0}$ is the standard Brownian motion and the parameter ε goes to zero. This means tha the signal/noise ratio goes to infinity. The logarithm of the likelihood ratio for the model (4.82) can be represented as

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

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

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

Therefore, the maximum likelihood estimate for $\widehat{q}=q^*$. So, if $q^*=\infty$ we obtain that $\widehat{q}=\infty$. So, this estimator does not work. For these reasons we propose to study the estimation problem for q for the process (4.82) in the nonparametric setting and to apply the model selection procedure (4.30). To this end we consider the model (4.1) with the unknown function S defined as

$$S(t)=\sum_{j=1}^{q}\theta_j\,\phi_j(t)\,. \tag{4.83}$$

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

$$\widehat{S}_d(x)=\sum_{j=1}^{d}\widehat{\theta}_{j,\varepsilon}\phi_j(x)\,. \tag{4.84}$$

This estimate can be obtained from (4.22) with the weights $\lambda_d(j)=\chi\{j\leq d\}$. The number of estimators ι is some function of ε, i.e. $\iota=\iota_\varepsilon$, such that

$$\lim_{\varepsilon\to 0}\iota_\varepsilon=\infty\quad\text{and}\quad\lim_{\varepsilon\to 0}\varepsilon^{\check{\delta}}\iota_\varepsilon=0 \tag{4.85}$$

for any $\check{\delta}>0$. As a risk for the signals number we use

$$\mathbf{D}_\varepsilon(d,q)=\mathcal{R}^*_\varepsilon(\widehat{S}_d,S)\,, \tag{4.86}$$

where the risk $\mathcal{R}^*_\varepsilon(\widehat{S},S)$ is defined in (4.2) and d is some integer number (maybe random) from the set $\{1,\ldots,\iota\}$. In this case the cost function (4.27) has the following form.

$$J_\varepsilon(d)=\sum_{j=1}^{d}\widehat{\theta}^2_{j,\varepsilon}-2\sum_{j=1}^{d}\widetilde{\theta}_{j,\varepsilon}+\delta\,\widehat{P}_\varepsilon(\lambda)\,. \tag{4.87}$$

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

$$\widehat{q}_\varepsilon=\mathrm{argmin}_{1\leq d\leq\iota}J_\varepsilon(d)\,. \tag{4.88}$$

Note that Theorem 4.3 implies that the robust risks of the procedure (4.30) with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfy the following oracle inequality

$$\mathbf{D}_\varepsilon(\widehat{q}_\varepsilon\,,\,q) \le \frac{1+3\delta}{1-3\delta}\min_{1\le d\le \iota}\mathbf{D}_\varepsilon(d,q) + \varepsilon^2\,\frac{\mathbf{U}^*_\varepsilon(S)}{\delta}\,, \tag{4.89}$$

where the last term satisfies the property (3.37).

4.7 Simulation

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (4.30). In (4.1) we chose

$$S(t) = \sum_{j=1}^{10}\frac{j}{j+1}\,\phi_j(t)\,, \tag{4.90}$$

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

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

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

We calculated the empirical quadratic risk defined as

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

and the relative quadratic risk

$$\overline{\mathbf{R}}_* = \overline{\mathbf{R}}/\|S\|_p^2\,.$$

The expectation is taken as an average over $N = 10000$ replications, i.e.

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

We used the cost function with

$$\delta = \frac{1}{(3+|\ln\varepsilon|)^2}\,.$$

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

Table 4.1: Empirical risks

In the following graphics the dashed line is the model selection procedure (4.30), the continuous line is the function (4.90) and the bold line is the corresponding observations (4.1).

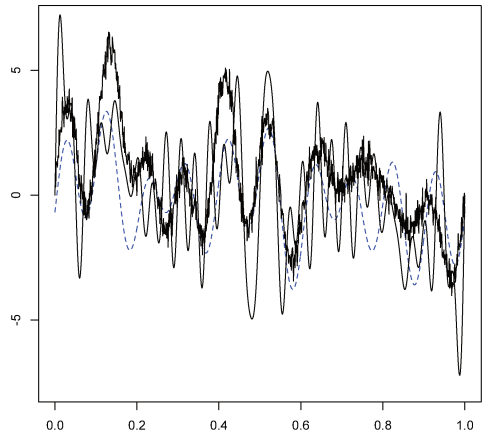

Figure 4.1: Estimator of S for $\varepsilon = 1/\sqrt{20}$

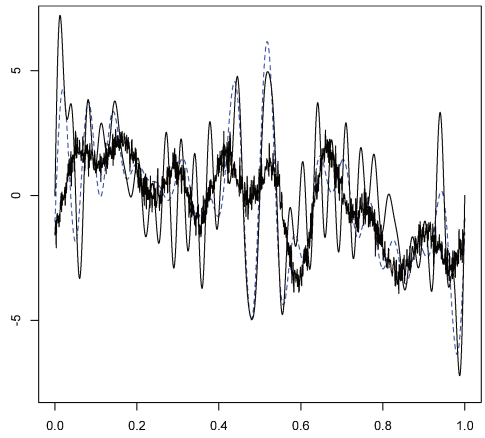

Figure 4.2: Estimator of S for $\varepsilon = 1/\sqrt{100}$

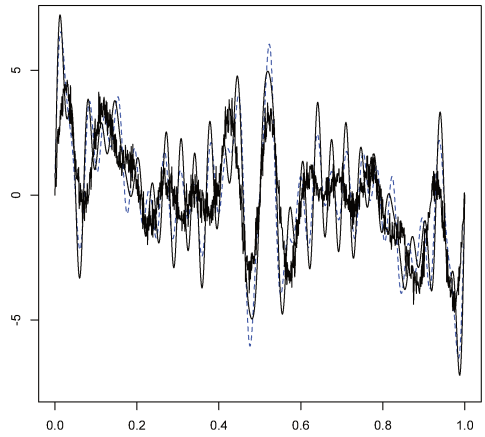

Figure 4.3: Estimator of S for $\varepsilon = 1/\sqrt{200}$

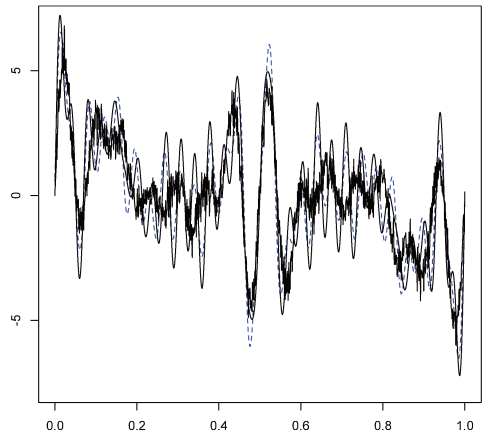

Figure 4.4: Estimator of S for $\varepsilon = 1/\sqrt{1000}$

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

Table 4.2: Estimation of the number of signals

To estimate the number of signals q we use two procedures. The first $\widehat{q}_1$ is (4.89) with $\nu = [\ln \varepsilon^{-2}]$. The second $\widehat{q}_2$ is defined through the shrinkage approach for the model selection procedure (4.90),

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

Remark 4.3. It should be noted that the LSE procedure (4.89) is more appropriate than the shrinkage method for such a number detection problem.

Now we give the algorithm of the model selection procedure given in Section 4.3

Algorithm 3 Model selection procedure

Require: $\varepsilon > 0$, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

ρ_1, ρ_2, ς_ε^*: satisfying Conditions (4.5) and (4.6)

$k^* \geq 1$, ϖ: satisfying Condition (4.33)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$n = [1/\varepsilon^2]$, $m = [1/\varpi]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varpi]$ *to* $[m\varpi]$ **do**

    for $k \longleftarrow 1$ *to* n **do**

      Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (4.35)

    end for

  end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varpi, \ldots, m\varpi\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,\varepsilon} = \int_0^1 \phi_k(t) \mathrm{d}\, \breve{y}_t$.

$\widetilde{\theta}_{k,\varepsilon} \longleftarrow \widehat{\theta}^2_{k,\varepsilon} - \varepsilon^2$.

The observation $(\breve{y}_t)_{0 \leq t \leq n}$ are given in (4.7) and $(\phi_k)_{k \geq 1}$ is the basis given in (4.9)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,\varepsilon}, ..., \widehat{\theta}_{n,\varepsilon})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,\varepsilon}, ..., \widetilde{\theta}_{n,\varepsilon})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varpi]$ *to* $[m\varpi]$ **do**

    $J_\varepsilon(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^n \lambda_{i,j}(l)\widetilde{\theta}_{j,\varepsilon} + \delta\, P_\varepsilon(\lambda)$.

    where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_ε is the penalty term given in (4.29)

  end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_\varepsilon(\lambda), \Lambda = \{\lambda_\alpha\,, \alpha \in \mathcal{A}\}$.

Appendix

Property of the penalty term

Lemma 4.1. *Assume that Proposition 4.1 holds. Then for any $n \geq 1$ and $\lambda \in \Lambda$,*

$$P_{\varepsilon}(\lambda) \leq \mathcal{R}(\widehat{S}_{\lambda}, S) + \varepsilon^2 \check{\varkappa}_Q \,,$$

where the coefficient $P_{\varepsilon}(\lambda)$ is defined in (4.43).

Proof. By the definition of $\mathrm{Err}_{\varepsilon}(\lambda)$ one has

$$\mathrm{Err}_{\varepsilon}(\lambda) = \sum_{j=1}^{n} \left((\lambda(j) - 1)\theta_j + \varepsilon^2 \lambda(j) \check{\xi}_j \right)^2 .$$

Through Proposition 4.1 it is easy to see that

$$\mathbf{E}_Q \, \mathrm{Err}_{\varepsilon}(\lambda) \geq \varepsilon^2 \sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q \, (\check{\xi}_j)^2 \geq P_{\varepsilon}(\lambda) - \varepsilon^2 \check{\varkappa}_Q \,.$$

Hence we obtain the desired result. □

Proof of the limit equality (4.78)

First, setting $\zeta_{\varepsilon} = \sum_{j=1}^{d} \kappa_j^2 \, a_j$, we obtain that

$$\left\{ S_{\kappa} \notin W_{\mathbf{r}}^{k} \right\} = \left\{ \zeta_{\varepsilon} > \mathbf{r} \right\} .$$

Moreover, note that one can check directly that

$$\lim_{\varepsilon \to 0} \mathbf{E} \, \zeta_{\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{v_{\varepsilon}} \sum_{j=1}^{d} s_j^* \, a_j = \check{\mathbf{r}} = (1 - \check{\gamma}) \mathbf{r} \,.$$

So, for sufficiently small ε we obtain that

$$\left\{ S_{\kappa} \notin W_{\mathbf{r}}^{k} \right\} \subset \left\{ \widetilde{\zeta}_{\varepsilon} > \mathbf{r}_1 \right\} ,$$

where $\mathbf{r}_1 = \mathbf{r}\check{\gamma}/2$, $\widetilde{\zeta}_{\varepsilon} = \zeta_{\varepsilon} - \mathbf{E} \, \zeta_{\varepsilon} = v_{\varepsilon}^{-1} \sum_{j=1}^{d} s_j^* a_j \widetilde{\eta}_j$ and $\widetilde{\eta}_j = \eta_j^2 - 1$ Through the correlation inequality (see, Proposition A.1 in [35]) we can get that for any $p \geq 2$

$$\mathbf{E} \, \widetilde{\zeta}_{\varepsilon}^{p} \leq (2p)^{p/2} \mathbf{E} |\widetilde{\eta}_1|^p \, v_{\varepsilon}^{-p} \left(\sum_{j=1}^{d} (s_j^*)^2 a_j^2 \right)^{p/2} = \mathrm{O}(\, v_{\varepsilon}^{-\frac{p}{4k+2}}) \,,$$

as $\varepsilon \to 0$. Therefore, for any $\check{\delta} > 0$ using the Chebychev inequality for $p > (4k+2)\check{\delta}$ we obtain that

$$v_{\varepsilon}^{\check{\delta}} \mathbf{P}(\widetilde{\zeta}_{\varepsilon} > \mathbf{r}_1) \to 0 \quad \text{as} \quad \varepsilon \to 0 \,.$$

Hence we obtain the equality (4.78). □

The absolute continuity of distributions for Lévy processes

In this section we study the absolute continuity for Lévy processes defined as

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le T\,, \tag{4.91}$$

where S is any arbitrary non-random square integrated function, i.e. from $\mathbf{L}_2[0,T]$ and $(\xi_t)_{0\le t\le T}$ is a Lévy process of the form (4.3) with nonzero constants ϱ_1 and ϱ_2. We denote by $\mathbf{P}_y$ and $\mathbf{P}_\xi$ the distributions of the processes $(y_t)_{0\le t\le 1}$ and $(\xi_t)_{0\le t\le 1}$ on the Skorokhod space $\mathbf{D}[0,T]$. Now for any $0 \le t \le T$ and $(x_t)_{0\le t\le T}$ from $\mathbf{D}[0,T]$ we set

$$\Upsilon_t(x) = \exp\left\{\int_0^t \frac{S(u)}{\varrho_1^2}\,\mathrm{d}x_u^c - \int_0^t \frac{S^2(u)}{2\varrho_1^2}\,\mathrm{d}u\right\}\,, \tag{4.92}$$

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

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

and, for any $t > 0$ and any measurable Γ from $\mathbb{R}\setminus\{0\}$,

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

Now we study the measures $\mathbf{P}_y$ and $\mathbf{P}_\xi$ in $\mathbf{D}[0,T]$.

Proposition 4.4. *For any $T > 0$, the measure $\mathbf{P}_y \ll \mathbf{P}_\xi$ in $\mathbf{D}[0,T]$ and the Radon-Nikodym derivative is*

$$\frac{\mathrm{d}\mathbf{P}_y}{\mathrm{d}\mathbf{P}_\xi}(\xi) = \Upsilon_T(\xi)\,.$$

Proof. Note that to show this proposition it suffices to check that for any $0 = t_0 < \ldots < t_n = T$ any $b_j \in \mathbb{R}$ for $1 \le j \le n$

$$\mathbf{E}\,\exp\left\{i\sum_{l=1}^n b_j(y_{t_j} - y_{t_{j-1}})\right\} = \mathbf{E}\,\exp\left\{i\sum_{l=1}^n b_j(\xi_{t_j} - \xi_{t_{j-1}})\right\}\Upsilon_T(\xi)\,.$$

Taking into account that the processes $(y_t)_{0\le t\le T}$ and $(\xi_t)_{0\le t\le T}$ have independent homogeneous increments, one needs to check only that for any $b \in \mathbb{R}$ and $0 \le s < t \le T$

$$\mathbf{E}\,\exp\{i\,b(y_t - y_s)\} = \mathbf{E}\,\exp\{i\,b(\xi_t - \xi_s)\}\,\frac{\Upsilon_t(\xi)}{\Upsilon_s(\xi)}\,. \tag{4.93}$$

To check this equality, note that the process

$$\Upsilon_t(\xi) = \exp\left\{\int_0^t \frac{S(u)}{\varrho_1}\,\mathrm{d}w_u - \int_0^t \frac{S^2(u)}{2\varrho_1^2}\,\mathrm{d}u\right\}$$

is a Gaussian martingale. From here we directly obtain Equation (4.93). □

The van Trees inequality for Lévy processes

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

$$\mathrm{d}y_t = S(t,\theta)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le 1\,, \tag{4.94}$$

where $S(t,\theta) = \sum_{i=1}^{d} \theta_i\,\phi_i(t)$ with the unknown parameters $\theta = (\theta_1,\dots,\theta_d)'$ and the process $(\xi_t)_{0\le t\le 1}$ is defined in (4.3). Note now that according to Proposition 4.4 the distribution $\mathbf{P}_\theta$ of the process (4.94) is absolutely continuous with respect to the $\mathbf{P}_\xi$ on $\mathbf{D}[0,1]$ and the corresponding Radon-Nikodym derivative is

$$f(x,\theta) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mathbf{P}_\xi}(x) = \exp\left\{\int_0^1 \frac{S(t,\theta)}{\varrho_1^2}\,\mathrm{d}x_t^c - \int_0^1 \frac{S^2(t,\theta)}{2\varrho_1^2}\,\mathrm{d}t\right\}, \tag{4.95}$$

where $x = (x_t)_{0\le t\le T}$ is an arbitrary function from $\mathbf{D}[0,1]$.

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

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

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

$$\lim_{|\theta_j|\to\infty} \mathbf{g}(\theta)\,\varphi_j(\theta_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |\mathbf{g}_j'(\theta)|\,\Phi(\theta)\,\mathrm{d}\theta < \infty\,, \tag{4.96}$$

where

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

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

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

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

Lemma 4.2. *For any $\mathcal{F}^y = \sigma\{y_t, 0 \le t \le 1\}$-measurable square integrable function $\widehat{\mathbf{g}}$ and for any $1 \le j \le d$, the following inequality holds*

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

where

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

Proof. First of all, note that the density (4.95) of the process ξ is bounded with respect to $\theta_j \in \mathbb{R}$ and, for any $1 \le j \le d$,

$$\limsup_{|\theta_j|\to\infty} f(\xi,\theta) \,=\, 0\,. \qquad \text{a.s.}$$

Now, we set

$$\widetilde{\Phi}_j = \widetilde{\Phi}_j(x,\theta) = \frac{\partial\,(f(x,\theta)\Phi(\theta))/\partial\theta_j}{f(x,\theta)\Phi(\theta)}\,.$$

Taking into account the condition (4.96) and integrating by parts yield

$$\widetilde{\mathbf{E}}\left((\widehat{\mathbf{g}}-\mathbf{g}(\theta))\widetilde{\Phi}_j\right) = \int_{\mathcal{X}\times\mathbb{R}^d} (\widehat{\mathbf{g}}(x)-\mathbf{g}(\theta))\frac{\partial}{\partial\theta_j}\,(f(x,\theta)\Phi(\theta))\,\mathrm{d}\theta\,\mathbf{P}_\xi(\mathrm{d}x)$$

$$= \int_{\mathcal{X}\times\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \mathbf{g}_j'(\theta)\,f(x,\theta)\Phi(\theta)\mathrm{d}\theta_j\right)\left(\prod_{i\neq j}\mathrm{d}\theta_i\right)\,\mathbf{P}_\xi(\mathrm{d}x) = \Lambda_j\,.$$

Now by the Cauchy's inequality we obtain the following lower bound for the quadratic risk

$$\widetilde{\mathbf{E}}(\widehat{\mathbf{g}}-\mathbf{g}(\theta))^2 \ge \frac{\Lambda_j^2}{\widetilde{\mathbf{E}}\Psi_j^2}\,.$$

To study the denominator in the left handside of this inequality note that, in view of the reprentation (4.95),

$$\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j} = \frac{1}{\varrho_1}\int_0^1 \psi_j(t)\,\mathrm{d}w_t\,.$$

Therefore, for each $\theta\in\mathbb{R}^d$,

$$\mathbf{E}_\theta\,\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j} \,=0$$

and

$$\mathbf{E}_\theta\left(\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j}\right)^2 = \frac{1}{\varrho_1^2}\int_0^1 \psi_j^2(t)\mathrm{d}t = \frac{1}{\varrho_1^2}\|\psi\|^2\,.$$

Taking into account that

$$\widetilde{\Phi}_j = \frac{1}{f(x,\theta)}\frac{\partial\,f(x,\theta)}{\partial\theta_j} + \frac{1}{\Phi(\theta)}\frac{\partial\,\Phi(\theta))}{\partial\theta_j}\,,$$

we get

$$\widetilde{\mathbf{E}}\Psi_j^2 = \frac{n}{\varrho_1^2}\,\|\psi\|^2 + \,I_j\,.$$

Hence we got the desired result. □

Chapter 5

Renewal theory

5.1 Renewal density

This section is concerned with results related to the renewal measure (1.8).

Theorem 5.1. *(Goldie's theorem)*
Let η be a probability law on $\mathbb{R}$ with finite second moment and positive first moment m, such that $\tilde{\eta}(\beta) < \infty$ for some $\beta > 0$. Suppose that η is spread out, so for some n_0 we have

$$\eta^{(n_0)} = (1-\delta)\phi_0 + \delta\phi_1,$$

where $\delta \in [0,1[$ is constant and ϕ_0, ϕ_1 are probability measures with ϕ_0 absolutely continuous. Suppose that β has been taken so small that $\delta\tilde{\phi}_1(\beta) < 1$. Suppose that $\hat{\eta}(\theta) \neq 1$ on the line $\Im\theta = -\beta$. Then the renewal measure $\nu := \sum_{j=0}^{\infty} \eta^{(n)}$ may be written $\nu = \nu_0 + \nu_1$, where ν_1 is a finite measure such that $\tilde{\nu_1}(\beta) < \infty$, and ν_0 is absolutlely continuous with a continuous bounded density $p(.)$ such that

$$p(t) = \frac{1}{m} - \frac{1}{2\pi}\int_C e^{-i\theta t}\frac{d\theta}{1-\hat{\eta}(\theta)} + o(e^{-\beta t}),\ t \longrightarrow \infty.$$

Here C is a simple closed contour in the domain $\mathbf{D} := \{\theta :\ -\beta < \Im\theta < 0\}$, enclosing all the zeroes of $1-\hat{\eta}$ in $\mathbf{D}$, $\tilde{\eta}(\theta) = \int_{\mathbb{R}} e^{\theta t}\eta(dt)$ and $\hat{\eta}(\theta) = \int_{\mathbb{R}} e^{i\theta t}\eta(dt)$.
The proof of this Theorem is given in [8].

Now we need to adapt this result to our framwork. We start with the following lemma.

Lemma 5.1. *Let τ be a positive random variable with a density g, such that $\mathbf{E}e^{\beta\tau} < \infty$ for some $\beta > 0$. Then there exists a constant β_1, $0 < \beta_1 \leq \beta$ for which,*

$$\mathbf{E}e^{(\beta_1 + i\omega)\tau} \neq 1 \qquad \forall\omega \in \mathbb{R}\,.$$

Proof. We will show this lemma by the contradiction, i.e. we assume that there exist some sequence of positive numbers going to zero $(\gamma_k)_{k\geq 1}$ and a sequence $(w_k)_{k\geq 1}$ such that

$$\mathbf{E}e^{(\gamma_k + i\omega_k)\tau} = 1 \tag{5.1}$$

for any $k \geq 1$. Firstly, assume that $\limsup_{k\to\infty} w_k = +\infty$. Note that in this case, for any $N \geq 1$,

$$\left|\int_0^N e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t\right| \leq \left|\int_0^N \cos(w_k t)\, g(t)\mathrm{d}t\right| + \left|\int_0^N (e^{\gamma_k t}-1)\cos(w_k t)\, g(t)\mathrm{d}t\right|,$$

i.e., in view of Lemma 5.2, for any fixed $N \geq 1$

$$\limsup_{k\to\infty} \int_0^N e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t = 0\,.$$

Since for some $\beta > 0$ the integral $\int_0^{+\infty} e^{\beta t}\, g(t)\mathrm{d}t < \infty$, we get

$$\lim_{k\to\infty} \int_0^{+\infty} e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t = 0\,.$$

Let now assume that $\limsup_{k\to\infty} w_k = \omega_\infty \neq 0$ and $0 < |\omega_\infty| < \infty$. In this case there exists a sequence $(l_k)_{k\geq 1}$ such that $\lim_{k\to\infty} w_{l_k} = \omega_\infty$, i.e.

$$1 = \limsup_{k\to\infty} \mathbf{E} e^{\gamma_{l_k}\tau}\cos(\tau w_{l_k}) = \mathbf{E}\,\cos(\tau w_\infty)\,.$$

It is clear that, for random variables having density, the last equality is possible if and only if $w_\infty = 0$. In this case, i.e. when $\limsup_{k\to\infty} w_{l_k} = 0$, the equation (5.1) implies

$$\limsup_{k\to\infty} \mathbf{E}\, e^{\gamma_{l_k}\tau}\frac{\sin(\tau w_{l_k})}{w_{l_k}} = \mathbf{E}\,\tau = 0\,.$$

But, under our conditions, $\mathbf{E}\tau > 0$. These contradictions imply the desired result. □

Proposition 5.1. *Let τ be a positive random variable with the distribution η having a density g which satisfies Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$). Then the renewal measure* (1.8) *is absolutely continuous with density ρ, for which*

$$\rho(x) = \frac{1}{\check{\tau}} + \Upsilon(x)\,, \tag{5.2}$$

where $\check{\tau} = \mathbf{E}\tau_1$ and $\Upsilon(\cdot)$ is some function defined on $\mathbb{R}_+$ with values in $\mathbb{R}$ such that

$$\sup_{x\geq 0} x^\gamma |\Upsilon(x)| < \infty \quad \textit{for all} \quad \gamma > 0\,.$$

Proof. First, note that we can represent the renewal measure $\check{\eta}$ as $\check{\eta} = \eta * \eta_0$ and $\eta_0 = \sum_{j=0}^\infty \eta^{(j)}$. It is clear that in this case the density ρ of $\check{\eta}$ can be written as

$$\rho(x) = \int_0^x g(x-y)\sum_{n\geq 0} g^{(n)}(y)\mathrm{d}y\,. \tag{5.3}$$

Now we use the arguments proposed in the proof of Lemma 9.5 from [8]. For any $0 < \epsilon < 1$ we set

$$\rho_\epsilon(x) = \int_0^x g(x-y) \left(\sum_{n\geq 0} (1-\epsilon)^n \, g^{(n)}(y) - \frac{(1-\epsilon)}{\check{\tau}} \, g_0(y) \right) \mathrm{d}y - g(x) \,, \tag{5.4}$$

where $g_0(y) = e^{-\epsilon y/\check{\tau}} 1_{\{y>0\}}$. It is easy to deduce that for any $x \in \mathbb{R}$

$$\lim_{\epsilon \to 0} \rho_\epsilon(x) \, = \, \rho(x) - \frac{1}{\check{\tau}} \int_0^x g(z) \, \mathrm{d}z - g(x) \,. \tag{5.5}$$

Moreover, in view of the condition $\mathbf{H}_1$) we obtain that the function $\rho_\epsilon(x)$ satisfies the condition **D**) from Section 5.2. So, through Proposition 5.3 we get

$$\rho_\epsilon(x+) + \rho_\epsilon(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{\rho}_\epsilon(\theta) \, \mathrm{d}\theta \,,$$

where $\widehat{\rho}_\epsilon(\theta) = \int_{\mathbb{R}} e^{i\theta x} \rho_\epsilon(x) \mathrm{d}x$. Note that

$$|\widehat{g}(\theta)| = \left| \int_{\mathbb{R}} e^{i\theta x} g(x) \mathrm{d}x \right| \leq \int_{\mathbb{R}} g(x) \mathrm{d}x = 1 \,,$$

i.e. for any $0 < \epsilon < 1$ we have $|(1-\epsilon)\widehat{g}(\theta)| < 1$ and therefore

$$\sum_{n=0}^{\infty} (1-\epsilon)^n (\widehat{g}(\theta))^n = \frac{1}{1-(1-\epsilon)\widehat{g}(\theta)} \,.$$

From this and, taking into account that

$$\widehat{g}_0(\theta) = \int_{\mathbb{R}} e^{i\theta x} g_0(x) \mathrm{d}x = \frac{\check{\tau}}{\epsilon - i\check{\tau}\theta} \,,$$

we obtain

$$\begin{aligned} \widehat{\rho}_\epsilon(\theta) &= \widehat{g}(\theta) \sum_{n=0}^{\infty} (1-\epsilon)^n (\widehat{g}(\theta))^n - \left(\frac{1-\epsilon}{\check{\tau}} \right) \widehat{g}(\theta) \widehat{g}_0(\theta) - \widehat{g}(\theta) \\ &= \widehat{g}(\theta) G_\epsilon(\theta) \quad \text{and} \quad G_\epsilon(\theta) = \frac{1}{1-(1-\epsilon)\widehat{g}(\theta)} - \frac{1-i\check{\tau}\theta}{\epsilon - i\check{\tau}\theta} \,, \end{aligned}$$

i.e.

$$\rho_\epsilon(x-) + \rho_\epsilon(x+) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{g}(\theta) G_\epsilon(\theta) \, \mathrm{d}\theta \,. \tag{5.6}$$

One can check directly that

$$\sup_{0<\epsilon<1, \theta \in \mathbb{R}} |G_\epsilon(\theta)| \, < \, \infty \,.$$

Therefore, using the condition $\mathbf{H}_3$) and the Lebesgue's dominated convergence theorem, we can pass to limit as $\epsilon \to 0$ in (5.6), i.e., we obtain that

$$\rho(x+) + \rho(x-) - \frac{2}{\check{\tau}} \int_0^x g(z) \, \mathrm{d}z - g(x+) - g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{g}(\theta) G_0(\theta) \, \mathrm{d}\theta \,,$$

where

$$G_0(\theta) = \frac{1}{1-\widehat{g}(\theta)} + \frac{1-i\check{\tau}\theta}{i\check{\tau}\theta}\,.$$

Using here again Proposition 5.3 we deduce that

$$\rho(x+) + \rho(x-) = \frac{2}{\check{\tau}} \int_0^x g(z)\,\mathrm{d}z + \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta \tag{5.7}$$

and

$$\check{G}(\theta) = \frac{1}{1-\widehat{g}(\theta)} + \frac{1}{i\check{\tau}\theta}\,.$$

Note now that we can represent the density (5.3) as

$$\rho(x) = g * \sum_{n\geq 0} g^{(n)} = \sum_{n\geq 1} g^{(n)}(x) = g(x) + \sum_{n\geq 2} g^{(n)}(x) =: g(x) + \rho_c(x)$$

and the function $\rho_c(x)$ is continuous for all $x \in \mathbb{R}$. This means that

$$\widetilde{\rho}(x) = \frac{\rho(x+) + \rho(x-)}{2} - \rho(x) = \frac{g(x+) + g(x-)}{2} - g(x)$$

and, therefore, the condition $\mathbf{H}_2$) implies that, for any $\gamma > 0$,

$$\sup_{x\geq 0} x^{\gamma}\,|\widetilde{\rho}(x)| < \infty.$$

Now we can rewrite (5.7) as

$$\rho(x) = \frac{1}{\check{\tau}} \int_0^x g(z)\,\mathrm{d}z + \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta - \widetilde{\rho}(x). \tag{5.8}$$

Taking into account that $\mathbf{E}\,e^{\beta\tau} < \infty$ for some $\beta > 0$ we can obtain that

$$\sup_{x\geq 0} x^{\gamma} \int_x^{+\infty} g(z)\,\mathrm{d}z < \infty\,.$$

To study the second term in (5.8) we will use Proposition 5.2. Indeed, Condition $\mathbf{H}_3$) implies the first limit equality in (5.10). The second one follows directly from Lemma 5.2. Therefore, in view of Proposition 5.2, there exists some $\beta^* > 0$ such that, for any $0 \leq \beta_0 \leq \beta^*$,

$$\int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta = e^{-\beta_0 x} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta - i\beta_0)\check{G}(\theta - i\beta_0)\,\mathrm{d}\theta\,.$$

Note that, due to Lemma 5.1, the function $1-\widehat{g}(\theta)$ has no zeros on the line $\{z \in \mathbb{C} : \mathrm{Im}(z) = -\beta_1\}$. Moreover, one can check directly that $\theta = 0$ is an isolated zero. So, this means that for any $N > 1$ there can be only finitely many zeros in $\{z \in \mathbb{C} : -\beta_1 < \mathrm{Im}(z) < 0\,,\, |\mathrm{Re}(z)| < N\}$ of the function $1 - \widehat{g}(\theta)$. Moreover, note that in view of lemma 5.2 for any $r > 0$

$$\lim_{\mathrm{Re}(\theta)\to\infty, |\mathrm{Im}(\theta)|\leq r} \widehat{g}(\theta) = 0\,.$$

This means that there exists $N > 0$ such that the function $1-\widehat{g}(\theta) \neq 0$ for $\theta \in \{z \in \mathbb{C} : -\beta_1 < \mathrm{Im}(z) < 0\,, |\mathrm{Re}(z)| \geq N\}$. So, there can be only finitely many zeros of the function $1 - \widehat{g}(\theta)$ in $\{z \in \mathbb{C} : -\beta_1 < Im(z) < 0\}$ for some fixed $0 < \beta_1 < \beta$. Therefore, there exists some $\beta_0 > 0$ for which the function $1 - \widehat{g}(\theta)$ has no zeros in $\{z \in \mathbb{C} : -\beta_0 < Im(z) < 0\}$, i.e. the function $\check{G}(\theta)$ will be bounded in this set and we obtain that

$$\sup_{x\geq 0} e^{\beta_0 x} \left| \int_{\mathbb{R}} e^{-ix\theta}\, \widehat{g}(\theta)\check{G}(\theta)\, \mathrm{d}\theta \right| < \infty\,.$$

Thus the conclusion follows. □

Using this proposition we can study the renewal process $(N_t)_{t\geq 0}$ introduced in (1.7).

Corollary 5.1. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any* $t > 0$,

$$\mathbf{E}\, N_t \leq |\rho|_*\, t \quad \textit{and} \quad \mathbf{E}\, N_t^2 \leq\ |\rho|_*\, t +\ |\rho|_*^2\, t^2\,. \tag{5.9}$$

Proof. First, by means of Proposition 5.1, note that we get

$$\mathbf{E}\, N_t = \mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} = \int_0^t \rho(v)\, \mathrm{d}v \leq |\rho|_*\, t\,.$$

Regarding the last bound in (5.9), we use the same reasoning as in the previous inequality, i.e., we obtain

$$\begin{aligned}\mathbf{E}\, N_t^2 &= \mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} + 2\mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} \sum_{j=k+1} \mathbf{1}_{\{T_j \leq t\}} \\ &= \mathbf{E}\, N_t + 2\mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}}\, \Theta(T_k) = \mathbf{E}\, N_t + \int_0^t \Theta(v)\, \rho(v)\, \mathrm{d}v\,,\end{aligned}$$

where, for $0 \leq v \leq t$, we defined the function $\Theta(v) = \mathbf{E}\, N_{t-v} \leq |\rho|_*(t - v)$. □

5.2 Properties of the Fourier transform

Theorem 5.2. *Cauchy (1825)*

Let U be a simply connected open subset of $\mathbb{C}$, let $g : U \to \mathbb{C}$ be a holomorphic function, and let γ be a rectifiable path in U whose start point is equal to its end point. Then

$$\oint_\gamma g(z)\mathrm{d}z = 0\,.$$

Proposition 5.2. *Let $g : \mathbb{C} \to \mathbb{C}$ be a holomorphic function in $U = \{z \in \mathbb{C} : -\beta_1 < Imz < \beta_2\}$ for some $\beta_1 > 0$ and $\beta_2 > 0$. Assume that, for any $-\beta_1 \leq t \leq 0$,*

$$\int_{\mathbb{R}} |g(\theta + it)|\, \mathrm{d}\theta < \infty \quad \textit{and} \quad \lim_{|\theta|\to\infty} g(\theta + it)\ = 0\,. \tag{5.10}$$

Then, for any $x \in \mathbb{R}$ and for any $0 < \beta < \beta_1$,

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta)\,\mathrm{d}\theta = e^{-\beta x} \int_{\mathbb{R}} e^{i\theta x} g(\theta - i\beta)\,\mathrm{d}\theta. \tag{5.11}$$

Proof. First note that the conditions of this theorem imply that

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta)\,\mathrm{d}\theta = \lim_{N\to\infty} \int_{-N}^{N} e^{i\theta x} g(\theta)\,\mathrm{d}\theta\,.$$

We fix now $0 < \beta < \beta_1$ and we set for any $N \geq 1$

$$\gamma = \{z \in \mathbb{C} : -N \leq \mathrm{Re}z \leq N\,,\, \mathrm{Im}z = 0\} \cup \{z \in \mathbb{C} : -N \leq \mathrm{Im}z \leq N\,,\, \mathrm{Re}z = N\}$$

$$\cup \{z \in \mathbb{C} : -N \leq \mathrm{Re}z \leq N\,,\, \mathrm{Im}z = -\beta\} \cup \{z \in \mathbb{C} : -\beta \leq \mathrm{Im}z \leq 0\,,\, \mathrm{Re}z = -N\}\,.$$

Now, in view of Theorem 5.2, we obtain that for any $N \geq 1$

$$\oint_{\gamma} e^{izx}\, g(z)\mathrm{d}z = \int_{-N}^{N} e^{i\theta x} g(\theta)\,\mathrm{d}\theta + \int_{0}^{-\beta} e^{i(N+it)x} g(N+it)\,\mathrm{d}t$$

$$+ \int_{N}^{-N} e^{i(-i\beta+\theta)x} g(-i\beta+\theta)\mathrm{d}\theta + \int_{-\beta}^{0} e^{i(-N+it)x} g(-N+it)\mathrm{d}t = 0\,. \tag{5.12}$$

Conditions (5.10) provide that

$$\lim_{N\to\infty} \int_{0}^{-\beta} e^{i(N+it)x} g(N+it)\,\mathrm{d}t = \lim_{N\to\infty} \int_{-\beta}^{0} e^{i(-N+it)x} g(-N+it)\,\mathrm{d}t = 0\,.$$

Therefore, letting $N \to \infty$ in (5.12) we obtain (5.11). Hence we get the desired result. □

The following technical lemma is also needed.

Lemma 5.2. *Let $g : [a,b] \to \mathbb{R}$ be a function from $\mathbf{L}_1[a,b]$. Then, for any fixed $-\infty \leq a < b \leq +\infty$,*

$$\lim_{N\to\infty} \int_{a}^{b} g(x)\,\sin(Nx)\mathrm{d}x = 0 \quad \textit{and} \quad \lim_{N\to\infty} \int_{a}^{b} g(x)\,\cos(Nx)\mathrm{d}x = 0\,. \tag{5.13}$$

Proof. Let first $-\infty < a < b < +\infty$. Assume that g is continuously differentiable, i.e. $g \in \mathbf{C}^1[a,b]$. Then integrating by parts gives us

$$\int_{a}^{b} g(x)\,\sin(Nx)\,\mathrm{d}x = \frac{1}{N}\left(g(b)\,\sin(Nb)\; - g(a)\,\sin(Na)\; - \int_{a}^{b} g^{'}(x)\,\cos(Nx)\,\mathrm{d}x\right).$$

So, from this we obtain that

$$\left|\int_{a}^{b} g(x)\,\sin(Nx)\,\mathrm{d}x\right| \leq \frac{|g(a)| + |g(a)| + (b-a)\max_{a\leq x\leq b}|g^{'}(x)|}{N}\,.$$

This implies the first limit in (5.13) for this case. The second one is obtained similarly. Let now g be any absolutely integrated function on $[a,b]$, i.e. $g \in \mathbf{L}_1[a,b]$. In this case there exists a sequence $g_n \in \mathbf{C}^1[a,b]$ such that

$$\lim_{n\to\infty} \int_a^b |g(x) - g_n(x)| \mathrm{d}x = 0\,.$$

Therefore, taking into account that for any $n \geq 1$

$$\lim_{N\to\infty} \int_a^b g_n(x)\, \sin(Nx) \mathrm{d}x = 0\,,$$

we obtain that

$$\limsup_{n\to\infty} \left| \int_a^b g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \int_a^b |g(x) - g_n(x)| \mathrm{d}x\,.$$

So, letting in this inequality $n \to \infty$ we obtain the first limit in (5.13) and, similarly, we obtain the second one. Let now $b = +\infty$ and $a = -\infty$. In this case we obtain that for any $-\infty < a < b < +\infty$

$$\left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| + \int_b^{+\infty} |g(x)\,| \mathrm{d}x + \int_{-\infty}^{a} |g(x)\,| \mathrm{d}x\,.$$

Using here the previous results we obtain that for any $-\infty < a < b < +\infty$

$$\limsup_{N\to\infty} \left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \int_b^{+\infty} |g(x)\,| \mathrm{d}x + \int_{-\infty}^{a} |g(x)\,| \mathrm{d}x\,.$$

Passing here to limit as $b \to +\infty$ and $a \to -\infty$ we obtain the first limit in (5.13). Similarly, we can obtain the second one. □

Let us now study the inverse Fourier transform. To this end, we need the following local Dini condition.

D) *Assume that, for some fixed $x \in \mathbb{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and there exists $\delta = \delta(x) > 0$ for which

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

Proposition 5.3. *Let $g : \mathbb{R} \to \mathbb{R}$ be a function from $\mathbf{L}_1(\mathbb{R})$. If, for some $x \in \mathbb{R}$, this function satisfies the condition* **D**, *then*

$$g(x+) + g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-i\theta x} \widehat{g}(\theta)\, \mathrm{d}\theta\,, \tag{5.14}$$

where

$$\widehat{g}(\theta) = \int_{\mathbb{R}} e^{i\theta t}\, g(t)\, \mathrm{d}t\,.$$

Proof. First, for any fixed $N > 0$ we set

$$J_N(x) = \frac{1}{2\pi} \int_{-N}^{N} e^{-i\theta x} \widehat{g}(\theta)\, \mathrm{d}\theta = \frac{1}{\pi} \int_{\mathbb{R}} g(z) \int_0^N \cos(\theta(z-x))\, \mathrm{d}\theta \mathrm{d}z\,,$$

i.e.,

$$J_N(x) = \frac{1}{\pi} \int_{\mathbb{R}} g(z)\, \frac{\sin(N(z-x))}{z-x}\, \mathrm{d}z = \frac{1}{\pi} \int_0^{\infty} (g(x+t) + g(x-t))\, \frac{\sin(Nt)}{t}\, \mathrm{d}t\,.$$

Taking into account that for any $N > 0$ the integral

$$\frac{2}{\pi} \int_0^{\infty} \frac{\sin(Nt)}{t}\, \mathrm{d}t = 1 \tag{5.15}$$

and denoting $B(x) = (g(x+) + g(x-))/2$, we obtain that

$$J_N(x) - B(x) = \frac{1}{\pi} \int_0^{\infty} \frac{\omega(x,t)\sin(Nt)}{t}\, \mathrm{d}t \quad \text{and} \quad \omega(x,t) = g(x+t) + g(x-t) - 2B(x)\,.$$

Now we represent the last integral as

$$\int_0^{\infty} \frac{\omega(x,t)\sin(Nt)}{t}\, \mathrm{d}t = I_{1,N} + I_{2,N} - 2B(x) I_{3,N}\,,$$

where

$$I_{1,N} = \int_0^{\delta} \frac{\omega(x,t)}{t}\, \sin(Nt)\mathrm{d}t\,, \quad I_{2,N} = \int_{\delta}^{\infty} G(t)\, \sin(Nt)\mathrm{d}t\,, \quad I_{3,N} = \int_{\delta}^{\infty} \frac{\sin(Nt)}{t}\, \mathrm{d}t$$

and $G(t) = (g(x+t) + g(x-t))/t$. Condition **D** and Lemma 5.2 imply directly the convergence $I_{1,N} \to 0$ as $N \to \infty$. Now note that, since $g \in \mathbf{L}_1(\mathbb{R})$, then the function G is absolutely integrated. Therefore, in view of Lemma 5.2, $I_{2,N} \to 0$ as $N \to \infty$. As to the last integral we use the property (5.15), i.e., the changing of the variables gives

$$I_{3,N} = \int_{\delta N}^{\infty} \frac{\sin t}{t}\, \mathrm{d}t \to 0 \quad \text{as} \quad N \to \infty\,.$$

Hence we have the desired result. □

Conclusion

The main purpose of this work is the non-parametric estimation for regression models in continuous time. First, we consider the problem of estimation an unknown fonction S on the basis of continuous observations, we define the noise in this model through a semi-Markov process which keeps the dependence for any duration n. So, we are in the case of dependent observations for which the dependence does not disapear for a sufficient large duration of observation. Second, we consider the same model when the estimation is based on discrete data and we obtain the sufficient conditions on the frequency observations under wich the robust effecient is shown. In the third model we consider a Lévy non-parametric regression with noise intensity and we estimate the unknown function S in the case where the noise level goes to 0 and the Lévy measure can go to infinity. In all of these models, we propose an adaptive model selection for the robust risk.

Bibliography

[1] V. S. Barbu, S. Beltaief and S. M. Pergamenchtchikov. Robust adaptive efficient estimation for semi - Markov nonparametric regression models. - Preprint, **15 April** (2016), *https://arxiv.org/pdf/1604.04516.pdf* submitted in *Statistical inference for stochastic processes*

[2] V. S. Barbu, S. Beltaief and S. M. Pergamenchtchikov. Robust adaptive efficient estimation for a semi - Markov continuous time regression from discrete data. - Preprint, **24 Mai** (2017),

[3] S. Beltaief, O. Chernoyaro and S. M. Pergamenchtchikov. Model selection for the robust efficient estimation in the Levy nonparametric regression. - Preprint, **23 Nov** (2016), *https://arxiv.org/pdf/1611.07378.pdf*

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

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

[6] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall, 2004.

[7] K. Bichteler and J. Jacod. *Calcul de Malliavin pour les diffusions avec sauts: existence d'une densité dans le cas unidimensionnel.* Séminaire de Probabilité, XVII, Lecture Notes in Math., **986**, Springer, Berlin, 132–157, 1983.

[8] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, **1** (1), 126–166, 1991.

[9] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation: Asymptotic Theory.* Springer, Berlin-New York, 1981.

[10] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes.* 2nd edition, Springer, Berlin, 2002.

[11] 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.

[12] 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.

[13] 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**, 215–235, 2003.

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

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

[16] 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**, 1083–1111, 2010.

[17] 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.

[18] V. V. Konev and S. M. Pergamenshchikov. Robust model selection for a semi-martingale continuous time regression from discrete data. *Stochastic Processes and their Applications*, **125**, 294–326, 2015.

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

[20] N. Limnios and G. Oprisan. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston, 2001.

[21] R. Sh. Liptser and A. N. Shiryaev. *Theory of Martingales*. Springer, 1986.

[22] T. Mikosch. *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes*. Springer, 2004.

[23] C. Mallows. Some comments on C_p. *Technometrics* **15**,661–675, 1973.

[24] C. Marinelli and M. Röckner. On maximal inequalities for purely discontinuous martingales in infinite dimensions. *Séminaire de Probabilités*, Lect. Notes Math., **XLVI**, 293–315, 2014.

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

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

[27] M. S. Pinsker. Optimal filtration of square integrable signals in Gaussian white noise. *Problems of Transmission Information*, **17**, 120–133, 1981.

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

[29] C. Mallows. Some comments on C_p. *Technometrics* **15**, 661–675, 1973.

[30] Akaike H. A new look at the statistical model identification. *IEEE Trans. on Automatic Control* **19**, 716–723, 1974.

[31] A. Barron, L. Birgé and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301–415, 1999.

[32] K. Bichteler and J. Jacod *Calcul de Malliavin pour les diffusions avec sauts: existence dune densité dans le cas unidimensionnel*. Séminaire de Probabilité, XVII, Lecture Notes in Math., **986**, Springer, Berlin, 132–157, 1983.

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

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

[35] L. Galtchouk, and S. Pergamenshchikov. Uniform concentration inequality for ergodic diffusion processes observed at discrete times. *Stochastic Processes and their Applications*, **123**, 91109, 2013.

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

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

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

[39] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection, 2009, *http://hal.archives-ouvertes.fr/hal-00326910/fr/.*

[40] A. Kneip. Ordered linear smoothers. *Annals of Statistcs* **22**, 835–866, 1994.

[41] Yu. A. Kutoyants. *Parameter Estimation for Stochastic Processes.* Heldeman-Verlag, Berlin, 1984.

[42] B. V. Manelis. Algorithms for tracking and demodulation of the mobile communication signal in the conditions of the Indiscriminating multipath propagation, *Radiotekhnika*, 4, 16–21, 2007 (in Russian).

[43] C. Marinelli and M. Röckner. On maximal inequalities for purely discontinuous martingales in infinite dimensions. *Séminaire de Probabilités*, Lect. Notes Math., **XLVI** , 293–315, 2014.

[44] A. A. Novikov. On discontinuous martingales. *Theory Probab. Appl.*, **20**, 1, 11–26, 1975.

[45] P. Massart. *A non-asymptotic theory for model selection.* European Congress of Mathematics, Eur. Math. Soc., Zrich, 2005.

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

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

[48] A. P. Trifonov and A. V. Kharin. Estimating the number of signals with unknown amplitudes . *Nelineinyi Mir*, **11** (12), 853 – 866, 2013 (in Russian).

[49] A. P. Trifonov, A. V. Kharin, O.V. Chernoyarov and K. S. Kalashnikov. Determining the number of radio signals with unknown phases. *International Journal on Communications Antenna and Propagation*, **5** (6), 367–374, 2015.

[50] A. P. Trifonov and Yu. S. Shinakov. *Joint distinction of signals and estimation of their parameters on the background noise*, Radio i Svyaz', Moscow, 1986 (in Russian).

[51] A. P. Trifonov and A. V. Kharin. Estimation of the number of radio signals with unknown amplitudes and phases. *Radiophysics and Quantum Electronics*, **58** (1), 57–70, 2015.

[52] A. P. Trifonov and A. V. Kharin . Estimation of the number of orthogonal signals with unknown non-energy parameters. *Radioelectronics and Communications Systems*, **58** (8), 362 – 370, 2015.

[53] A. P. Trifonov, A. V. Kharin and O. V. Chernoyarov. Estimation of the number of orthogonal signals with unknown parameters. *Journal of Communications Technology and Electronics*, **61** (9), 1026 –1033, 2016.

[54] L. I. Volkov, , M. S. Nemirovskii and Yu. S. Shinakov. *Digital Radio Communication Systems.* Eko-Trendz, Moscow, 2005 (in Russian).

[55] M. Wax and M. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Signal Process*, **33** (2), 387 – 392, 1985.

Résumé

Ce travail est consacré au problème d'estimation non paramétrique dans des modèles de regression en temps continu. On considère le problème d'estimation d'une fonction inconnue S supposée périodique. Cette estimation est basée sur des observations générées par un processus stochastique; ces observations peuvent êtres en temps continu ou discret. Pour ce faire, nous construisons une série d'estimateurs par projection et nous approchons la fonction inconnue S par une série de Fourier finie. Dans cette thèse, nous considérons le problème d'estimation dans le cadre adaptatif, c'est-à-dire le cas où la régularité de la fonction S est inconnue. Pour ce problème, nous développons une nouvelle méthode d'adaptation basée sur la procédure de sélection de modèle proposée par Konev et Pergamenshchikov (2012). Tout d'abord, cette procédure nous donne une famille d'estimateurs; après nous choisissons le meilleur estimateur possible en minimisant une fonction coût. Nous donnons également une inégalité d'Oracle pour le risque de nos estimateurs et nous donnons la vitesse de convergence minimax.

Mots Clés : Estimation non asymptotique, sélection de modèle, inégalité d'Oracle pointue, risque robuste , efficacitée asymptotique.

Abstract

This thesis is devoted to the problem of non parametric estimation for continuous-time regression models. We consider the problem of estimating an unknown periodoc function S. This estimation is based on observations generated by a stochastic process; these observations may be in continuous or discrete time. To this end, we construct a series of estimators by projection and thus we approximate the unknown function S by a finite Fourier series. In this thesis we consider the estimation problem in the adaptive setting, i.e. in situation when the regularity of the fonction S is unknown. In this way, we develop a new adaptive method based on the model selection procedure proposed by Konev and Pergamenshchikov (2012). Firstly, this procedure give us a family of estimators, then we choose the best possible one by minimizing a cost function. We give also an oracle inequality for the risk of our estimators and we give the minimax convergence rate.

Key Words: Non asymptotic estimation, Robust risk, Model selection, Sharp oracle inequality, Assymptotic efficiency.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problem of non parametric estimation in regression models has an important role in theorical and applied statistics. In this thesis, we consider regression models in continuous time of the form

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

where S is an unknown periodic function from $\mathbf{L}_2[0,n]$, wich we want to estimate on the basis of observations y_t. This observations can be in continuous time or in discrete time. This problem was considered in many frameworks, for example, in the framework of the "signal+white noise" models (see, for example, [9] or [47]). Later, in order to study dependent observations in continuous time, were introduced "signal+color noise" regressions based on Ornstein-Uhlenbeck processes (cf. [11], [12], [13], [16]).

Moreover, to include jumps in such models, the papers [17] and [18] used non Gaussian Ornstein-Uhlenbeck processes introduced in [4] for modeling the risky assets in the stochastic volatility financial markets. Unfortunately, the dependence of the stable Ornstein-Uhlenbeck type decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain very quickly the same "signal+white noise" model.

The main goal of this thesis is to develop the model (1.1) for the noise process with large dependence. This allows us to consider the signal estimation problem for signals observed with noises of complex structure "against-signal". To achieve this goal, we use semi-Markov processes to model the dependent noises, because it is well known that such processes keep the dependence for a long time.

In our work we use the robust estimation approach introduced in [17] for such problems. To this end, we denote by Q the distribution of $(\xi_t)_{0\le t\le n}$ in the Skorokhod space $\mathcal{D}[0,n]$. We assume that Q is unknown and belongs to some distribution family $\mathcal{Q}_n$ wich will be specified later. In this thesis we use the quadratic risk

$$\mathcal{R}_Q(\widetilde{S}_n, S) = \mathbf{E}_{Q,S}\, \|\widetilde{S}_n - S\|^2\,, \tag{1.2}$$

where $\|f\|^2 = \int_0^1 f^2(s)\mathrm{d}s$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1.1) corresponding to the noise distribution Q. Since the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

$$\mathcal{R}^*_n(\widetilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widetilde{S}_n, S)\,, \tag{1.3}$$

which enables us to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate $\widetilde{S}_n$ for all distributions in the family $\mathcal{Q}_n$.

In order to estimate the function S belonging to $\mathbf{L}_2[0,n]$, one can consider a projection estimator and thus approximate S by a finite Fourier series. Following Pinsker's approach (1981), we use the weighted least square estimators which provide efficient estimation, but, in order to obtain efficient estimation, one needs to use regularity properties of the function S. Our approach is to consider the estimation problem in the adaptive setting, i.e. in situation when the regularity of the function S is unknown. In this way, we develop a new adaptive method based on the model selection procedure proposed by Konev and Pergamenshchikov (2012). To show the efficiency, one needs to obtain the corresponding sharp oracle inequality; note that this is a crucial tool in order to be able to show the adaptive efficiency. The "sharp" oracle inequality means that the upper bound for the risk has the coefficient of the main term close to one.

We recall that the main term is the minimal risk over the family of estimators that we consider. To obtain the oracle inequality one needs to develop renewal theory for the model (1.1). In our thesis we obtain a new asymptotic development for the renewal density. In fact, this result is a version of Goldie's theorem (1991). Unfortunately, we cannot use directly the Goldie's theorem, since in that result there is a singular component of the renewal distribution, which makes the use of that result impossible for the estimation purposes. In our work we give sufficient conditions for having an asymptotic development for the renewal density without a singular component.

The effeciency of the estimator will be also proved. To this end, we assume that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W^k_{\mathbf{r}} = \{f \in \mathcal{C}^k_{per}[0,1] \, : \, \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq \mathbf{r}\}\,, \tag{1.4}$$

where $\mathbf{r} > 0$ and $k \geq 1$ are some unknown parameters, $\mathcal{C}^k_{per}[0,1]$ 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 $0 \leq i \leq k$. Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (1.3) is given by

$$\mathbf{r}^*_k = \, ((2k+1)\mathbf{r})^{1/(2k+1)} \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)} . \tag{1.5}$$

Note that this is the well-known Pinsker's constant obtained for the nonadaptive filtration problem in "signal + small white noise" model (see, for example, [47]), generated by the process (1.1).

Main results of the thesis

In this thesis we study three types of regression models in continuous time, the observations are generated mainly by a semi-Markov process and Lévy process. In this way, our model is

capable to take into account "small" jumps, thanks to the Lévy process, as well as "big" jumps, thanks to the semi-Markov process.

1.1 Semi-Markov process

In our work, we consider a pure jump process $(z_t)_{t\geq 0}$ as a semi-Markov process with the following form

$$z_t = \sum_{i=1}^{N_t} Y_i, \tag{1.6}$$

where $(Y_i)_{i\geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E}\, Y_i = 0\,, \quad \mathbf{E}\, Y_i^2 = 1 \quad \text{and} \quad \mathbf{E}\, Y_i^4 < \infty\,.$$

Here N_t is a general counting process (see, for example, [22]) defined as

$$N_t = \sum_{k=1}^{\infty} 1_{\{T_k \leq t\}} \quad \text{and} \quad T_k = \sum_{l=1}^{k} \tau_l\,, \tag{1.7}$$

where $(\tau_l)_{l\geq 1}$ is an i.i.d. sequence of positive integrated random variables with distribution η and mean $\check{\tau} = \mathbf{E}\,\tau_1 > 0$. We assume that the processes $(N_t)_{t\geq 0}$ and $(Y_i)_{i\geq 1}$ are independent. Note that the process $(z_t)_{t\geq 0}$ is a special case of a semi-Markov process (see, e.g., [5] and [20]). It should be noted that if τ_j are exponential random variables, then $(N_t)_{t\geq 0}$ is a Poisson process and, in this case, the process $(\xi_t)_{t\geq 0}$ given in (1.14) is a Lévy process for which this model has been studied in [14], [15] and [17]. But, in the general case when the process (1.6) is not a Lévy process, this process has a memory and cannot be treated in the framework of semi-martingales with independent increments. In this case, we need to develop new tools based on renewal theory arguments, what we do in Chapter.5. This tools will be intensively used in the proofs of our main results.
Let ρ be the density of the renewal measure $\check{\eta}$ defined as

$$\check{\eta} = \sum_{l=1}^{\infty} \eta^{(l)}\,, \tag{1.8}$$

where $\eta^{(l)}$ is the lth convolution power of η. To study this renewal measure we assume that the measure η has a density g which satisfies the following conditions.

$\mathbf{H}_1$) *Assume that, for any $x \in \mathbb{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and, for any $K > 0$, *there exists* $\delta = \delta(K) > 0$ *for which*

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

$\mathbf{H}_2$) *For any* $\gamma > 0$,

$$\sup_{z\ge 0} z^\gamma |2g(z) - g(z-) - g(z+)| \, < \, \infty.$$

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

Remark 1.1. It should be noted that the condition $\mathbf{H}_3$) means that there exists an exponential moment for the random variable $(\tau_j)_{j\ge 1}$, i.e. these random variables are not too large. This is a natural constraint since these random variables define the intervals between jumps, i.e., the frequency of the jumps. So, to study the influence of the jumps in the model (1.13) one needs to consider the noise process (1.14) with "small" interval between jumps or large jump frequency.

For the next condition we need to introduce the Fourier transform of any function f from $\mathbf{L}_1(\mathbb{R})$, $f : \mathbb{R} \to \mathbb{R}$, defined as

$$\widehat{f}(\theta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\theta x}\, f(x)\, \mathrm{d}x. \tag{1.9}$$

$\mathbf{H}_4$) *There exists* $t^* > 0$ *such that the function* $\widehat{g}(\theta - it)$ *belongs to* $\mathbf{L}_1(\mathbb{R})$ *for any* $0 \le t \le t^*$.

Remark 1.2. It is clear that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true for any continuously differentiable function g, for example for the exponential density.

Proposition 1.1. *Let* τ *be a positive random variable with the distribution* η *having a density* g *which satisfies Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$). *Then the renewal measure* (1.8) *is absolutely continuous with density* ρ, *for which*

$$\rho(x) = \frac{1}{\check{\tau}} + \Upsilon(x)\,, \tag{1.10}$$

where $\check{\tau} = \mathbf{E}\tau_1$ *and* $\Upsilon(\cdot)$ *is some function defined on* $\mathbb{R}_+$ *with values in* $\mathbb{R}$ *such that*

$$\sup_{x\ge 0} x^\gamma |\Upsilon(x)| < \infty \quad \textit{for all} \quad \gamma > 0\,.$$

It should be noted that in view of this proposition, Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) imply

$$|\rho|_* = \sup_{t\ge 0} |\rho(t)| < \infty \quad \text{and} \quad \|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)|\, \mathrm{d}x \, < \infty \tag{1.11}$$

1.2 Lévy process

In this thesis we assume that the Lévy process L_t is defined as

$$L_t = \breve{\varrho}\, w_t + \sqrt{1-\breve{\varrho}^2}\, \check{L}_t\,, \quad \check{L}_t = x * (\mu - \widetilde{\mu})_t\,, \tag{1.12}$$

where, $0 \le \breve{\varrho} \le 1$ is an unknown constant, $(w_t)_{t\ge 0}$ is a standard Brownian motion, $\mu(\mathrm{d}s, \mathrm{d}x)$ is the jump measure with the deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, where $\Pi(\cdot)$ is some positive measure on $\mathbb{R}$ (see, for example, [10, 6] for details). In this thesis, we use the usual notation

$$\Pi(|x|^m) = \int_{\mathbb{R}} |z|^m\, \Pi(\mathrm{d}z) \quad \text{for any} \quad m > 0.$$

Note that $\Pi(\mathbb{R})$ may be equal to $+\infty$.

Remark 1.3. In this thesis, we assume that the processes $(N_t)_{t\ge 0}$ and $(Y_i)_{i\ge 1}$ in (1.6) are independent between them and are also independent of $(L_t)_{t\ge 0}$.

1.3 Non-parametric estimation based on continuous data

We consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{1.13}$$

where S is an unknown 1-periodic function from $\mathbf{L}_2[0,1]$ defined on $\mathbb{R}$ with values in $\mathbb{R}$, the noise process $(\xi_t)_{t\ge 0}$ is defined as

$$\xi_t = \varrho_1 L_t + \varrho_2 z_t\,, \tag{1.14}$$

where ϱ_1 and ϱ_2 are unknown coefficients, the pure jump process $(z_t)_{t\ge 1}$ is the semi-Markov process defined in (1.6) and $(L_t)_{t\ge 0}$ is the Levy process defined in (1.12), for which we assume that

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

The problem is to estimate the unknown function S in the model (1.13) on the basis of observations $(y_t)_{0\le t\le n}$. The main goal is to consider continuous time regression models with dependent observations for which the dependence does not disappear for a sufficient large duration of observations. To this end we define the noise in the model through a semi-Markov process which keeps the dependence for any duration n. This type of models allows, for example, to estimate the signals observed under long impulse noise impact with a memory or "against signals".
To estimate the function S we use here the model selection procedure for continuous time

regression models from [17] based on the Fourrier expansion. We recall that for any function S from $\mathbf{L}_2[0,1]$ we can write

$$S(t) = \sum_{j=1}^{\infty} \theta_j \, \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j) = \int_0^1 S(t)\phi_j(t)\mathrm{d}t \, ,$$

where $(\phi_j)_{j\geq 1}$ is an orthonormal uniformly bounded basis in $\mathbf{L}_2[0,1]$, i.e., for some constant $\phi_* \geq 1$, which may be depend on n,

$$\sup_{0\leq j\leq n} \sup_{0\leq t\leq 1} |\phi_j(t)| \, \leq \, \phi_* < \infty \, .$$

We extend the functions $\phi_j(t)$ by periodicity, i.e., we set $\phi_j(t) := \phi_j(\{t\})$, where $\{t\}$ is the fractional part of $t \geq 0$. For example, we can take the trigonometric basis $(\mathrm{Tr}_j)_{j\geq 1}$ in $\mathbf{L}_2[0,1]$ defined as $\mathrm{Tr}_1 \equiv 1$ and, for $j \geq 2$,

$$\mathrm{Tr}_j(x) = \sqrt{2} \begin{cases} \cos(2\pi[j/2]x) & \text{for even} \quad j, \\ \\ \sin(2\pi[j/2]x) & \text{for odd} \quad j, \end{cases} \tag{1.15}$$

where $[x]$ denotes the integer part of x.
Now, for obtaining efficient estimation one needs to use weighted least square estimators defined as

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(t) \, ,$$

where the coefficients $\lambda = (\lambda(j))_{1\leq j\leq n}$ belong to some finite set Λ from $[0,1]^n$. In this thesis we consider the adaptive case, i.e. we assume that the regularity of the function S is unknown. In this case we chose the weight coefficients on the basis of the model selection procedure proposed in [17] for the general semi-martingale regression model in continuous time.
Now, to choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}} \, . \tag{1.16}$$

Now, in order to study the asymptotic effeciency for our procedure, we give the following oracle inequality for the robust risk defined in (1.3) and through a specific distribution family.

Theorem 1.1. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_4$) *hold and that the unknown function S is continuously differentiable. Then, for any $n \geq 1$ and $0 < \delta < 1/6$, the procedure* (1.16) *satisfy the following oracle inequality*

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

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that, under some conditions, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\le r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0.$$

It follows directly, by using the oracle inequality, that our procedure is efficient with the efficient robust rate $\upsilon_n^{2k/(2k+1)}$ with $\upsilon_n = n/\varsigma^*(n)$ such that, for any $\check{\epsilon} > 0$,

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

Corollary 1.1.

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^*\,. \tag{1.17}$$

We prove in this thesis that the robust minimax convergence rate may be faster or slower than the classical one $(n^{2k/(2k+1)})$.

1.4 Non-parametric estimation based on discrete data

In this chapter we consider the regression model (1.13) with the noise (1.14). The problem is to estimate the unknown function S in the case when continuous observation cannot be provided and only discret time measurement are available, the observations are in the forme

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

where the integer $p \ge 1$ is the observation frequency. We will use the trigonometric basis defined in (1.15). By making use of this basis we consider the discrete Fourier transformation of S

$$S(t) = \sum_{j=1}^{p} \theta_{j,p}\,\mathrm{Tr}_j(t), \quad t \in \{t_1, ..., t_p\},$$

where the Fourier coefficients are defined by

$$\theta_{j,p} = (S, \mathrm{Tr}_j)_p = \frac{1}{p}\sum_{i=1}^{p} S(t_i)\mathrm{Tr}_j(t_i).$$

Then, we estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\Psi_{j,p}(t)\,,$$

where the weight vector $\lambda = (\lambda(1), ..., \lambda(n))$ belongs to some finite set Λ from $[0,1]^n$,

$$\widehat{\theta}_{j,p} = \frac{1}{n}\int_0^n \Psi_{j,p}(t)\mathrm{d}\,y_t \quad \text{and} \quad \Psi_{j,p}(t) = \sum_{l=1}^{np} \mathrm{Tr}_j(t_l)1_{\{t_{l-1}<t\le t_l\}}\,.$$

In order to find a proper weight sequence λ in the set Λ one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. Let us consider the empirical squared error

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}} . \tag{1.18}$$

Now, To obtain the oracle inequality and to prove the effeciency of our procedure, we obtain in this thesis the sufficient condition for the frequency observations p.
$\mathbf{H}_5$) We assume that there exists $\check{\delta} > 0$ such that for any $n \geq 3$

$$p \geq n^{5/6} .$$

Now, in order to study the asymptotic effeciency for our procedure, we give the following oracle inequality for the robust risk defined in (1.3) and through a specific distribution family.

Theorem 1.2. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_5$) *hold and that the unknown function* S *is continuously differentiable. Then, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, the procedure* (1.18) *satisfy the following oracle inequality*

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

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that, under some conditions, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\leq r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0.$$

1.5 Non-parametric estimation for Lévy regression models

We consider a regression model in continuous time with the Lévy noise

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \varepsilon\, \mathrm{d}\xi_t , \quad 0 \leq t \leq 1 , \tag{1.19}$$

where S is an unknown function defined on $\mathbb{R}$ with values in $\mathbb{R}$, $(\xi_t)_{0\leq t\leq 1}$ is a Lévy process and $\varepsilon > 0$ is the noise intensity. The problem is to estimate the function S based on the continuous observations $(y_t)_{0\leq t\leq 1}$ when $\varepsilon \to 0$. We assume that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^4) < \infty . \tag{1.20}$$

When constructing the sharp model selection procedures, in this model, we will use the approach close to that of the papers [14], [15], [16], [18] developed for the estimation of a 1-periodic function in continuous time on a large time interval, i.e.

$$\mathrm{d}x_t = S(t)\mathrm{d}t + \mathrm{d}\eta_t\,, \quad 0 \le t \le n\,.$$

Note that, for any $0 < t < 1$, setting $y_t = n^{-1}\sum_{j=1}^{n}(x_{t+j} - x_j)$, we can represent this model as a model with small parameter of form (1.19)

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon\,\mathrm{d}\xi_t\,,$$

where $\varepsilon = n^{-1/2}$ and $\xi_t = n^{-1/2}\sum_{j=1}^{n}(\eta_{t+j} - \eta_t)$. The main difference between this model and the original one is that the jumps are small, i.e.

$$\Delta\xi_t = \mathrm{O}(n^{-1/2}) = \mathrm{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0\,,$$

but we have not such property in the model (1.19). Therefore, unfortunately, we cannot use directly the method developed for the estimation problem on the large time interval to the model (1.19). So, the main goal of this chapter is to develop a new sharp model selection method for the estimation problem of the function S as $\varepsilon \to 0$. Let $(\phi_j)_{j\ge 1}$ be an orthonormal basis in $\mathbf{L}_2[0,1]$ with $\phi_1 \equiv 1$. We assume that this basis is uniformly bounded, i.e. for some constant $\phi^* \ge 1$, which may depend on $\varepsilon > 0$,

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

where $n = n_\varepsilon = [1/\varepsilon^2]$ and $[x]$ denotes the integer part of x. To estimate the function S we use the following Fourier series

$$S(t) = \sum_{j\ge 1}\theta_j\,\phi_j(t).\,,$$

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

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

where $n = [1/\varepsilon^2]$, the weights $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$. To choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}\,. \tag{1.21}$$

Theorem 1.3. *Assume that for the model* (1.19) *the condition* (1.20) *holds and the unknown function S is continuously differentiable. Then, for any $0 < \delta < 1/6$, the procedure* (1.21) *satisfy the following oracle inequality*

$$\mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*_\varepsilon(\widehat{S}_\lambda, S) + \varepsilon^2 \, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta}\,, \tag{1.22}$$

where the term $\mathbf{U}^*_\varepsilon(S) > 0$ is such that under some conditions, for any $r > 0$ and $\check{\delta} > 0$

$$\lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \sup_{\|\dot{S}\|\le r} \mathbf{U}^*_\varepsilon(S) \, = 0\,. \tag{1.23}$$

As an application of the sharp model selection method in this thesis, we consider the estimation problem for the number of signals in the multi-path connexion channel. In the framework of the statistical radio - physics models we study the telecommunication system in which we observe in the multi-path channel the summarized signal with the noise on the time interval $[0,1]$,

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

where $(n_t)_{t\ge 0}$ is some noise, usually this "white noise". The energetic parameters $(\theta_j)_{j\ge 1}$ and the number of the signals q are unknown and the signals $(\phi_j)_{j\ge 1}$ are known orthonormal functions, i.e.

$$\int_0^1 \phi_i(t)\,\phi_j(t)\,\mathrm{d}t = \mathbf{1}_{\{i\ne j\}}\,.$$

The problem is to estimate q, when the signal noise ratio goes to infinity. To describe this problem in the framework of the mathematical model we use the following stochastic differential equation

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \phi_j(t)\right) \mathrm{d}t + \varepsilon \mathrm{d}w_t\,, \tag{1.24}$$

where $(w_t)_{t\ge 0}$ is the standard Brownian motion and the parameter ε goes to zero. This means that the ratio signal/noise goes to infinity. The logarithm of the likelihood ratio for the model (1.24) can be represented as

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

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

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

Therefore, the maximum likelihood estimate for $\widehat{q} = q^*$. So, if $q^* = \infty$ we obtain that $\widehat{q} = \infty$. So, this estimator does not work. For this reason we propose to study the estimation problem

for q for the process (1.24) in the nonparametric setting and to apply the model selection procedure (1.21).

Chapter 2

Non-parametric estimation for semi-Markov regression models based on continuous data

2.1 Introduction

Let us consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{2.1}$$

where S is an unknown 1-periodic function from $\mathbf{L}_2[0,1]$ defined on $\mathbb{R}$ with values in $\mathbb{R}$, the noise process $(\xi_t)_{t\ge 0}$ is defined as

$$\xi_t = \varrho_1 L_t + \varrho_2 z_t\,, \tag{2.2}$$

where ϱ_1 and ϱ_2 are unknown coefficients, the pure jump process $(z_t)_{t\ge 1}$ is the semi-Markov process defined in (1.6) and $(L_t)_{t\ge 0}$ is the Lévy process defined in (1.12), for which we assume that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^8) < \infty\,. \tag{2.3}$$

The problem is to estimate the unknown function S in the model (2.1) on the basis of observations $(y_t)_{0\le t\le n}$. Firstly, this problem was considered in the framework of the "signal+white noise" models (see, for example, [9] or [47]). Later, in order to study dependent observations in continuous time, were introduced "signal+color noise" regressions based on Ornstein-Uhlenbeck processes (cf. [11], [12], [13], [16]).

Moreover, to include jumps in such models, the papers [17] and [18] used non Gaussian Ornstein-Uhlenbeck processes introduced in [4] for modeling of the risky assets in the stochastic volatility financial markets. Unfortunately, the dependence of the stable Ornstein-Uhlenbeck type decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain very quickly the same "signal+white noise" model.

The main goal of this chapter is to consider continuous time regression models with dependent observations for which the dependence does not disappear for a sufficient large duration of observations. To this end, we define the noise in the model (2.1) through a semi-Markov process which keeps the dependence for any duration n. This type of models allows, for example, to estimate the signals observed under long impulse noise impact with a memory or "against signals".

In this chapter we use the robust estimation approach introduced in [17] for such problems. To this end, we denote by Q the distribution of $(\xi_t)_{0\le t\le n}$ in the Skorokhod space $\mathcal{D}[0,n]$. We assume that Q is unknown and belongs to some distribution family $\mathcal{Q}_n$ specified in Section 2.2. In this chapter we use the quadratic risk

$$\mathcal{R}_Q(\widetilde{S}_n, S) = \mathbf{E}_{Q,S}\,\|\widetilde{S}_n - S\|^2\,, \tag{2.4}$$

where $\|f\|^2 = \int_0^1 f^2(s)\mathrm{d}s$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (2.1) corresponding to the noise distribution Q. Since the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

$$\mathcal{R}^*_n(\widetilde{S}_n, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widetilde{S}_n, S)\,, \tag{2.5}$$

which enables us to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate $\widetilde{S}_n$ for all distributions in the family $\mathcal{Q}_n$.

To summarize, the goal of this chapter is to develop robust efficient model selection methods for the model (2.1) with the semi-Markov noise having unknown distribution, based on the approach proposed by Konev and Pergamenshchikov in [17] and [18] for continuous time regression models with semi-martingale noises. Unfortunately, we cannot use directly this method for semi-Markov regression models, since their tool essentially uses the fact that the Ornstein-Uhlenbeck dependence decreases with geometrical rate and the "white noise" case is obtained sufficiently quickly.

Thus in the thesis we propose new analytical tools based on renewal methods to obtain the sharp non-asymptotic oracle inequalities. As a consequence, we obtain the robust efficiency for the proposed model selection procedures in the adaptive setting.

2.2 Model selection

Note that for any function f from $\mathbf{L}_2[0,n]$, $f:[0,n]\to\mathbb{R}$, for the noise process $(\xi_t)_{t\ge 0}$ defined in (2.2), with $(z_t)_{t\ge 0}$ given in (1.6), the integral

$$I_n(f) = \int_0^n f(s)\mathrm{d}\xi_s \tag{2.6}$$

is well defined with $\mathbf{E}_Q\, I_n(f) = 0$. Moreover, as it is shown in Lemma 2.2,

$$\mathbf{E}_Q\, I_n^2(f) \le \varkappa_Q \int_0^n f^2(s)\mathrm{d}\,s\,, \tag{2.7}$$

where $\varkappa_Q = \varrho_1^2 + \varrho_2^2\,|\rho|_*$ and $|\rho|_* = \sup_{t\ge 0}|\rho(t)| < \infty$. Let us define the family of the noise distributions for the model (2.1) which is used in the robust risk (2.5). Note that any distribution Q from $\mathcal{Q}_n$ is defined by the unknown parameters in (2.2) and (1.12). We assume that

$$\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau} \le \varsigma^*, \tag{2.8}$$

where the unknown bounds ς^* are functions of n, i.e. $\varsigma^* = \varsigma^*(n)$, such that for any $\check{\epsilon} > 0$,

$$\lim_{n\to\infty} n^{\check{\epsilon}}\,\varsigma^*(n) = +\infty \quad \text{and} \quad \lim_{n\to\infty} \frac{\varsigma^*(n)}{n^{\check{\epsilon}}} = 0\,. \tag{2.9}$$

Remark 2.1. As we will see later, the parameter σ_Q is the limit of the Fourier transform of the noise process (2.2). Such limit is called variance proxy (see [17]).

Remark 2.2. Note that, generally (but it is not necessary) the parameters ϱ_1 and ϱ_2 can be dependent on n. The conditions (2.9) mean that we consider all possible cases, i.e. these parameters may go to infinity or be constant or go to zero as well. See, for example, the conditions (3.32) in [18].

Now, let $(\phi_j)_{j\geq 1}$ be an orthonormal uniformly bounded basis in $\mathbf{L}_2[0,1]$, i.e., for some constant $\phi_* \geq 1$, which may depend on n,

$$\sup_{0\leq j\leq n}\ \sup_{0\leq t\leq 1} |\phi_j(t)| \,\leq\, \phi_* < \infty\,. \tag{2.10}$$

We extend the functions $\phi_j(t)$ by periodicity, i.e., we set $\phi_j(t) := \phi_j(\{t\})$, where $\{t\}$ is the fractional part of $t \geq 0$. For example, we can take the trigonometric basis defined in (1.15). To estimate the function S we use here the model selection procedure for continuous time regression models from [17], based on the Fourrier expansion. We recall that, for any function S from $\mathbf{L}_2[0,1]$, we can write

$$S(t) = \sum_{j=1}^{\infty} \theta_j\,\phi_j(t) \quad \text{and} \quad \theta_j = (S,\phi_j) = \int_0^1 S(t)\phi_j(t)\mathrm{d}t\,. \tag{2.11}$$

So, to estimate the function S it suffices to estimate the coefficients θ_j and to replace them in this representation by their estimators. Using the fact that the function S and ϕ_j are 1 - periodic, we can write that

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

If we replace here the differential $S(t)\mathrm{d}t$ by the stochastic observed differential $\mathrm{d}y_t$ we obtain the natural estimate for θ_j on the time interval $[0,n]$

$$\widehat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\mathrm{d}\,y_t\,, \tag{2.12}$$

which can be represented, in view of the model (2.1), as

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j)\,. \tag{2.13}$$

Now (see, for example, [9]) we can estimate the function S by the projection estimators, i.e.

$$\widehat{S}_m(t) = \sum_{j=1}^{m} \widehat{\theta}_{j,n}\,\phi_j(t)\,, \quad 0\leq t\leq 1\,, \tag{2.14}$$

for some number $m \to \infty$ as $n \to \infty$. It should be noted that Pinsker in [47] shows that the projection estimators of the form (2.14) are not efficient. For obtaining efficient estimation one needs to use weighted least square estimators defined as

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,n}\phi_j(t)\,, \tag{2.15}$$

where the coefficients $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$. As it is shown in [47], in order to obtain efficient estimators, the coefficients $\lambda(j)$ in (2.15) need to be chosen depending on the regularity of the unknown function S. In this thesis we consider the adaptive case, i.e. we assume that the regularity of the function S is unknown. In this case we chose the weight coefficients on the basis of the model selection procedure proposed in [17] for the general semi-martingale regression model in continuous time. These coefficients will be obtained later in (2.28). To this end, first we set

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = 1 + \max_{\lambda\in\Lambda} \check{L}(\lambda)\,, \tag{2.16}$$

where $\#(\Lambda)$ is the cardinal number of Λ and $\check{L}(\lambda) = \sum_{j=1}^{n}\lambda(j)$. Now, to choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

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

which in our case is equal to

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n}\lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n}\lambda(j)\widehat{\theta}_{j,n}\theta_j + \sum_{j=1}^{\infty}\theta_j^2. \tag{2.17}$$

Since the Fourier coefficients $(\theta_j)_{j\ge1}$ are unknown, we replace the terms $\widehat{\theta}_{j,n}\theta_{j,n}$ by

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{\widehat{\sigma}_n}{n}\,, \tag{2.18}$$

where $\widehat{\sigma}_n$ is an estimate for the variance proxy σ_Q defined in (2.8). If it is known, we take $\widehat{\sigma}_n = \sigma_Q$; otherwise, we can choose it, for example, as in [17], i.e.

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \widehat{\mathbf{T}}^2_{j,n}\,, \tag{2.19}$$

where $\widehat{\mathbf{T}}_{j,n}$ are the estimators for the Fourier coefficients $(\mathbf{T}_j)_{j\ge1}$ with respect to the trigonometric basis (1.15), i.e.

$$\widehat{\mathbf{T}}_{j,n} = \frac{1}{n}\int_0^n Tr_j(t)\mathrm{d}y_t\,. \quad \text{and} \quad \mathbf{T}_j = \int_0^1 Tr_j(t)\,S(t)\mathrm{d}t\,. \tag{2.20}$$

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

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda), \tag{2.21}$$

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

$$P_n(\lambda) = \frac{\widehat{\sigma}_n |\lambda|^2}{n}. \tag{2.22}$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,, \tag{2.23}$$

where

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_n(\lambda). \tag{2.24}$$

We recall that the set Λ is finite so $\widehat{\lambda}$ exists. In the case when $\widehat{\lambda}$ is not unique, we take one of them. Let us now specify the weight coefficients $(\lambda(j))_{1\le j\le n}$. Consider, for some fixed $0 < \varepsilon < 1$, a numerical grid of the form

$$\mathcal{A} = \{1, \dots, k^*\} \times \{\varepsilon, \dots, m\varepsilon\}\,, \tag{2.25}$$

where $m = [1/\varepsilon^2]$. We assume that both parameters $k^* \ge 1$ and ε are functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that

$$\begin{cases} \lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0\,, \\ \\ \lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\check{\delta}}\varepsilon(n) = +\infty, \end{cases} \tag{2.26}$$

for any $\check{\delta} > 0$. One can take, for example, for $n \ge 2$,

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k^*_0 + \sqrt{\ln n}\,, \tag{2.27}$$

where $k^*_0 \ge 0$ is some fixed constant and the threshold $\varsigma^*(n)$ is introduced in (2.8). For each $\alpha = (\beta, l) \in \mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha = (\lambda_\alpha(j))_{1\le j\le n}$$

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\le j<j_*\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}, \tag{2.28}$$

where $j_* = 1 + [\ln \upsilon_n]$, $\omega_\alpha = (\mathrm{d}_\beta\, l\upsilon_n)^{1/(2\beta+1)}$,

$$\mathrm{d}_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad \upsilon_n = n/\varsigma^*. \tag{2.29}$$

Now we define the set Λ as

$$\Lambda \,=\, \{\lambda_{\alpha}\,,\, \alpha \in \mathcal{A}\}\,. \tag{2.30}$$

It will be noted that in this case the cardinal of the set Λ is

$$\check{\iota} = k^{*}m\,. \tag{2.31}$$

Moreover, taking into account that $\mathrm{d}_{\beta} < 1$ for $\beta \geq 1$ we obtain for the set (2.30)

$$|\Lambda|_{*} \,\leq\, 1 + \sup_{\alpha\in\mathcal{A}} \omega_{\alpha} \leq 1 + (\upsilon_{n}/\varepsilon)^{1/3}\,. \tag{2.32}$$

Remark 2.3. Note that the form (2.28) for the weight coefficients in (2.15) was proposed by Pinsker in [47] for the efficient estimation in the nonadaptive case, i.e. when the regularity parameters of the function S are known. In the adaptive case these weight coefficients are used in [17, 18] to show the asymptotic efficiency for model selection procedures.

2.3 Oracle inequality

In this section we obtain in Theorem 2.2 the non-asymptotic oracle inequality for the quadratic risk (2.4) for the model selection procedure (2.23) and in Theorem 2.3 the non-asymptotic oracle inequality for the robust risk (2.5) for the same model selection procedure (2.23), considered with the coefficients (2.28).

In order to prove the oracle inequality, the following conditions will be needed for the noise $(\xi_t)_{t\geq 0}$. Here we use the conditions introduced in [17] for the general semi-martingale model (2.1).

$\mathbf{C}_1$) *For all $n \geq 1$ and Q there exist a variance proxy $\sigma_Q > 0$ and the constant $\mathbf{C}_{1,Q,n} \geq 0$ such that, for any basis functions with the bound* (2.10),

$$\sup_{x\in[-1,1]^n} \left|B_{1,Q,n}(x)\right| \,\leq\, \mathbf{C}_{1,Q,n} < \infty\,,$$

where $B_{1,Q,n}(x) = \sum_{j=1}^{n} x_j \,\Big(\mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q\Big)$.

$\mathbf{C}_2$) *For all $n \geq 1$ and Q there exists a constant $\mathbf{C}_{2,Q,n} \geq 1$ such that, for any basis functions with the bound* (2.10),

$$\sup_{|x|\leq 1} \mathbf{E}_Q\, B_{2,Q,n}^2(x) \,\leq\, \mathbf{C}_{2,Q,n} < \infty,$$

where $|x|^2 = \sum_{j=1}^{n} x_j^2$ *and* $B_{2,Q,n}(x) = \sum_{j=1}^{n} x_j \,\Big(\xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2\Big)$.

Before stating the non-asymptotic oracle inequality, let us first introduce the following parameters which will be used for describing the rest term in the oracle inequalities. For the

renewal density ρ defined in (1.8) we set

$$\Upsilon(x) = \rho(x) - \frac{1}{\check{\tau}} \quad \text{and} \quad \|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)|\,\mathrm{d}x\,, \tag{2.33}$$

where $\check{\tau} = \mathbf{E}\,\tau_1$. In Proposition 5.1 we show that $|\rho|_* = \sup_{t\geq 0}|\rho(t)| < \infty$ and $\|\Upsilon\|_1 < \infty$. So, using this, we can introduce the following parameters

$$\Psi_Q = 4\varkappa_Q \check{\iota} + \left(5 + \frac{4\check{\iota}}{\sigma_Q}\right)\left(\sigma_Q\,\check{\tau}\,\phi^2_{max}\,\|\Upsilon\|_1 + \phi^4_{max}(1+\sigma_Q^2)^3\,\check{l}\right) \tag{2.34}$$

and

$$\mathbf{c}^*_Q = \sigma_Q + 2\varkappa_Q + \sigma_Q\,\check{\tau}\,\phi^2_{max}\,\|\Upsilon\|_1 + \phi^4_{max}(1+\sigma_Q^2)^2\,\check{l}\,, \tag{2.35}$$

where $\check{l} = (4\check{\tau}^2+8)\,\|\Upsilon\|_1 + 5 + 13(1+\check{\tau})^2(1+|\rho|^2_*)(\mathbf{E}Y_1^4) + 4\Pi(x^4)$. First, let us state the non-asymptotic oracle inequality for the quadratic risk (2.4) for the model selection procedure (2.23).

Theorem 2.1. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold. Then, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, the estimator of* S *given in* (2.23) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{\Psi_Q + 10|\Lambda|_*\,\mathbf{E}_S|\widehat{\sigma}_n - \sigma_Q|}{n\delta}\,. \tag{2.36}$$

Proof. First, note that we can rewrite the empirical squared error in (2.17) as follows

$$\mathrm{Err}_n(\lambda) = J_n(\lambda) + 2\sum_{j=1}^{\infty}\lambda(j)\check{\theta}_{j,n} + ||S||^2 - \delta P_n(\lambda), \tag{2.37}$$

where $\check{\theta}_{j,n} = \widetilde{\theta}_{j,n} - \theta_j\widehat{\theta}_{j,n}$. Using the definition of $\widetilde{\theta}_{j,n}$ in (2.18) we obtain that

$$\check{\theta}_{j,n} = \frac{1}{\sqrt{n}}\theta_j\xi_{j,n} + \frac{1}{n}\widetilde{\xi}_{j,n} + \frac{1}{n}\varsigma_{j,n} + \frac{\sigma_Q - \widehat{\sigma}_n}{n}\,,$$

where $\varsigma_{j,n} = \mathbf{E}_Q\xi^2_{j,n} - \sigma_Q$ and $\widetilde{\xi}_{j,n} = \xi^2_{j,n} - \mathbf{E}_Q\xi^2_{j,n}$. Putting

$$M(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\lambda(j)\theta_j\xi_{j,n} \quad \text{and} \quad P^0_n = \frac{\sigma_Q|\lambda|^2}{n}\,, \tag{2.38}$$

we can rewrite (2.37) as

$$\begin{aligned}\mathrm{Err}_n(\lambda) =& J_n(\lambda) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n}\,\check{L}(\lambda) + 2M(\lambda) + \frac{2}{n}B_{1,Q,n}(\lambda)\\ &+ 2\sqrt{P^0_n(\lambda)}\frac{B_{2,Q,n}(e(\lambda)}{\sqrt{\sigma_Q n}} + \|S\|^2 - \rho P_n(\lambda),\end{aligned} \tag{2.39}$$

where $e(\lambda) = \lambda/|\lambda|$, the function $\check{L}(\cdot)$ is defined in (2.16) and the functions $B_{1,Q,n}(\cdot)$ and $B_{2,Q,n}(\cdot)$ are given in conditions $\mathbf{C}_1$) and $\mathbf{C}_2$).

Let $\lambda_0 = (\lambda_0(j))_{1\le j\le n}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be as in (2.24). Substituting λ_0 and $\widehat{\lambda}$ in Equation (2.39), we obtain

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) - \mathrm{Err}_n(\lambda_0) =& J(\widehat{\lambda}) - J(\lambda_0) + 2\frac{\sigma_Q - \widehat{\sigma}_Q}{n}\,\check{L}(\varpi) + \frac{2}{n}B_{1,Q,n}(\varpi) + 2M(\varpi)\\ &+ 2\sqrt{P_n^0(\widehat{\lambda})}\frac{B_{2,Q,n}(\widehat{e})}{\sqrt{\sigma_Q n}} - 2\sqrt{P_n^0(\lambda_0)}\frac{B_{2,Q,n}(e_0)}{\sqrt{\sigma_Q n}}\\ &- \delta P_n(\widehat{\lambda}) + \delta P_n(\lambda_0), \end{aligned} \tag{2.40}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{e} = e(\widehat{\lambda})$ and $e_0 = e(\lambda_0)$. Note that, by (2.16),

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

Applying the inequality

$$2|ab| \le \delta a^2 + \delta^{-1}b^2 \tag{2.41}$$

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

$$2\sqrt{P_n^0(\lambda)}\frac{|B_{2,Q,n}(e(\lambda))|}{\sqrt{\sigma_Q n}} \le \delta P_n^0(\lambda) + \frac{B_{2,Q,n}^2(e(\lambda))}{\delta\sigma_Q\, n}.$$

Taking into account the bound (2.59), we get

$$\begin{aligned}\mathrm{Err}_n(\hat{\lambda}) \le& \mathrm{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta\sigma_Q\, n}\\ &+ \frac{1}{n}|\widehat{\sigma} - \sigma_Q|(|\widehat{\lambda}|^2 + |\lambda_0|^2) + 2\delta P_n(\lambda_0)\,,\end{aligned}$$

where $B_{2,Q,n}^* = \sup_{\lambda\in\Lambda} B_{2,Q,n}^2((e(\lambda))$. Moreover, noting that in view of (2.16) $\sup_{\lambda\in\Lambda}|\lambda|^2 \le |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) \le& \mathrm{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta\sigma_Q n}\\ &+ \frac{4|\Lambda|_*}{n}|\widehat{\sigma} - \sigma_Q| + 2\delta P_n(\lambda_0).\end{aligned} \tag{2.42}$$

To estimate the second term in the right side of this inequality we set

$$S_x = \sum_{j=1}^{n} x(j)\theta_j\phi_j\,, \quad x = (x(j))_{1\le j\le n} \in \mathbb{R}^n\,.$$

Thanks to (2.7) we estimate the term $M(x)$ for any $x \in \mathbb{R}^n$ as

$$\mathbf{E}_Q M^2(x) \le \varkappa_Q \frac{1}{n}\sum_{j=1}^{n} x^2(j)\theta_j^2 = \varkappa_Q \frac{1}{n}||S_x||^2. \tag{2.43}$$

To estimate this function for a random vector $x \in \mathbb{R}^n$ we set

$$Z^* = \sup_{x \varepsilon \Lambda_1} \frac{nM^2(x)}{||S_x||^2}, \quad \Lambda_1 = \Lambda - \lambda_0 .$$

So, through Inequality (2.41), we get

$$2|M(x)| \le \delta ||S_x||^2 + \frac{Z^*}{n\delta}. \tag{2.44}$$

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

$$\mathbf{E}_Q Z^* \le \sum_{x \in \Lambda_1} \frac{n\mathbf{E}_Q M^2(x)}{||S_x||^2} \le \sum_{x \in \Lambda_1} \varkappa_Q = \varkappa_Q \check{\iota} , \tag{2.45}$$

where $\check{\iota} = \mathrm{card}(\Lambda)$. Moreover, note that, for any $x \in \Lambda_1$,

$$||S_x||^2 - ||\widehat{S}_x||^2 = \sum_{j=1}^{n} x^2(j)(\theta_j^2 - \widehat{\theta}_j^2) \le -2M_1(x), \tag{2.46}$$

where $M_1(x) = n^{-1/2} \sum_{j=1}^{n} x^2(j)\theta_j \xi_{j,n}$. Taking into account that, for any $x \in \Lambda_1$ the components $|x(j)| \le 1$, we can estimate this term as in (2.43), i.e.,

$$\mathbf{E}_Q M_1^2(x) \le \varkappa_Q \frac{||S_x||^2}{n} .$$

Similarly to the previous reasoning we set

$$Z_1^* = \sup_{x \varepsilon \Lambda_1} \frac{nM_1^2(x)}{||S_x||^2}$$

and we get

$$\mathbf{E}_Q Z_1^* \le \varkappa_Q \check{\iota} . \tag{2.47}$$

Using the same type of arguments as in (2.44), we can derive

$$2|M_1(x)| \le \delta ||S_x||^2 + \frac{Z_1^*}{n\delta}. \tag{2.48}$$

From here and (2.46), we get

$$||S_x||^2 \le \frac{||\widehat{S}_x||^2}{1-\delta} + \frac{Z_1^*}{n\delta(1-\delta)} \tag{2.49}$$

for any $0 < \delta < 1$. Using this bound in (2.44) yields

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

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

$$2M(\varpi) \le \frac{2\delta(\mathrm{Err}_n(\widehat{\lambda}) + \mathrm{Err}_n(\lambda_0))}{1-\delta} + \frac{Z^* + Z_1^*}{n\delta(1-\delta)}.$$

Using this bound in (2.42) we obtain

$$\mathrm{Err}_n(\widehat{\lambda}) \leq \frac{1+\delta}{1-3\delta}\mathrm{Err}_n(\lambda_0) + \frac{Z^* + Z_1^*}{n\delta(1-3\delta)} + \frac{2\mathbf{C}_{1,Q,n}}{n(1-3\delta)} + \frac{2B^*_{2,Q,n}}{\delta(1-3\delta)\sigma_Q n}$$
$$+ \frac{(4|\Lambda|_* + 2)}{n(1-3\delta)}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).$$

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

$$\mathrm{Err}_n(\widehat{\lambda}) \leq \frac{1+\delta}{1-3\delta}\mathrm{Err}_n(\lambda_0) + \frac{2(Z^* + Z_1^*)}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4B^*_{2,Q,n}}{\delta\sigma_Q n}$$
$$+ \frac{(8|\Lambda|_* + 2)}{n}|\widehat{\sigma}_n - \sigma_Q| + \frac{2\delta}{(1-3\delta)}\, P_n^0(\lambda_0).$$

In view of Proposition 2.3 we estimate the expectation of the term $B^*_{2,Q,n}$ in (2.42) as

$$\mathbf{E}_Q\, B^*_{2,Q,n} \leq \sum_{\lambda\in\Lambda} \mathbf{E}_Q B^2_{2,Q,n}(e(\lambda)) \leq \check{\iota}\mathbf{C}_{2,Q,n}\,.$$

Taking into account that $|\Lambda|_* \geq 1$, we get

$$\mathcal{R}(\widehat{S}_*, S) \leq \frac{1+\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\iota}}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4\check{\iota}\mathbf{C}_{2,Q,n}}{\delta\sigma_Q n}$$
$$+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).$$

Using the upper bound for $P_n(\lambda_0)$ in Lemma 2.6, one obtains (2.36), that finishes the proof. □

Now we study the estimate (2.19).

Proposition 2.1. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold and that the function S is continuously differentiable. Then, for any $n \geq 2$,*

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \sigma_Q| \leq \frac{6\|\dot{S}\|^2 + \mathbf{c}_Q^*}{\sqrt{n}}\,. \tag{2.50}$$

Proof. We use here the same method as in [14]. First of all note that Definition (2.20) implies that

$$\widehat{\mathbf{T}}_{j,n} = \mathbf{T}_j + \frac{1}{\sqrt{n}}\,\eta_{j,n}\,, \tag{2.51}$$

where

$$\mathbf{T}_j = \int_0^1 S(t)\, Tr_j(t)\mathrm{d}t \quad \text{and} \quad \eta_{j,n} = \frac{1}{\sqrt{n}}\int_0^n \mathrm{Tr}_j(t)\,\mathrm{d}\xi_t\,.$$

So, we have

$$\widehat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j^2 + 2M_n + \frac{1}{n}\sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2\,, \tag{2.52}$$

where

$$M_n = \frac{1}{\sqrt{n}} \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j \, \eta_{j,n} \, .$$

Note that, for continuously differentiable functions (see, for example, Lemma A.6 in [14]), the Fourier coefficients $(\mathbf{T}_j)_{\{j\geq 1\}}$ satisfy the following inequality, for any $n \geq 1$,

$$\sum_{j=[\sqrt{n}]+1}^{\infty} \mathbf{T}_j^2 \leq \frac{4\left(\int_0^1 |\dot{S}(t)|\mathrm{d}t\right)^2}{\sqrt{n}} \leq \frac{4\|\dot{S}\|^2}{\sqrt{n}} \, . \tag{2.53}$$

In the same way as in (2.43) we estimate the term M_n, i.e.,

$$\mathbf{E}_Q \, M_n^2 \leq \frac{\varkappa_Q}{n} \sum_{j=[\sqrt{n}]+1}^{n} \mathbf{T}_j^2 \leq \frac{4\varkappa_Q \|\dot{S}\|^2}{n\sqrt{n}},$$

while the absolute value of this term for $n \geq 1$ can be estimated as

$$|\mathbf{E}_Q \, M_n| \leq \frac{\varkappa_Q + \|\dot{S}\|^2}{\sqrt{n}} \, .$$

Moreover, using Propositions 2.2 and 2.3 we can represent the last term in (2.52) as

$$\frac{1}{n} \sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2 = \frac{\sigma_Q(n-\sqrt{n})}{n} + \frac{B_{1,Q,n}(x')}{n} + \frac{B_{2,Q,n}(x'')}{\sqrt{n}},$$

with $x'_j = \mathbf{1}_{\{\sqrt{n}<j\leq n\}}$ and $x''_j = \mathbf{1}_{\{\sqrt{n}<j\leq n\}}/\sqrt{n}$. Therefore,

$$\mathbf{E}_Q \left| \frac{1}{n} \sum_{j=[\sqrt{n}]+1}^{n} \eta_{j,n}^2 - \sigma_Q \right| \leq \frac{\sigma_Q}{\sqrt{n}} + \frac{\mathbf{C}_{1,Q,n}}{n} + \frac{\sqrt{\mathbf{C}_{2,Q,n}}}{\sqrt{n}}.$$

Taking into account that $\mathbf{C}_{2,Q,n} \geq 1$, we obtain the bound (2.50) and hence the desired result. □

Theorem 2.1 and Proposition 2.1 implies the following result.

Theorem 2.2. *Assume that Conditions* $\mathbf{C}_1$*) and* $\mathbf{C}_2$*) hold and that the function* S *is continuously differentiable. Then, for any* $n \geq 1$ *and* $0 < \delta \leq 1/6$*, the procedure* (2.23), (2.19) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{60\widetilde{\Lambda}_n \, \|\dot{S}\|^2 + \widetilde{\Psi}_{Q,n}}{n\delta} \, , \tag{2.54}$$

where $\widetilde{\Psi}_{Q,n} = 10\widetilde{\Lambda}_n \mathbf{c}_Q^* + \Psi_Q$ *and* $\widetilde{\Lambda}_n = |\Lambda|_*/\sqrt{n}$.

Remark 2.4. Note that the coefficient $\varkappa_Q$ can be estimated as $\varkappa_Q \le (1+\check{\tau}|\rho|_*)\sigma_Q$. Therefore, taking into account that $\phi^4_{max} \ge 1$, the remainder term in (2.54) can be estimated as

$$\widetilde{\Psi}_{Q,n} \le \mathbf{C}_* \left(1+\sigma_Q^6+\frac{1}{\sigma_Q}\right)(1+\widetilde{\Lambda}_n)\check{\iota}\phi^4_{max}\,, \tag{2.55}$$

where $\mathbf{C}_* > 0$ is some constant which is independent of the distribution Q.

Furthermore, let us study the robust risk (2.5) for the procedure (2.23). In this case, the distribution family $\mathcal{Q}_n$ consists in all distributions on the Skorokhod space $\mathcal{D}[0,n]$ of the process (2.2) with the parameters satisfying the conditions (2.8) and (2.9).

Moreover, we assume also that the number of the weight vectors and the upper bound for the basis functions in (2.10) may depend on $n \ge 1$, i.e. $\check{\iota} = \check{\iota}(n)$ and $\phi_* = \phi_*(n)$, such that for any $\check{\epsilon} > 0$

$$\lim_{n\to\infty} \frac{\check{\iota}(n)}{n^{\check{\epsilon}}} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\phi_*(n)}{n^{\check{\epsilon}}} = 0\,. \tag{2.56}$$

The next result presents the non-asymptotic oracle inequality for the robust risk (2.5) for the model selection procedure (2.23), considered with the coefficients (2.28).

Theorem 2.3. *Assume that Conditions* $\mathbf{H}_1$) – $\mathbf{H}_4$) *hold and that the unknown function* S *is continuously differentiable. Then, for the robust risk defined in* (2.5) *through the distribution family* (2.8) – (2.9)*, the procedure* (2.23) *with the coefficients* (2.28) *for any* $n \ge 1$ *and* $0 < \delta < 1/6$*, satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{U}^*_n(S)}{n\delta}, \tag{2.57}$$

where the sequence $\mathbf{U}^*_n(S) > 0$ *is such that, under the conditions* (2.9), (2.26) *and* (2.56)*, for any* $r > 0$ *and* $\check{\delta} > 0$,

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\le r} \frac{\mathbf{U}^*_n(S)}{n^{\check{\delta}}} = 0. \tag{2.58}$$

Proof. First note, that in view of (2.31) and (2.26)

$$\lim_{n\to\infty} \frac{\check{\iota}}{n^{\check{\epsilon}}} = \lim_{n\to\infty} \frac{k^* m}{n^{\check{\epsilon}}} = 0 \qquad \text{for any} \quad \check{\epsilon} > 0\,.$$

Furthermore, the bound (2.32) and the conditions (2.9) and (2.26) yield

$$\lim_{n\to\infty} \frac{|\Lambda|_*}{n^{1/3+\check{\epsilon}}} = 0 \quad \text{for any} \quad \check{\epsilon} > 0\,.$$

So, from here we obtain the convergence (2.58). □

Now we need to check the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) for the process (2.2)

Proposition 2.2. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold. Then Condition* $\mathbf{C}_1$) *holds with*

$$\mathbf{C}_{1,Q,n} = \sigma_Q\, \check{\tau}\, \phi^2_{max}\, \|\Upsilon\|_1. \tag{2.59}$$

Proof. First, note that from (2.93) we have

$$\xi_{j,n} = \frac{\varrho_1}{\sqrt{n}} I_n^L(\phi_j) + \frac{\varrho_2}{\sqrt{n}} I_n^z(\phi_j)\,.$$

So, using (2.95) we can write that

$$\mathbf{E}\xi_{j,n}^2 = \frac{\varrho_1^2}{n}\int_0^n \phi_j^2(t)\mathrm{d}\,t + \frac{\varrho_2^2}{n}\mathbf{E}\sum_{l=1}^{\infty}\phi_j^2(T_l)1_{\{T_l\le n\}}\,. \tag{2.60}$$

Proposition 5.1 implies

$$\begin{aligned}\mathbf{E}\sum_{l=1}^{\infty}\phi_j^2(T_l)1_{\{T_l\le n\}} &= \int_0^n \phi_j^2(x)\,\rho(x)\mathrm{d}x \\ &= \frac{1}{\breve{\tau}}\int_0^n \phi_j^2(x)\mathrm{d}\,x + \int_0^n \phi_j^2(x)\Upsilon(x)\mathrm{d}\,x\,.\end{aligned}$$

Note that $\int_0^n \phi_j^2(t)\mathrm{d}\,t = n$. So, in view of the condition (2.10), we obtain

$$\left|\mathbf{E}\xi_{j,n}^2 - \sigma_Q\right| = \frac{\varrho_2^2}{n}\left|\int_0^n \phi_j^2(x)\Upsilon(x)\mathrm{d}\,x\right| \le \frac{\varrho_2^2}{n}\,\phi_{max}^2\,\|\Upsilon\|_1\,. \tag{2.61}$$

Estimating here ϱ_2^2 by $\sigma_Q\breve{\tau}$ we obtain the inequality (2.59) and hence the conclusion follows. □

Proposition 2.3. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold. Then Condition* $\mathbf{C}_2$) *holds with*

$$\mathbf{C}_{2,Q,n} = \phi_{max}^4(1+\sigma_Q^2)^3\,\breve{l} \tag{2.62}$$

and $\breve{l}$ *is given in* (2.35).

Proof. By Ito's formula one gets

$$\mathrm{d}I_t^2(f) = 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\breve{\varrho}^2\,f^2(t)\mathrm{d}\,t + \sum_{0\le s\le t} f^2(s)(\Delta\xi_s^d)^2\,, \tag{2.63}$$

where $\xi_t^d = \varrho_3\,\breve{L}_t + \varrho_2 z_t$ and $\varrho_3 = \varrho_1\sqrt{1-\breve{\varrho}^2}$. Taking into account that the processes $(\breve{L}_t)_{t\ge0}$ and $(z_t)_{t\ge0}$ are independent and the time of jumps T_k defined in (1.7) has a density, we have $\Delta z_s\Delta\breve{L}_s = 0$ a.s. for any $s\ge0$. Therefore, we can rewrite the differential (2.63) as

$$\begin{aligned}\mathrm{d}I_t^2(f) =& 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\breve{\varrho}^2\,f^2(t)\mathrm{d}\,t \\ &+ \varrho_3^2\mathrm{d}\sum_{0\le s\le t} f^2(s)(\Delta\breve{L}_s)^2 + \varrho_2^2\mathrm{d}\sum_{0\le s\le t} f^2(s)(\Delta z_s)^2\,.\end{aligned} \tag{2.64}$$

From Lemma 2.2 it follows that

$$\mathbf{E}I_t^2(f) = \varrho_1^2\int_0^t f^2(s)ds + \varrho_2^2\int_0^t f^2(s)\rho(s)\mathrm{d}s\,.$$

Therefore, putting

$$\widetilde{I}_t(f) = I_t^2(f) - \mathbf{E} I_t^2(f)\,, \tag{2.65}$$

we obtain

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

where $\check{m}_t = \sum_{0\le s\le t}(\Delta\check{L}_s)^2 - t$ and $m_t = \sum_{0\le s\le t}(\Delta z_s)^2 - \int_0^t \rho(s)\mathrm{d}s$. For any non-random vector $x = (x_j)_{1\le j\le n}$ with $\sum_{j=1}^n x_j^2 \le 1$, we set

$$\bar{I}_t(x) = \sum_{j=1}^n x_j \widetilde{I}_t(\phi_j). \tag{2.66}$$

Denoting

$$A_t(x) = \sum_{j=1}^n x_j I_t(\phi_j)\phi_j(t) \quad \text{and} \quad B_t(x) = \sum_{j=1}^n x_j \phi_j^2(t)\,, \tag{2.67}$$

we get the following stochastic differential equation for (2.66)

$$\mathrm{d}\bar{I}_t(x) = 2A_{t-}(x)\mathrm{d}\xi_t + B_t(x)\mathrm{d}\widetilde{m}_t\,, \quad \bar{I}_0(x) = 0\,.$$

Applying the Ito's formula one obtains

$$\begin{aligned}\mathbf{E}\,\bar{I}_n^2(x) =& 2\mathbf{E}\int_0^n \bar{I}_{t-}(x)\mathrm{d}\bar{I}_t(x) + 4\varrho_1^2\check{\varrho}^2\mathbf{E}\int_0^n A_t^2(x)\mathrm{d}\,t \\ &+ \varrho_3^2\ \mathbf{E}\,\check{D}_n(x) + \varrho_2^2\,\mathbf{E}\,D_n(x)\,,\end{aligned} \tag{2.68}$$

where $\check{D}_n(x) = \sum_{0\le t\le n}\left(2A_{t-}(x)\Delta\check{L}_t + \varrho_3^2 B_t(x)(\Delta\check{L}_t)^2\right)^2$ and
$D_n(x) = \sum_{k=1}^{+\infty}\left(2A_{T_k-}(x)Y_k + \varrho_2 B_{T_k-}(x)Y_k^2\right)^2\ \mathbf{1}_{\{T_k\le n\}}$. Let us now show that

$$\left|\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x)\right| \le 2\,\varrho_2^4\phi_{max}^3\,\|\Upsilon\|_1\,n^2\,. \tag{2.69}$$

To this end, note that

$$\begin{aligned}\int_0^n \bar{I}_{t-}(x)\mathrm{d}\bar{I}_t(x) =& 2\sum_{1\le j,l\le n} x_j x_l \int_0^n \widetilde{I}_{t-}(\phi_j)\, I_{t-}(\phi_l)\phi_l(t)\mathrm{d}\xi_t \\ &+ \sum_{j=1}^n x_j \int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}\widetilde{m}_t\,.\end{aligned}$$

Using here Lemma 2.5, we get $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\, I_{t-}(\phi_i)\phi_i(t)\mathrm{d}\xi_t = 0$. Moreover, the process $(\check{m}_t)_{t\ge 0}$ is a martingale, i.e. $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}\widetilde{m}_t = 0$. Therefore,

$$\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x) = \varrho_2^2\sum_{j=1}^n x_j\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)B_t(x)\mathrm{d}m_t\,.$$

Taking into account here that for any non-random bounded function f

$$\mathbf{E}\int_0^n f(t)\mathrm{d}m_t = 0,$$

we obtain $\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\,B_t(x)\,\mathrm{d}m_t = \mathbf{E}\int_0^n I^2_{t-}(\phi_j)\,B_t(x)\,\mathrm{d}m_t$. So, Lemma 2.4 yields

$$\left|\mathbf{E}\int_0^n \widetilde{I}_{t-}(\phi_j)\,B_t(x)\mathrm{d}m_t\right| = \left|\sum_{l=1}^n x_l\mathbf{E}\int_0^n I^2_{t-}(\phi_j)\phi^2_l(t)\mathrm{d}m_t\right|$$

$$\le\ 2\,\varrho_2^2\phi^3_{max}\,\|\Upsilon\|_1\sum_{l=1}^n |x_l|n\,.$$

Therefore,

$$\left|\mathbf{E}\int_0^n \bar{I}_{t-}(x)d\bar{I}_t(x)\right| \le 2\,\varrho_2^4\phi^3_{max}\,\|\Upsilon\|_1\,n\sum_{1\le l,j\le n}|x_l|\,|x_j|$$

$$= 2\,\varrho_2^4\phi^3_{max}\,\|\Upsilon\|_1\,n\left(\sum_{l=1}^n |x_l|\right)^2.$$

Taking into account here that $\left(\sum_{l=1}^n |x_l|\right)^2 \le\ n\sum_{l\ge 1} x_l^2 \le n$, we obtain (2.69). Reminding that $\Pi(x^2)=1$ we can calculate directly that

$$\mathbf{E}\,\check{D}_n(x) = 4\,\mathbf{E}\int_0^n A^2_t(x)\mathrm{d}t + \varrho_3^4\,\Pi(x^4)\int_0^n B^2_t(x)\mathrm{d}t\,. \tag{2.70}$$

Note that, thanks to Lemma 2.1, we obtain that

$$\mathbf{E}\int_0^n A^2_t(x)\mathrm{d}\,t = \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\mathbf{E}I_t\phi_i(t)I_t\phi_j(t)\mathrm{d}\,t$$

$$= \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\int_0^t \phi_i(v)\phi_j(v)(\varrho_1^2+\varrho_2^2\rho(v))\mathrm{d}v$$

$$= \frac{1}{2}\varrho_1^2\sum_{i,j} x_ix_j\left(\int_0^n \phi_i(t)\phi_j(t)\mathrm{d}t\right)^2 + \varrho_2^2\,A_{1,n}(x)$$

$$\le \frac{n^2}{2}\varrho_1^2 + \varrho_2^2\,A_{1,n}(x)\,,$$

where $A_{1,n}(x) = \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\,\left(\int_0^t \phi_i(v)\phi_j(v)\,\rho(v)\mathrm{d}v\right)\mathrm{d}t$. This term can be estimated through Proposition 5.1 as

$$\left|A_{1,n}(x)\right| = \left|\frac{n^2}{2\check{\tau}} + \sum_{i,j} x_ix_j\int_0^n \phi_i(t)\phi_j(t)\,\left(\int_0^t \phi_i(v)\phi_j(v)\,\Upsilon(v)\mathrm{d}v\right)\mathrm{d}t\right|$$

$$\le \frac{n^2}{2\check{\tau}} + n\,\phi^4_{max}\,\|\Upsilon\|_1\sum_{i,j}|x_i||x_j| \le \left(\frac{1}{2\check{\tau}} + \phi^4_{max}\,\|\Upsilon\|_1\right)n^2\,.$$

So, reminding that $\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau}$ and that $\phi_{max} \geq 1$, we obtain that

$$\mathbf{E}\int_0^n A_t^2(x)\mathrm{d}\,t \leq \left(\frac{\sigma_Q}{2} + \phi_{max}^4 \|\Upsilon\|_1\right) n^2 \leq \left(\frac{1}{4} + \|\Upsilon\|_1\right) \phi_{max}^4 \,(1+\sigma_Q^2)\, n^2\,. \tag{2.71}$$

Taking into account that

$$\sup_{t\geq 0} B_t^2(x) \;\leq\; \phi_{max}^4 \left(\sum_{j=1}^n |x_j|\right)^2 \;\leq\; \phi_{max}^4\, n\,, \tag{2.72}$$

that $\phi_{max} \geq 1$ and that $\varrho_1^4 \leq \sigma_Q^2$, we estimate the expectation in (2.70) as

$$\mathbf{E}\check{D}_n \leq 4\phi_{max}^4(1+\sigma_Q^2)\left(1 + \|\Upsilon\|_1 + \Pi(x^4)\right)\, n^2\,. \tag{2.73}$$

Moreover, taking into account that the random variable Y_k is independent of $A_{T_{k^-}}(x)$ and of the field $\mathcal{G} = \sigma\{T_j\,,\, j \geq 1\}$ and that $\mathbf{E}\left(A_{T_{k^-}}(x)\,|\mathcal{G}\right) = 0$, we get

$$\mathbf{E}\sum_{k=1}^{+\infty} B_{T_k-}(x)\, A_{T_{k^-}}(x) Y_k^3 1_{\{T_k\leq n\}} = \sum_{k=1}^{+\infty} \mathbf{E}\mathbf{E}\left(B_{T_k-}(x)\, A_{T_{k^-}}(x) Y_k^3 1_{\{T_k\leq n\}}|\mathcal{G}\right)$$
$$= \mathbf{E}\, Y_1^3\, \mathbf{E}\sum_{k=1}^{+\infty} B_{T_k-}(x) 1_{\{T_k\leq n\}}\, \mathbf{E}(A_{T_{k^-}}(x)|\mathcal{G}) = 0\,.$$

Therefore,

$$\mathbf{E}\, D_n(x) = \varrho_2^2 \mathbf{E} Y_1^4 D_{1,n}(x) + 4D_{2,n}(x), \tag{2.74}$$

where

$$D_{1,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E}\, B_{T_k-}^2(x) 1_{\{T_k\leq n\}} \quad \text{and} \quad D_{2,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E}\, A_{T_{k^-}}^2(x) \mathbf{1}_{\{T_k\leq n\}}\,.$$

Using the bound (2.72) we can estimate the term $D_{1,n}$ as $D_{1,n}(x) \leq \phi_{max}^4 n \mathbf{E}\, N_n$. Using here Corollary 5.1, we obtain

$$D_{1,n}(x) \leq\; |\rho|_* \phi_{max}^4\, n^2\,. \tag{2.75}$$

Now, to estimate the last term in (2.74), note that the process $A_t(x)$ can be rewritten as

$$A_t(x) = \int_0^t Q_x(t,s) d\xi_s, \quad \text{with} \quad Q_x(t,s) = \sum_{j=1}^n x_j \phi_j(s)\phi_j(t). \tag{2.76}$$

Applying Lemma 2.3 again, we obtain for any $k \geq 1$

$$\mathbf{E}\left(A_{T_{k^-}}^2(x)|\mathcal{G}\right) = \varrho_1^2 \int_0^{T_k} Q_x^2(T_k,s)\mathrm{d}s + \varrho_2^2 \sum_{j=1}^{k-1} Q_x^2(T_k,T_j)\,.$$

So, we can represent the last term in (2.74) as

$$D_{2,n} = \varrho_1^2\, D_{2,n}^{(1)} + \varrho_2^2\, D_{2,n}^{(2)}\,, \tag{2.77}$$

where

$$D_{2,n}^{(1)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \int_0^{T_k} Q_x^2(T_k,s)\mathrm{d}s$$

and

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \sum_{j=1}^{k-1} Q_x^2(T_k,T_j)\,.$$

Thanks to Proposition 5.1 we obtain

$$D_{2,n}^{(1)} = \int_0^n \left(\int_0^t Q_x^2(t,s)\mathrm{d}s\right) \rho(t)\,\mathrm{d}t \;\le\; |\rho|_* \int_0^n \int_0^n Q_x^2(t,s)\mathrm{d}s\,\mathrm{d}t\,.$$

In view of the definition of Q_x in (2.76), we can rewrite the last integral as

$$\begin{aligned}\int_0^n Q_x^2(t,s)\mathrm{d}s &= \sum_{1\le i,j\le n} x_i\, x_j\, \phi_i(t)\,\phi_j(t) \int_0^n \phi_i(s)\,\phi_j(s)\,\mathrm{d}s \\ &= \sum_{i=1}^n x_i^2\,\phi_i^2(t) \int_0^n \phi_i^2(s)\,\mathrm{d}s = n\sum_{i=1}^n x_i^2\,\phi_i^2(t)\,.\end{aligned}$$

Since $\sum_{j=1}^n x_j^2 \le 1$, we obtain that,

$$\int_0^n Q_x^2(t,s)\mathrm{d}s \;\le\; \phi_{max}^2\, n \quad\text{and}\quad D_{2,n}^{(1)} \le \phi_{max}^2\,|\rho|_*\, n^2\,. \tag{2.78}$$

Let us estimate now the last term in (2.77). First, note that we can represent this term as

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}} \sum_{j=1}^{k-1} Q_x^2(T_k,T_j) = \sum_{j=1}^{\infty} \mathbf{1}_{\{T_j\le n\}}\, G(T_j) = \int_0^n G(t)\,\rho(t)\mathrm{d}t\,,$$

where

$$\begin{aligned}G(t) &= \sum_{k=1}^{+\infty} \mathbf{E}\,\mathbf{1}_{\{T_k\le n\}}\, Q_x^2((t+T_k),t) = \int_0^n Q_x^2(t+v,t)\,\rho(v)\mathrm{d}v \\ &= \int_t^{n+t} Q_x^2(u,t)\,\rho(u-t)\mathrm{d}u\,.\end{aligned}$$

It is clear that, for any $0\le t\le n$,

$$\int_t^{n+t} Q_x^2(u,u-t)\,\rho(u)\,\mathrm{d}u \;\le |\rho|_* \int_0^{2n} Q_x^2(v,t)\,\mathrm{d}v\,.$$

In view of the inequality (2.78) we obtain

$$\int_0^{2n} Q_x^2(u,t)\,\mathrm{d}u = \int_0^{2n} Q_x^2(t,u)\,\mathrm{d}u \,\le 2\phi_{max}^2\, n\,.$$

Therefore,

$$\max_{0\le t\le n} G(t) \,\le\, 2|\rho|_*\,\phi_{max}^2\, n \quad \text{and} \quad D_{2,n}^{(2)} \le\, 2|\rho|_*^2\,\phi_{max}^2\, n^2\,.$$

So, estimating ϱ_2^2 by $\check{\tau}\sigma_Q$ and taking into account that $\mathbf{E}Y_1^4 \ge 1$, we obtain that we obtain that

$$\mathbf{E}\, D_n(x) \le 13\,(1+\check{\tau})\phi_{max}^4\ \mathbf{E}Y_1^4(1+|\rho|_*^2)\, n^2\sigma_Q\,.$$

Using all these bounds in (2.68) we obtain (2.62) and thus the conclusion follows. □

Remark 2.5. The properties (2.59) and (2.62) are used to obtain the oracle inequalities given in Section 2.3 (see, for example, [17]).

2.4 Efficiency

Now we study the asymptotic efficiency for the procedure (2.23) with the coefficients (2.28), with respect to the robust risk (2.5) defined by the distribution family (2.8)–(2.9). To this end, we assume that the unknown function S in the model (2.1) belongs to the Sobolev ball

$$W_{\mathbf{r}}^k = \{f \in \mathcal{C}_{per}^k[0,1] \,:\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le \mathbf{r}\}\,, \tag{2.79}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some unknown parameters, $\mathcal{C}_{per}^k[0,1]$ 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 $W_{\mathbf{r}}^k$ can be written as an ellipsoid in $\mathbf{L}_2[0,1]$, i.e.,

$$W_{\mathbf{r}}^k = \{f \in \mathcal{C}_{per}^k[0,1] \,:\, \sum_{j=1}^{\infty} a_j\,\theta_j^2 \,\le \mathbf{r}\}, \tag{2.80}$$

where $a_j = \sum_{i=0}^{k} (2\pi[j/2])^{2i}$ and $\theta_j = \int_0^1 f(v)\mathrm{Tr}_j(v)\mathrm{d}v$. We recall that the trigonometric basis $(\mathrm{Tr}_j)_{j\ge1}$ is defined in (1.15).

Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (2.5) is given by

$$\mathbf{r}_k^* = \,((2k+1)\mathbf{r})^{1/(2k+1)}\,\left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)}\,. \tag{2.81}$$

Note that this is the well-known Pinsker constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]). Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the σ-field $\sigma\{y_t\,,\, 0 \le t \le n\}$ generated by the process (2.1).

The following two results give the lower and upper bound for the robust risk in our case.

Theorem 2.4. *Under Conditions* (2.8) *and* (2.9),

$$\liminf_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) \geq \mathbf{r}_k^*\,, \tag{2.82}$$

where $\upsilon_n = n/\varsigma^*$.

Proof. First, we denote by Q_0 the distribution of the noise (2.2) with the parameter $\varrho_1 = \varsigma^*$, $\breve{\varrho} = 1$ and $\varrho_2 = 0$, i.e. the distribution for the "signal + white noise" model. So, we can estimate as below the robust risk

$$\mathcal{R}_n^*(\widetilde{S}_n, S) \geq \mathcal{R}_{Q_0}(\widetilde{S}_n, S)\,.$$

Now Theorem 6.1 from [15] yields the lower bound (2.82). Hence this finishes the proof. □

Note that if the parameters $\mathbf{r}$ and k are known, i.e. for the non-adaptive estimation case, then to obtain the efficient estimation for the "signal+white noise" model. Pinsker in [47] proposed to use the estimate $\widehat{S}_{\lambda_0}$ defined in (2.15) with the weights (2.28) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, l_0)\,, \tag{2.83}$$

where $l_0 = [\mathbf{r}/\varepsilon]\varepsilon$. For the model (2.1) – (2.2) we show the same result.

Proposition 2.4. *The estimator* $\widehat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^*\,.$$

Proof. Putting $\lambda_0(j) = 0$ for $j \geq n$ we can represent the quadratic risk for the estimator (2.15) as

$$\| \widehat{S}_{\lambda_0} - S \|^2 = \sum_{j=1}^{\infty} (1-\lambda_0(j))^2\theta_j^2 - 2H_n + \frac{1}{n}\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2\,,$$

where $H_n = n^{-1/2}\sum_{j=1}^{n}(1-\lambda_0(j))\lambda_0(j)\theta_j\xi_{j,n}$. Note that $\mathbf{E}_Q H_n = 0$ for any $Q \in Q_n$, therefore,

$$\mathbf{E}_Q \| \widehat{S}_{\lambda_0} - S \|^2 = \sum_{j=1}^{\infty} (1-\lambda_0(j))^2\theta_j^2 + \frac{1}{n}\mathbf{E}_Q\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2\,.$$

Proposition 2.2 and the last inequality in (2.8) imply that for any $Q \in \mathcal{Q}_n$

$$\mathbf{E}_Q\sum_{j=1}^{n}\lambda_0^2(j)\xi_{j,n}^2 \leq \varsigma^*\sum_{j=1}^{n}\lambda_0^2(j) + \frac{\phi_{max}^2\varsigma^*\|\Upsilon\|_1}{\check{\tau}} := \varsigma^*\sum_{j=1}^{n}\lambda_0^2(j) + \mathbf{C}_{1,n}^*\,.$$

Therefore,

$$\mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \sum_{j=j_*}^{\infty}(1-\lambda_0(j))^2\theta_j^2 + \frac{1}{\upsilon_n}\sum_{j=1}^{n}\lambda_0^2(j) + \frac{\mathbf{C}_{1,n}^*}{n},$$

where j_* and υ_n are defined in (2.28). Setting

$$\Upsilon_{1,n}(S) = \upsilon_n^{2k/(2k+1)} \sum_{j=j_*}^{\infty} (1-\lambda_0(j))^2 \theta_j^2 \quad \text{and} \quad \Upsilon_{2,n} = \frac{1}{\upsilon_n^{1/(2k+1)}} \sum_{j=1}^{n} \lambda_0^2(j)\,,$$

we rewrite the last inequality as

$$\upsilon_n^{2k/(2k+1)}\, R_n^*(\widehat{S}_{\lambda_0}, S) \le \Upsilon_{1,n}(S) + \Upsilon_{2,n} + \check{\mathbf{C}}_n\,, \tag{2.84}$$

where $\check{\mathbf{C}}_n = \upsilon_n^{2k/(2k+1)} \mathbf{C}_{1,n}^*/n$. Note that Conditions (2.9) and (2.56) imply that $\mathbf{C}_{1,n}^* = (n^{\check{\delta}})$ as $n \to \infty$ for any $\check{\delta} > 0$; therefore, $\check{\mathbf{C}}_n \to 0$ as $n \to \infty$. Putting

$$u_n = \upsilon_n^{2k/(2k+1)} \sup_{j \ge j_*} (1-\lambda_0(j))^2/a_j\,,$$

with a_j defined in (2.80), we estimate the first term in (2.84) as

$$\sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \le \sup_{S \in W_{\mathbf{r}}^k} u_n \sum_{j \ge 1} a_j \theta_j \le u_n \mathbf{r}\,.$$

Taking into account that $a_j/(\pi^{2k} j^{2k}) \to 1$ as $j \to \infty$ and $l_0 \to \mathbf{r}$ as $\varepsilon \to 0$ and using the definition of ω_{α_0} in (2.28), we obtain that

$$\begin{aligned}\limsup_{n\to\infty} u_n &\le \lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{j \ge j_*} \frac{(1-\lambda_0(j))^2}{(\pi\, j)^{2k}} \\ &= \lim_{n\to\infty} \frac{\upsilon_n^{2k/(2k+1)}}{\pi^{2k} \omega_{\alpha_0}^{2k}} = \frac{1}{\pi^{2k}\,(\mathrm{d}_k \mathbf{r})^{2k/(2k+1)}}\,.\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty} \sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \le \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k} (\mathrm{d}_k)^{2k/(2k+1)}} =: \Upsilon_1^*\,. \tag{2.85}$$

As to the second term in (2.84), note that

$$\lim_{n\to\infty} \frac{1}{\omega_{\alpha_0}} \sum_{j=1}^{n} \lambda_0^2(j) = \int_0^1 (1-t^k)^2 \mathrm{d}t = \frac{2k^2}{(k+1)(2k+1)}\,.$$

So, taking into account that $\omega_{\alpha_0}/\upsilon_n^{1/(2k+1)} \to (\mathrm{d}_k \mathbf{r})^{1/(2k+1)}$ as $n \to \infty$, the limit of $\Upsilon_{2,n}$ can be calculated as

$$\lim_{n\to\infty} \Upsilon_{2,n} = \frac{2(\mathrm{d}_k \mathbf{r})^{1/(2k+1)}\, k^2}{(k+1)(2k+1)} =: \Upsilon_2^*\,.$$

Moreover, since $\Upsilon_1^* + \Upsilon_2^* =: \mathbf{r}_k^*$, we obtain

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \le \mathbf{r}_k^*$$

and get the desired result. □

For the adaptive estimation we use the model selection procedure (2.23) with the parameter δ defined as a function of n satisfying

$$\lim_n \delta_n = 0 \quad \text{and} \quad \lim_n n^{\check{\delta}}\, \delta_n = 0 \tag{2.86}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$.

Theorem 2.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then the robust risk defined in* (2.5) *through the distribution family* (2.8)–(2.9) *for the procedure* (2.23) *based on the trigonometric basis* (1.15) *with the coefficients* (2.28) *and the parameter* $\delta = \delta_n$ *satisfying* (2.86) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le \mathbf{r}_k^* \,. \tag{2.87}$$

Theorem 2.4 and Theorem 2.5 allow us to compute the optimal convergence rate.

Corollary 2.1. *Under the assumptions of Theorem 2.5, we have*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^* \,. \tag{2.88}$$

Remark 2.6. 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, [47], [46]). We see here that the efficient robust rate is $\upsilon_n^{2k/(2k+1)}$, i.e., if the distribution upper bound $\varsigma^* \to 0$ as $n \to \infty$, we obtain a faster rate with respect to $n^{2k/(2k+1)}$, and, if $\varsigma^* \to \infty$ as $n \to \infty$, we obtain a slower rate. In the case when ς^* is constant, than the robust rate is the same as the classical non robust convergence rate.

2.5 Simulations

In this section we report the results of a Monte Carlo experiment in order to assess the performance of the proposed model selection procedure (2.23). In (2.1) we chose a 1-periodic function which is defined, for $0 \le t \le 1$, as

$$S(t) = t\sin(2\pi t) + t^2(1-t)\cos(4\pi t)\,. \tag{2.89}$$

We simulate the model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,,$$

where $\xi_t = 0.5\mathrm{d}wt + 0.5\mathrm{d}z_t$.

Here z_t is the semi-Markov process defined in (1.6) with a Gaussian $\mathcal{N}(0,1)$ sequence $(Y_j)_{j\ge1}$ and $(\tau_k)_{k\ge1}$ used in (1.7) taken as $\tau_k \sim \chi_3^2$.

n	$\overline{\mathbf{R}}$	$\overline{\mathbf{R}_*}$
20	0.04430	0.235
100	0.01290	0.068
200	0.00812	0.043
1000	0.00196	0.010

Table 2.1: Empirical risks

We use the model selection procedure (2.23) with the weights (2.28) in which $k^* = 100 + \sqrt{\ln(n)}$, $t_i = i/\ln(n)$, $m = [\ln^2(n)]$ and $\delta = (3 + \ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathbf{R}} = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\left(\widehat{S}_n(t_j) - S(t_j)\right)^2, \tag{2.90}$$

where the observation frequency $p = 100001$ and the expectation was taken as an average over $N = 10000$ replications, i.e.,

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

We set the relative quadratic risk as

$$\overline{\mathbf{R}_*} = \overline{\mathbf{R}}/||S||_p^2, \quad \text{with} \quad ||S||_p^2 = \frac{1}{p}\sum_{j=0}^{p} S^2(t_j)\,. \tag{2.91}$$

In our case $||S||_p^2 = 0.1883601$.

Table 2.1 gives the values for the sample risks (2.90) and (2.91) for different numbers of observations n.

Figures 2.1–2.4 show the behaviour of the regression function and its estimates by the model selection procedure (2.23) depending on the values of observation periods n. The black full line is the regression function (2.89) and the red dotted line is the associated estimator.

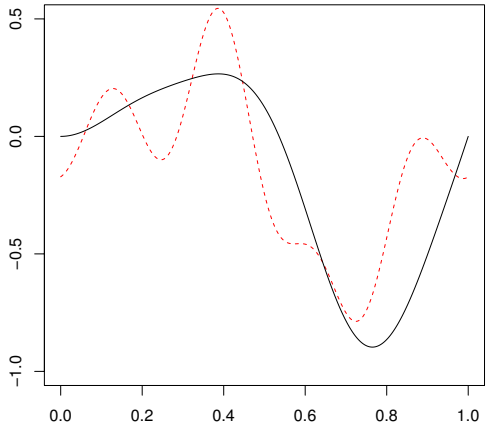

Figure 2.1: Estimator of S for $n = 20$

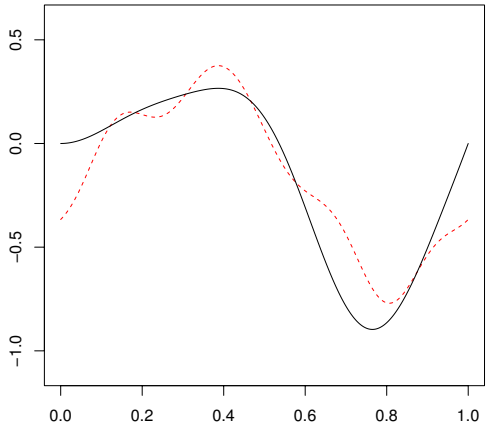

Figure 2.2: Estimator of S for $n = 100$

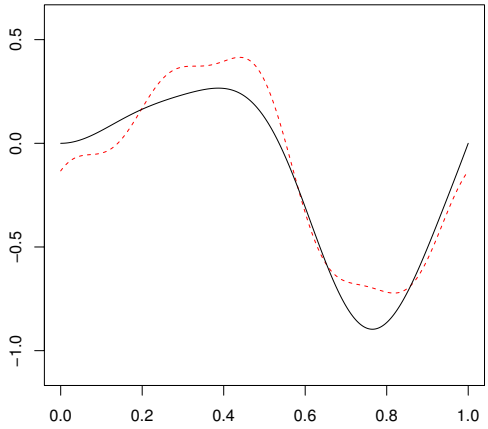

Figure 2.3: Estimator of S for $n = 200$

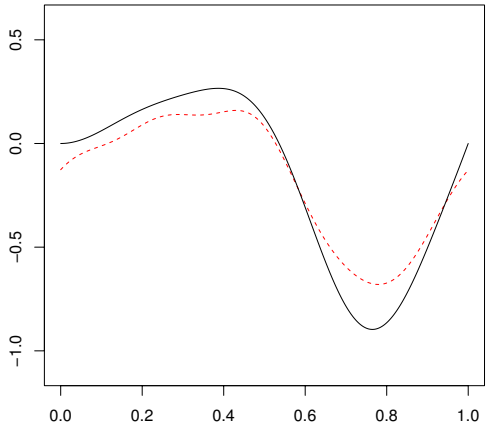

Figure 2.4: Estimator of S for $n = 1000$

Remark 2.7. From numerical simulations of the procedure (2.23) with various observation numbers n we may conclude that the quality of the proposed procedure: (i) is good for practical

needs, i.e. for reasonable (non large) number of observations; (ii) is improving as the number of observations increases.

Now we give the algorithm of the model selection procedure given in Section 2.2

Algorithm 1 Model selection procedure

Require: n, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

ρ_1, ρ_2, ς^*: satisfying Conditions (2.8) and (2.9)

$k^* \geq 1$, ε: satisfying Condition (2.26)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$m = [1/\varepsilon^2]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

    for $k \longleftarrow 1$ *to* n **do**

      Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (2.28)

    end for

  end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varepsilon, \ldots, m\varepsilon\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,n} = \frac{1}{n} \int_0^n \phi_k(t) \mathrm{d}\, y_t\,.$

$\widetilde{\theta}_{k,n} \longleftarrow \widehat{\theta}^2_{k,n} - \frac{1}{n}\,.$

The observation $(y_t)_{0 \leq t \leq n}$ are given in (2.1) with the noise process (2.2) and $(\phi_k)_{k \geq 1}$ is the basis given in (2.10)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,n}, ..., \widehat{\theta}_{n,n})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,n}, ..., \widetilde{\theta}_{n,n})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

    $J_n(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^n \lambda_{i,j}(l)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda)$.

    where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_n is the penalty term given in (2.22)

  end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_n(\lambda), \Lambda = \{\lambda_\alpha\,,\, \alpha \in \mathcal{A}\}$.

2.6 Stochastic calculus for semi-Markov processes

In this section we give some results of stochastic calculus for the process $(\xi_t)_{t\geq 0}$ given in (2.2), needed all along this paper. As the process ξ_t is the combination of a Lévy process and a semi-Markov process, these results are not standard and need to be provided.

Lemma 2.1. *Let f and g be any non-random functions from $\mathbf{L}_2[0,n]$ and $(I_t(f))_{t\geq 0}$ be the process defined in* (2.6)*. Then, for any $0 \leq t \leq n$,*

$$\mathbf{E}\, I_t(f) I_t(g) = \varrho_1^2\,(f,g)_t \,+ \varrho_2^2\,(f, g\rho)_t\,, \tag{2.92}$$

where $(f,g)_t = \int_0^t f(s)\, g(s)\mathrm{d}s$ and ρ is the density defined in (1.8)*.*

Proof. First, note that we can represent the stochastic integral $I_t(f)$ as

$$I_t(f) = \varrho_1 I_t^L(f) + \varrho_2 I_t^z(f)\,, \tag{2.93}$$

where

$$I_t^L(f) = \int_0^t f(s)\mathrm{d}L_s \quad \text{and} \quad I_t^z(f) = \int_0^t f(s)\mathrm{d}z_s\,.$$

Note that the mutual covariation for the martingales $I_t^L(f)$ and $I_t^L(g)$ (see, for example, [21]) may be calculated as

$$[I^L(f), I^L(g)]_t = \check{\varrho}^2 \int_0^t f(s)g(s)\mathrm{d}s + (1-\check{\varrho}^2) \sum_{0\leq s\leq t} f(s)g(s)\left(\Delta \check{L}_s\right)^2\,, \tag{2.94}$$

where $\Delta\check{L}_s = \check{L}_s - \check{L}_{s-}$. Taking into account that $\mathbf{E}\, I_t^L(f)\, I_t^L(g) = \mathbf{E}\,[I^L(f), I^L(g)]_t$ and that in view of the first condition in (2.3) $\Pi(x^2) = 1$, we obtain that

$$\begin{aligned}\mathbf{E}\, I_t^L(f)\, I_t^L(g) &= \check{\varrho}^2 \int_0^t f(s)g(s)\mathrm{d}s + (1-\check{\varrho}^2)\,\Pi(x^2) \int_0^t f(s)\, g(s)\mathrm{d}s \\ &= \int_0^t f(s)\, g(s)\mathrm{d}s\,.\end{aligned} \tag{2.95}$$

Moreover, note that

$$\begin{aligned}\mathbf{E} I_t^z(f) I_t^z(g) &= \mathbf{E}\left(\sum_{l=1}^{\infty} f(T_l)g(T_l)Y_l^2 1_{\{T_l\leq t\}}\right) \\ &= \mathbf{E}\left(\sum_{l=1}^{\infty} f(T_l)g(T_l)1_{\{T_l\leq t\}}\right) = \int_0^t f(s)g(s)\rho(s)\mathrm{d}s\,.\end{aligned}$$

Hence the conclusion follows. □

Lemma 2.2. *Assume that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then, for any $n \geq 1$ and for any non random function f from $\mathbf{L}_2[0,n]$, the stochastic integral* (2.6) *exists and satisfies the properties* (2.7) *with the coefficient $\varkappa_Q$ given in* (2.7)*.*

Proof. This lemma follows directly from Lemma 2.1 with $f = g$ and Proposition 5.1. □

Lemma 2.3. *Let f and g be bounded functions defined on $[0,\infty)\times\mathbb{R}$. Then, for any $k \geq 1$,*

$$\mathbf{E}\left(I_{T_{k-}}(f)\, I_{T_{k-}}(g) \mid \mathcal{G}\right) = \varrho_1^2(f\,,\, g)_{T_k} + \varrho_2^2 \sum_{l=1}^{k-1} f(T_l)\, g(T_l),$$

where $\mathcal{G}$ is the σ-field generated by the sequence $(T_l)_{l\geq 1}$, i.e., $\mathcal{G} = \sigma\{T_l\,,\, l \geq 1\}$.

Proof. Using (2.93), (2.95) and, taking into account that the process $(L_t)_{t\geq 0}$ is independent of $\mathcal{G}$, we obtain

$$\mathbf{E}\left(I_{T_{k-}}(f)\, I_{T_{k-}}(g) \mid \mathcal{G}\right) = \varrho_1^2(f\,,\, g)_{T_k} + \mathbf{E}\left(I^z_{T_{k-}}(f)\, I^z_{T_{k-}}(g) \mid \mathcal{G}\right).$$

Moreover,

$$\begin{aligned}\mathbf{E}\left(I^z_{T_{k-}}(f)\, I^z_{T_{k-}}(g) \mid \mathcal{G}\right) &= \mathbf{E}\left(\left(\sum_{l=1}^{k-1} f(T_l)Y_l\right)\left(\sum_{l=1}^{k-1} g(T_l)Y_l\right) \mid \mathcal{G}\right)\\ &= \sum_{l=1}^{k-1} f(T_l)\, g(T_l)\,.\end{aligned}$$

Thus we obtain the desired result. □

Lemma 2.4. *Assume that Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$) hold true. Then, for any measurable bounded non-random functions f and g, we have*

$$\left|\mathbf{E}\int_0^n I^2_{t-}(f)\, g(t)\,\mathrm{d}m_t\right| \leq 2\varrho_2^2 |g|_*|f|_*^2\, \|\Upsilon\|_1\, n.$$

Proof. Using the definition of the process $(m_t)_{t\geq 0}$ we can represent this integral as

$$\begin{aligned}\int_0^n I^2_{t-}(f)\, g(t)\,\mathrm{d}m_t &= \sum_{k\geq 1} I^2_{T_k-}(f)\, g(T_k)\, Y_k^2\, \mathbf{1}_{\{T_k\leq n\}}\\ &\quad - \int_0^n I_t^2(f)\, g(t)\, \rho(t)\,\mathrm{d}t =: V_n - U_n\,. \qquad (2.96)\end{aligned}$$

Note now that

$$\mathbf{E}\, V_n \;= \mathbf{E}\sum_{k\geq 1} g(T_k)\, \mathbf{E}\left(I^2_{T_k-}(f)\;\mid \mathcal{G}\right)\, \mathbf{1}_{\{T_k\leq n\}}\,.$$

Now, using Lemma 2.3 we can represent the last expectation as

$$\mathbf{E}\, V_n \;= \varrho_1^2\, \mathbf{E}\, V_n' + \varrho_2^2\, \mathbf{E}\, V_n'', \qquad (2.97)$$

where

$$V_n^{'} = \sum_{k\geq 1} g(T_k)\, \|f\|^2_{T_k}\, \mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad V_n^{''} = \sum_{k\geq 2} g(T_k)\, \mathbf{1}_{\{T_k\leq n\}} \sum_{l=1}^{k-1} f^2(T_l)\,.$$

The first term in (2.97) can be represented as

$$\mathbf{E}\, V_n^{'} = \int_0^n g(t)\, \|f\|_t^2 \rho(t)\mathrm{d}t\,.$$

To estimate the last expectation in (2.97), note that

$$\mathbf{E}\, V_n^{''} = \mathbf{E} \sum_{l\geq 1} f^2(T_l)\, \bar{g}(T_l) \mathbf{1}_{\{T_l\leq n\}} = \int_0^n f^2(v)\, \bar{g}(v)\, \rho(v)\mathrm{d}v\,,$$

where

$$\bar{g}(v) = \mathbf{E} \sum_{k\geq 1} g(v+T_k)\, \mathbf{1}_{\{T_k\leq n-v\}} = \int_v^n g(t)\, \rho(t-v)\mathrm{d}t\,.$$

Moreover, using now the representation (2.92), we calculate the expectation of the last term in (2.96)

$$\mathbf{E}\, U_n = \varrho_1^2 \int_0^n \|f\|_t^2\, g(t)\, \rho(t)\, \mathrm{d}t + \varrho_2^2 \int_0^n \check{f}(t)\, g(t)\, \rho(t)\, \mathrm{d}t\,,$$

where $\check{f}(t) = \int_0^t f^2(s)\, \rho(s)\, \mathrm{d}s$. This implies that

$$\mathbf{E} \int_0^n I_{t-}^2(f)\, g(t)\, \mathrm{d}m_t = \varrho_2^2 \int_0^n g(t)\, \delta(t)\mathrm{d}t\,,$$

where $\delta(t) = \int_0^t f^2(v)\, (\rho(t-v) - \rho(t))\, \rho(v)\, \mathrm{d}v$. Note that, in view of Proposition 5.1, the function δ can be estimated as

$$|\delta(t)| \leq |f|_*^2\, |\rho|_* \int_0^t |\Upsilon(t-v) - \Upsilon(t)|\, \mathrm{d}v \leq |f|_*^2\, |\rho|_*\, (\|\Upsilon\|_1 + t|\Upsilon(t)|)\,.$$

Therefore,

$$\left| \mathbf{E} \int_0^n I_{t-}^2(f)\, g(t)\, \mathrm{d}m_t \right| \leq 2\varrho_2^2 |g|_* |f|_*^2\, \|\Upsilon\|_1\, n$$

and this finishes the proof. □

Lemma 2.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any measurable bounded non-random functions* f *and* g*, one has*

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = 0.$$

Proof. First, note that

$$\int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = \varrho_1 \int_0^n I_t^2(f) I_t(g) g(t) \mathrm{d}L_t + \varrho_2 \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) \mathrm{d}z_t.$$

Second, we will show that

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) \mathrm{d}L_t = 0\,. \tag{2.98}$$

Using the notations (2.93), we set

$$J_1 = \int_0^n I_t^2(f) I_t^L(g) g(t) \mathrm{d}L_t \quad \text{and} \quad J_2 = \int_0^n I_t^2(f) I_t^z(g) g(t) \mathrm{d}L_t,$$

we obtain that

$$\int_0^n I_t^2(f) I_t(g) g(t) \mathrm{d}L_t = \varrho_1 J_1 + \varrho_2 J_2 \,. \tag{2.99}$$

Now let us recall the Novikov inequalities, [44], also referred to as the Bichteler–Jacod inequalities (see [32, 43]) providing bound moments of supremum of purely discontinuous local martingales for any predictable function h and any $p \geq 2$

$$\mathbf{E} \sup_{0 \leq t \leq n} \left| \int_{[0,t] \times \mathbb{R}} h \, \mathrm{d}(\mu - \nu) \right|^p \leq C_p^* \mathbf{E} \, \check{J}_{p,n}(h) \,, \tag{2.100}$$

where C_p^* is some positive constant and

$$\check{J}_{p,n}(h) = \left(\int_{[0,n] \times \mathbb{R}} h^2 \, \mathrm{d}\nu \right)^{p/2} + \int_{[0,n] \times \mathbb{R}} h^p \, \mathrm{d}\nu \,.$$

By applying this inequality for the non-random function $h(s,x) = g(s)x$, and, recalling that $\Pi(x^8) < \infty$, we obtain,

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^{\check{L}}(g) \right|^8 < \infty \,.$$

Taking into account that, for any non random square integrated function f, the integral $\left(\int_0^t f(s) \mathrm{d}w_s \right)$ is Gaussian with the parameters $\left(0, \int_0^t f^2(s) \mathrm{d}s \right)$, we obtain

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^L(g) \right|^8 < \infty.$$

Finally, by using the Cauchy's inequality, we can estimate for any $0 < t \leq n$ the following expectation as

$$\mathbf{E} \, (I_t^L(f))^4 (I_t^L(g))^2 \leq \sqrt{\mathbf{E} \, (I_t^L(f))^8} \sqrt{\mathbf{E} \, (I_t^L(f))^4}$$

i.e.,

$$\sup_{0 \leq t \leq n} \mathbf{E} \, (I_t^L(f))^4 (I_t^L(g))^2 < \infty \,.$$

Moreover, taking into account that the processes $(L_t)_{t \geq 0}$ and $(z_t)_{t \geq 0}$ are independent, we obtain that

$$\mathbf{E} \, (I_t^z(f))^4 (I_t^L(g))^2 = \mathbf{E} \, (I_t^z(f))^4 \mathbf{E} \, (I_t^L(g))^2 = \int_0^t g^2(s) \mathrm{d}s \, \mathbf{E} \, (I_t^z(f))^4 \,.$$

One can check directly here that, for $t > 0$,

$$\mathbf{E} \, |I_t^z(f)|^4 \leq |f|_*^4 \, \mathbf{E} \, Y_1^4 \, \mathbf{E} N_t^2 \,.$$

Note that the last bound in Corollary 5.1 yields $\sup_{0\le t\le n}\mathbf{E}\,(I_t^z(f))^4<\infty$ and, therefore,

$$\sup_{0\le t\le n}\mathbf{E}\,(I_t(f))^4(I_t^L(g))^2<\infty\,.$$

It follows directly that $\mathbf{E}J_1=0$. Now we study the last term in (2.99). To this end, first note that similarly to the previous reasoning we obtain that

$$\mathbf{E}\int_0^n(I_t^L(f))^2I_t^z(g)g(t)\mathrm{d}L_t=0\quad\text{and}\quad\mathbf{E}\int_0^n I_t^L(f)I_t^z(f)I_t^z(g)g(t)\mathrm{d}L_t=0\,.$$

Therefore, to show (2.98) one needs to show that

$$\mathbf{E}\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t=0\,. \tag{2.101}$$

To check this, note that, for any $0<t\le n$ and for any bounded function f,

$$I_t^z(f)=\sum_{k=1}^{\infty}f(T_k)\,Y_k\,\mathbf{1}_{\{T_k\le t\}}=\sum_{k=1}^{N_n}f(T_k)\,Y_k\,\mathbf{1}_{\{T_k\le t\}}\,,$$

i.e.,

$$\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t=\sum_{k=1}^{N_n}\sum_{l=1}^{N_n}\sum_{j=1}^{N_n}f(T_k)\,f(T_l)\,g(T_j)\,Y_jY_l\,Y_k\,I_{klj}\,,$$

where

$$I_{klj}=\int_0^n\mathbf{1}_{\{T_k\le t\}}\mathbf{1}_{\{T_l\le t\}}\mathbf{1}_{\{T_j\le t\}}\mathrm{d}L_t\,.$$

Taking into account that the $(L_t)_{t\ge0}$ is independent of the field $\mathcal{G}_z=\sigma\{z_t\,,t\ge0\}$, we obtain that $\mathbf{E}\left(I_{klj}|\mathcal{G}_z\right)=0$. Therefore,

$$\begin{aligned}&\mathbf{E}\int_0^n(I_t^z(f))^2I_t^z(g)g(t)\,\mathrm{d}L_t\\ =\;&\mathbf{E}\sum_{k=1}^{N_n}\sum_{l=1}^{N_n}\sum_{j=1}^{N_n}f(T_k)\,f(T_l)\,g(T_j)\,Y_jY_l\,Y_k\,\mathbf{E}\left(I_{klj}|\mathcal{G}_z\right)=0.\end{aligned}$$

So, we obtain (2.101) and hence the proof is achieved. □

Appendix

Property of the penalty term

Lemma 2.6. *For any $n\ge1$ and $\lambda\in\Lambda$,*

$$P_n^0(\lambda)\le\mathbf{E}_QErr_n(\lambda)+\frac{\mathbf{C}_{1,Q,n}}{n},$$

where the coefficient $P_n^0(\lambda)$ was defined in (2.38).

Proof. By the definition of $\mathrm{Err}_n(\lambda)$ one has

$$\mathrm{Err}_n(\lambda) = \sum_{j=1}^{n} \left((\lambda(j) - 1)\theta_j + \frac{\lambda(j)}{n} \xi_{j,n} \right)^2 .$$

In view of Proposition 2.2, this leads to the desired result

$$\mathbf{E}_Q \, \mathrm{Err}_n(\lambda) \geq \frac{1}{n} \sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q \, \xi_{j,n}^2 \geq P_n^0(\gamma) - \frac{\mathbf{C}_{1,Q,n}}{n}.$$

Chapter 3

Non-parametric estimation for semi-Markov regression models based on discrete data

3.1 Introduction

Let us consider a regression model in continuous time

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \mathrm{d}\, \xi_t\,, \quad 0 \le t \le n\,, \tag{3.1}$$

where S is an unknown 1-periodic function defined on $\mathbb{R}$ with values on $\mathbb{R}$, $(\xi_t)_{t\ge 0}$ is the unobserved noise process (2.2) . The problem is to estimate the unknown function S in model (3.1) on the basis of observations

$$(y_{t_j})_{0\le j\le np}, \quad t_j = j\Delta, \quad \Delta = \frac{1}{p}, \tag{3.2}$$

where integer $p \ge 1$ is the observation frequency. In this chapter we use the risks defined in (2.4) and (2.5) for the distribution family $\mathcal{Q}_n$.

The goal of this chapter is to develop a robust efficient model selection method for the model (3.1) with the semi-Markov dependence having unknown distribution. We use the approach proposed by Konev and Pergamenshchikov in [17] and [18] for continuos time regression models with non martingale noises. Unfortunately, we cannot use directly their method for the semi-Markov regression models, since their tool essentially uses the fact that the Ornstein - Uhlenbeck dependence decreases with geometrical rate and obtain sufficiently quickly the"white noise" case. In this chapter we propose new analytical tools based on renewal methods, to obtain the sharp non-asymptotic oracle inequalities. And, as a consequence, we obtain robust efficiency for proposed model selection procedures.

3.2 Model selection

In this chapter we will use the trigonometric basis $(\mathrm{Tr}_j)_{j\ge 1}$ in $\mathbf{L}_2[0,1]$ defined in (1.15). By making use of this basis, we consider the discrete Fourier transform of S

$$S(t) = \sum_{j=1}^{p} \theta_{j,p}\, \mathrm{Tr}_j(t), \quad t \in \{t_1, ..., t_p\}, \tag{3.3}$$

where the Fourier coefficients are defined by

$$\theta_{j,p} = (S, \mathrm{Tr}_j)_p = \frac{1}{p}\sum_{i=1}^{p} S(t_i)\mathrm{Tr}_j(t_i). \tag{3.4}$$

In the sequel the corresponding norm will be denoted by $\|x\|_p^2 = (x,x)_p$. These Fourier coefficients $\theta_{j,p}$ can be estimated by

$$\widehat{\theta}_{j,p} = \frac{1}{n}\int_0^n \Psi_{j,p}(t)\mathrm{d}\, y_t \quad \text{and} \quad \Psi_{j,p}(t) = \sum_{l=1}^{np} \mathrm{Tr}_j(t_l) 1_{\{t_{l-1}<t\le t_l\}}\,. \tag{3.5}$$

We note that the system of functions $(\Psi_{j,p})_{1\le j\le p}$ is orthonormal in $\mathbf{L}_2[0,1]$ because

$$\int_0^1 \Psi_{j,p}(t)\Psi_{i,p}(t)\mathrm{d}\, t \ = (\mathrm{Tr}_j, \mathrm{Tr}_i)_p = 1_{\{i=j\}}\,.$$

In the sequel we need the Fourier coefficients for the function S with respect to the new basis $(\Psi_{j,p})_{1\le j\le p}$. These coefficiens can be writen as

$$\overline{\theta}_{j,p} = \int_0^1 S(t)\Psi_{i,p}(t)\mathrm{d}\, t \ = \theta_{j,p} + h_{j,p}, \tag{3.6}$$

where

$$h_{j,p}(S) = \sum_{l=1}^{p}\int_{t_{l-1}}^{t_l} \mathrm{Tr}_j(t_l)(S(t) - S(t_l))\mathrm{d}\, t\,.$$

From (3.1) it follows directly that these Fourier coefficients satisfy the equation

$$\widehat{\theta}_{j,p} = \overline{\theta}_{j,p} + \frac{1}{\sqrt{n}}\xi_{j,p}, \quad \text{where} \quad \xi_{j,p} = \frac{1}{\sqrt{n}} I_n(\Psi_{j,p})\,. \tag{3.7}$$

For any $0 \le t \le 1$ we estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(t) = \sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\Psi_{j,p}(t)\,, \tag{3.8}$$

where the weight vector $\lambda = (\lambda(1), ..., \lambda(n))$ belongs to some finite set Λ from $[0,1]^n$, $\widehat{\theta}_{j,p}$ was defined in (3.5). Now let us consider

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = \max_{\lambda\in\Lambda} L(\lambda)\,, \tag{3.9}$$

where $\#(\Lambda)$ is the cardinal number of Λ and $L(\lambda) = \sum_{j=1}^{n}\lambda(j)$. In the sequel we assume that $|\Lambda|_* \ge 1$ and $\lambda(j) = 0$ for $j \ge p$.

In order to find a proper weight sequence λ in the set Λ one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. Let as consider the empirical squared error

$$\mathrm{Err}(\lambda) = \|\widehat{S}_\lambda - S\|^2\,, \tag{3.10}$$

which in our case is equal to

$$\mathrm{Err}(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,p} - 2\sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,p}\overline{\theta}_{j,p} + \|S\|^2 . \tag{3.11}$$

Since the Fourier coefficients $(\theta_j)_{j\geq 1}$ are unknown, the weight coefficients $(\lambda(j))_{1\leq j\leq p}$ cannot be determined by minimizing this quality. To circumvent this difficulty, one needs to replace the terms $\widehat{\theta}_{j,p}\overline{\theta}_{j,p}$ by their estimators $\widetilde{\theta}_{j,p}$. Let us set

$$\widetilde{\theta}_{j,p} = \widehat{\theta}^2_{j,p} - \frac{\widehat{\sigma}_n}{n}, \tag{3.12}$$

where $\widehat{\sigma}_n$ is an estimate of the proxy variance σ_Q defined in (2.8). For, example, we can take it as

$$\widehat{\sigma}_n = \frac{n}{\check{p}} \sum_{j=l}^{\check{p}} \widehat{\theta}^2_{j,p} \quad \text{and} \quad \check{p} = \min(p, n) , \tag{3.13}$$

where $l = [\sqrt{n}]$, and we set $\widehat{\sigma}_n = 0$ *for* $l > p$. For this change in the empirical squared error, one has to pay some penalty. Thus, we obtain the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,n} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,n} + \delta\, P_n(\lambda), \tag{3.14}$$

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

$$P_n(\lambda) = \frac{\widehat{\sigma}_n|\lambda|^2}{n}. \tag{3.15}$$

Minimizing the cost function, that is

$$\hat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_n(\lambda), \tag{3.16}$$

and substituting the obtained weight coefficients $\hat{\lambda}$ in (3.8), lead to the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}. \tag{3.17}$$

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

3.3 Oracle inequality

In order to prove the oracle inequality, the following conditions will be needed for the noise $(\xi_t)_{t\geq 0}$. Here we use the conditions introduced in [17] for the general semi-martingale model (2.1).

$\mathbf{L}_1$) *For all* $n \geq 1$ *and* Q *there exist a variance proxy* $\sigma_Q > 0$ *and a constant* $\mathbf{L}_{1,Q} \geq 0$ *such that*

$$\sup_{p\geq 3} \sup_{x\in[-1,1]^n} \left|B_{1,Q}(x)\right| \leq \mathbf{L}_{1,Q} < \infty\,,$$

where $B_{1,Q}(x) = \sum_{j=1}^n x_j \left(\mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q\right)$.

$\mathbf{L}_2$) *For all* $n \geq 1$ *and* Q *there exists a constant* $\mathbf{L}_{2,Q} \geq 1$ *such that*

$$\sup_{p\geq 3} \sup_{|x|\leq 1} \mathbf{E}\, B_{2,Q}^2(x) \leq \mathbf{L}_{2,Q} < \infty,$$

where $B_{2,Q}(x) = \sum_{j=1}^n x_j\, \widetilde{\xi}_{j,p}$ *and* $\widetilde{\xi}_{j,p} = \xi_{j,p}^2 - \mathbf{E}_Q \xi_{j,p}^2$.

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

$$\mathbf{g}_{n,p} = 1 + |\Lambda|_* \left(\frac{\sqrt{n}}{\check{p}} + \frac{1}{\sqrt{\check{p}}}\right). \tag{3.18}$$

Firstly, we obtain the non asymptotic oracle inequality for the model selection procedure (3.17).

Theorem 3.1. *Assume that Conditions* $\mathbf{L}_1$) *and* $\mathbf{L}_2$) *hold. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, the weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17)*, satisfies the following oracle inequality*

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + l^* \frac{\check{\iota}}{\delta n} \left(\sigma_Q + |\Lambda|_* \mathbf{E}_Q |\widehat{\sigma}_n - \sigma_Q|\right).\end{aligned} \tag{3.19}$$

Proof. Using the cost function given in (3.14), we can rewrite the empirical squared error in (3.11) as follows

$$\mathrm{Err}(\lambda) = J_n(\lambda) + 2\sum_{j=1}^n \lambda(j)\check{\theta}_{j,p} + \|S\|^2 - \rho\hat{P}_n(\lambda), \tag{3.20}$$

where

$$\check{\theta}_{j,p} = \widetilde{\theta}_{j,p} - \overline{\theta}_{j,p}\widehat{\theta}_{j,p} = \frac{1}{\sqrt{n}}\overline{\theta}_{j,p}\xi_{j,p} + \frac{1}{n}\widetilde{\xi}_{j,p} + \frac{1}{n}\varsigma_{j,n} + \frac{\sigma_Q - \widehat{\sigma}_n}{n},$$

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

$$M(\lambda) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \lambda(j)\theta_j\xi_{j,p} \quad \text{and} \quad P_n^0 = \frac{\sigma_Q|\lambda|^2}{n}\,, \tag{3.21}$$

we can rewrite (3.20) as

$$\begin{aligned}\mathrm{Err}(\lambda) = J_n(\lambda) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n} L(\lambda) + 2M(\lambda) + \frac{2}{n} B_{1,Q}(\lambda) \\ + 2\sqrt{P_n^0(\lambda)}\frac{B_{2,Q}(e(\lambda))}{\sqrt{\sigma_Q n}} + \|S\|^2 - \rho P_n(\lambda),\end{aligned} \tag{3.22}$$

where $e(\lambda) = \lambda/|\lambda|$ and the function L was defined in (3.9). Let $\lambda_0 = (\lambda_0(j))_{1\le j\le p}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be defined as in (3.16). Substituting λ_0 and $\widehat{\lambda}$ in Equation (3.22), we obtain

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) - \mathrm{Err}(\lambda_0) =& J(\widehat{\lambda}) - J(\lambda_0) + 2\frac{\sigma_Q - \widehat{\sigma}_n}{n} L(\varpi) + \frac{2}{n} B_{1,Q}(\varpi) + 2M(\varpi) \\ &+ 2\sqrt{P_n^0(\widehat{\lambda})}\frac{B_{2,Q}(\widehat{e})}{\sqrt{\sigma_Q n}} - 2\sqrt{P_n^0(\lambda_0)}\frac{B_{2,Q}(e_0)}{\sqrt{\sigma_Q n}} \\ &- \delta P_n(\widehat{\lambda}) + \delta P_n(\lambda_0),\end{aligned} \tag{3.23}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{e} = e(\widehat{\lambda})$ and $e_0 = e(\lambda_0)$. Note that, by (3.9),

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

The inequality

$$2|ab| \le \delta a^2 + \delta^{-1} b^2 \tag{3.24}$$

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

$$2\sqrt{P_n^0(\lambda)}\frac{|B_{2,Q}(e(\lambda))|}{\sqrt{\sigma_Q n}} \le \delta P_n^0(\lambda) + \frac{B_{2,Q}^2(e(\lambda))}{\delta \sigma_Q\, n}.$$

Taking into account that $0 < \delta < 1$, we get

$$\begin{aligned}\mathrm{Err}(\hat{\lambda}) \le \mathrm{Err}(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{L}_{1,Q}}{n} + \frac{2B_{2,Q}^*}{\delta\sigma_Q\, n} \\ + \frac{1}{n}|\widehat{\sigma}_n - \sigma_Q|(|\widehat{\lambda}|^2 + |\lambda_0|^2) + 2\delta P_n(\lambda_0)\,,\end{aligned}$$

where $B_{2,Q}^* = \sup_{\lambda\in\Lambda} B_{2,Q}^2((e(\lambda))$. Moreover, noting that in view of (3.9) $\sup_{\lambda\in\Lambda} |\lambda|^2 \le |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \mathrm{Err}(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{L}_{1,Q}}{n} + \frac{2B_{2,Q}^*}{\delta\sigma_Q n} \\ &+ \frac{4|\Lambda|_*}{n}|\widehat{\sigma} - \sigma_Q| + 2\delta P_n(\lambda_0).\end{aligned} \tag{3.25}$$

To estimate the second term in the right side of this inequality we set

$$S_x = \sum_{j=1}^{n} x(j)\overline{\theta}_{j,p}\mathrm{Tr}_j\,, \quad x = (x(j))_{1\le j\le n} \in \mathbb{R}^n\,.$$

Thanks to (2.7) we estimate the term $M(x)$ for any $x \in \mathbb{R}^n$ as

$$\mathbf{E}_Q M^2(x) \le \varkappa_Q \frac{1}{n}\sum_{j=1}^{n} x^2(j)\overline{\theta}^2_{j,p} = \varkappa_Q \frac{1}{n}\|S_x\|^2. \tag{3.26}$$

To estimate this function for a random vector $x \in \mathbb{R}^n$, we set

$$Z^* = \sup_{x\varepsilon\Lambda_1} \frac{nM^2(x)}{\|S_x\|^2}\,, \quad \Lambda_1 = \Lambda - \lambda_0\,.$$

So, through Inequality (3.24), we get

$$2|M(x)| \le \delta\|S_x\|^2 + \frac{Z^*}{n\delta}. \tag{3.27}$$

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

$$\mathbf{E}_Q Z^* \le \sum_{x\in\Lambda_1} \frac{n\mathbf{E}_Q M^2(x)}{\|S_x\|^2} \le \sum_{x\in\Lambda_1} \varkappa_Q = \varkappa_Q \check{\iota}\,, \tag{3.28}$$

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

$$\|S_x\|^2 - \|\widehat{S}_x\|^2 = \sum_{j=1}^{n} x^2(j)(\overline{\theta}^2_{j,p} - \widehat{\theta}^2_j) \le -2M_1(x), \tag{3.29}$$

where $M_1(x) = n^{-1/2}\sum_{j=1}^{n} x^2(j)\overline{\theta}^2_{j,p}\xi_{j,n}$. Taking into account now that, for any $x \in \Lambda_1$, the components $|x(j)| \le 1$, we can estimate this term as in (3.26), i.e.

$$\mathbf{E}_Q\, M_1^2(x) \le \varkappa_Q\, \frac{\|S_x\|^2}{n}\,.$$

Similarly to the previous reasoning we set

$$Z_1^* = \sup_{x\varepsilon\Lambda_1} \frac{nM_1^2(x)}{\|S_x\|^2}$$

and we get

$$\mathbf{E}_Q\, Z_1^* \le \varkappa_Q\, \check{\iota}\,. \tag{3.30}$$

Using the same type of arguments as in (3.27), we can derive

$$2|M_1(x)| \le \delta\|S_x\|^2 + \frac{Z_1^*}{n\delta}. \tag{3.31}$$

From here and (3.29), we get

$$\|S_x\|^2 \le \frac{\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z_1^*}{n\delta(1-\delta)} \tag{3.32}$$

for any $0 < \delta < 1$. Using this bound in (3.27) yields

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

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

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

Using this bound in (3.25) we obtain

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{Z^* + Z_1^*}{n\delta(1-3\delta)} + \frac{2\mathbf{L}_{1,Q}}{n(1-3\delta)} + \frac{2B_{2,Q}^*}{\delta(1-3\delta)\sigma_Q n}\\ &+ \frac{(4|\Lambda|_* + 2)}{n(1-3\delta)}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).\end{aligned}$$

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

$$\begin{aligned}\mathrm{Err}(\widehat{\lambda}) \le& \frac{1+\delta}{1-3\delta}\mathrm{Err}(\lambda_0) + \frac{2(Z^* + Z_1^*)}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4B_{2,Q}^*}{\delta\sigma_Q n}\\ &+ \frac{(8|\Lambda|_* + 2)}{n}|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}\,P_n^0(\lambda_0).\end{aligned}$$

Now, in view of Condition $\mathbf{L}_2$), we estimate the expectation of the term $B_{2,Q}^*$ in (3.25) as

$$\mathbf{E}_Q\, B_{2,Q}^* \le \sum_{\lambda\in\Lambda} \mathbf{E}_Q B_{2,Q}^2(e(\lambda)) \le \imath\mathbf{L}_{2,Q}\,.$$

Now, taking into account that $|\Lambda|_* \ge 1$, we get

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \le& \frac{1+\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\imath}}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4\imath\mathbf{L}_{2,Q}}{\delta\sigma_Q n}\\ &+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)}P_n^0(\lambda_0).\end{aligned}$$

By using the upper bound for $P_n(\lambda_0)$ in Lemma 3.1, we obtain that

$$\begin{aligned}\mathcal{R}_Q(\widehat{S}_*, S) \le& \frac{1+3\delta}{1-3\delta}\mathcal{R}_Q(\widehat{S}_{\lambda_0}, S) + \frac{4\varkappa_Q\check{\imath}}{n\delta} + \frac{4\mathbf{L}_{1,Q}}{n} + \frac{4\imath\mathbf{L}_{2,Q}}{\delta\sigma_Q n}\\ &+ \frac{10|\Lambda|_*}{n}\,\mathbf{E}_Q\,|\widehat{\sigma} - \sigma_Q| + \frac{2\delta\mathbf{L}_{1,Q}}{(1-3\delta)n}\,.\end{aligned}$$

Taking into account that $1 - 3\delta \ge 1/2$ for $0 < \delta < 1/3$ and that $\kappa_Q \le (1 + \check{\tau}|\rho|_*)\sigma_Q$ and using the bounds (3.38) and (3.60) we obtain the inequality (3.19). Hence we get the desired result. □

Corollary 3.1. *Assume that Conditions* $\mathbf{L}_1$*) and* $\mathbf{L}_2$*) hold and the proxy variance* σ_Q *is known. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, for any weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17) *with* $\widehat{\sigma}_n = \sigma_Q$*, satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) + l^* \frac{\sigma_Q \check{\iota}}{\delta n} \,. \tag{3.33}$$

Now we study the model selection procedure (3.17) using the proxy estimate (3.13).

Theorem 3.2. *Assume that the function* S *is continuously differentiable and the conditions* $\mathbf{L}_1$*) and* $\mathbf{L}_2$*) hold true. Then there exists some constant* $l^* > 0$ *such that for any noise distribution* Q*, for any weight vectors set* Λ*, for any periodic function* S *for any* $n \geq 1$*,* $p \geq 3$ *and* $0 < \delta \leq 1/6$*, the procedure* (3.17)*, satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_Q(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + l^* \frac{\nu}{\delta n} (1+\sigma_Q)^3 \left(1 + \|\dot{S}\|^2\right) \mathbf{g}_{n,p} \,. \end{aligned} \tag{3.34}$$

Let us study the robust risks (3.4) for the procedure (3.17). In this case this family consists of all distributions on the Skorokhod space $\mathcal{D}[0,n]$ with the parameters satisfying the conditions (2.8) – (2.9) . Now, to obtain the efficiency property we use the weight coefficients $(\lambda(j))_{1\leq j\leq n}$ specified in (2.28).

Our goal is to bound asymptotically the term (3.18) by any power of n. To this end we assume the following condition.

$\mathbf{H}_5$) *Assume that there exists* $\check{\delta} > 0$ *such that for any* $n \geq 3$

$$p \geq n^{5/6} \,. \tag{3.35}$$

Now Theorem 3.2 implies the following oracle inequality.

Theorem 3.3. *Assume that the unknown function* S *is continuously differentiable. Moreover, assume that Conditions* $\mathbf{H}_1$*)–*$\mathbf{H}_5$*) hold. Then for the robust risks defined in* (3.4) *through the distribution family* (2.8) – (2.9)*, the procedure* (3.17) *with the coefficients* (2.28)*, for any* $n \geq 1$ *and* $0 < \delta < 1/6$*, satisfies the following oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \leq \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*(\widehat{S}_\lambda, S) + \frac{\mathbf{U}_n^*(S)}{n\delta} \,, \tag{3.36}$$

where the sequence $\mathbf{U}_n^*(S) > 0$ *is such that under condition* (2.26)*, for any* $r > 0$ *and* $\check{\delta} > 0$*,*

$$\lim_{n\to\infty} \sup_{\|\dot{S}\|\leq r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0 \,. \tag{3.37}$$

Now we need to check the conditions $\mathbf{L}_1$) and $\mathbf{L}_2$) for the process (2.2).

Proposition 3.1. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then*

$$\mathbf{L}_{1,Q} = 2\check{\tau}\,\|\Upsilon\|_1\,\sigma_Q\,. \tag{3.38}$$

Proof. Firstly, we set

$$I_n^L(f) = \int_0^n f(t)\mathrm{d}L_t \quad \text{and} \quad I_n^z(f) = \int_0^n f(t)\mathrm{d}z_t\,. \tag{3.39}$$

In view of (1.6) the last integral can be represented as

$$I_n^z(f) = \sum_{l=1}^{\infty} f(T_l)Y_l\mathbf{1}_{\{T_l\le n\}}\,. \tag{3.40}$$

Therefore,

$$\xi_{j,n} = \frac{\varrho_1}{\sqrt{n}}I_n^L(\Psi_{j,p}) + \frac{\varrho_2}{\sqrt{n}}I_n^z(\Psi_{j,p})$$

and

$$\mathbf{E}\xi_{j,n}^2 = \frac{\varrho_1^2}{n}\int_0^n \Psi_{j,p}^2(t)\mathrm{d}\,t + \frac{\varrho_2^2}{n}\mathbf{E}\sum_{l=1}^{\infty}\Psi_{j,p}^2(T_l)1_{\{T_l\le n\}}\,. \tag{3.41}$$

Using Proposition 5.1 we get

$$\begin{aligned}\mathbf{E}\sum_{l=1}^{\infty}\Psi_{j,p}^2(T_l)1_{\{T_l\le n\}} &= \int_0^n \Psi_{j,p}^2(x)\,\rho(x)\mathrm{d}x\\ &= \frac{1}{\check{\tau}}\int_0^n \Psi_{j,p}^2(x)\mathrm{d}x + \int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x\,,\end{aligned}$$

where ρ is the renewal density introduced in (1.8). Then we obtain,

$$\mathbf{E}\xi_{j,n}^2 = \sigma_Q + \frac{\varrho_2^2}{n}\int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x$$

and

$$\sup_{j\ge1}\left|\int_0^n \Psi_{j,p}^2(x)\Upsilon(x)\mathrm{d}x\right| \le 2\|\Upsilon\|_1\,, \tag{3.42}$$

where $\sigma_Q = \varrho_1^2 + \varrho_2^2/\check{\tau}$. This directly implies the desired result. □

To study the function $B_{2,Q}(x)$, we have to analyze the correlation properties for the following stochastic integrals

$$\widetilde{I}_n(f) = I_n^2(f) - \mathbf{E}I_n^2(f)\,. \tag{3.43}$$

To do this we set

$$\check{c}_1 = 1 + \Pi(x^4) + \|\Upsilon\|_1^2 + |\rho|_* \quad \text{and} \quad \check{c}_2 = 12(1+\check{\tau})^2\,(1+\check{c}_1)\,. \tag{3.44}$$

Now we investigate the behavior of the integrals defined in (3.43) as functions of f.

Proposition 3.2. *For any left continuous functions $f, g : (0, \infty) \longrightarrow R$ such that $\|f\|_* \leq 1$, $\|g\|_* \leq 1$, we have*

$$|\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g)| \leq 12\sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\check{c}_1\right). \tag{3.45}$$

Proof. By Ito's formula one gets

$$\mathrm{d}I_t^2(f) = 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\check{\varrho}^2 f^2(t)\mathrm{d}\,t + \sum_{0\leq s\leq t} f^2(s)(\Delta\xi_s^d)^2\,, \tag{3.46}$$

where $\xi_t^d = \varrho_3\check{L}_t + \varrho_2 z_t$ and $\varrho_3 = \varrho_1\sqrt{1-\check{\varrho}^2}$. Taking into account that the processes $(\check{L}_t)_{t\geq 0}$ and $(z_t)_{t\geq 0}$ are independent and the time of jumps T_k defined in (1.7) has a density, we have $\Delta z_s \Delta\check{L}_s = 0$ a.s. for any $s \geq 0$. Therefore, we can rewrite the differential (3.46) as

$$\begin{aligned}\mathrm{d}I_t^2(f) =& 2I_{t-}(f)\mathrm{d}I_t(f) + \varrho_1^2\check{\varrho}^2 f^2(t)\mathrm{d}\,t \\ &+ \varrho_3^2\mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta\check{L}_s)^2 + \varrho_2^2\mathrm{d}\sum_{0\leq s\leq t} f^2(s)(\Delta z_s)^2\,.\end{aligned} \tag{3.47}$$

Therefore, using Lemma 2.1 we obtain

$$\mathbf{E}I_t^2(f) = \varrho_1^2\,\|f\|_t^2 + \varrho_2^2\|f\sqrt{\rho}\|_t^2\,,$$

where $\|f\|_t^2 = \int_0^t f^2(u)\mathrm{d}u$, ρ is the density of the renewal measure $\sum_{j=1}^{\infty}\eta^{(j)}$ and with η the distribution of τ_1. Therefore,

$$\mathrm{d}\widetilde{I}_t(f) = 2I_{t-}(f)f(t)d\xi_t + f^2(t)\mathrm{d}\widetilde{m}_t\,, \quad \widetilde{m}_t = \varrho_3^2\check{m}_t + \varrho_2^2 m_t\,, \tag{3.48}$$

where $\check{m}_t = \sum_{0\leq s\leq t}(\Delta\check{L}_s)^2 - t$ and $m_t = \sum_{0\leq s\leq t}(\Delta z_s)^2 - \int_0^t\rho(s)\mathrm{d}s$. By the Ito's formula we get

$$\begin{aligned}\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g) =& \mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g) \\ &+ \mathbf{E}\int_0^n \widetilde{I}_{t-}(g)\mathrm{d}\widetilde{I}_t(f) + \mathbf{E}\left[\widetilde{I}(f),\widetilde{I}(g)\right]_n.\end{aligned} \tag{3.49}$$

First, note that the process $(\check{m}_t)_{t\geq 0}$ is a martingale and, using Lemma 2.5, we get

$$\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g) = \rho_2^2\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)g^2(t)\mathrm{d}m_t = \rho_2^2\mathbf{E}\int_0^n I_{t-}^2(f)g^2(t)\mathrm{d}m_t\,.$$

The last integral can be represented as

$$\mathbf{E}\int_0^n I_{t-}^2(f)g^2(t)\mathrm{d}m_t = J_1 - J_2\,,$$

where

$$J_1 = \mathbf{E}\sum_{k\geq 1} I_{T_k-}^2(f)g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad J_2 = \int_0^n \mathbf{E}\,I_t^2(f)g^2(t)\rho(t)\mathrm{d}t\,.$$

By Lemma 2.3 we get

$$J_1 = \mathbf{E}\sum_{k\geq 1} \mathbf{E}\left(I^2_{T_k-}(f)|\mathcal{G}\right) g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} = \varrho_1^2 J_{1,1} + \varrho_2^2 J_{1,2}\,,$$

where

$$J_{1,1} = \mathbf{E}\sum_{k\geq 1} \|f\|^2_{T_k} g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} \quad \text{and} \quad J_{1,2} = \mathbf{E}\sum_{k\geq 1}\sum_{l=1}^{k-1} f^2(T_l)\, g^2(T_k)\mathbf{1}_{\{T_k\leq n\}}\,.$$

We obtain directly that

$$J_{1,1} = \int_0^n \|f\|^2_t g^2(t)\rho(t)\mathrm{d}t$$

and

$$J_{1,2} = \mathbf{E}\sum_{l\geq 1} f^2(T_l)\sum_{k\geq l+1} g^2(T_k)\mathbf{1}_{\{T_k\leq n\}} = \int_0^n f^2(x)\left(\int_0^{n-x} g^2(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x.$$

From Lemma 2.1 we obtain that

$$J_2 = \varrho_1^2\int_0^n \|f\|^2_t g^2(t)\rho(t)\mathrm{d}t + \varrho_2^2\int_0^n \|f\sqrt{\rho}\|^2_t g^2(t)\rho(t)\mathrm{d}t\,.$$

Therefore,

$$\mathbf{E}\int_0^n I^2_{t-}(f)g^2(t)\mathrm{d}m_t = \varrho_2^2\int_0^n f^2(x)\left(\int_x^n g^2(t)(\rho(t-x)-\rho(t))\mathrm{d}t\right)\rho(x)\mathrm{d}x\,.$$

Taking into account that $\rho(t-x)-\rho(t) = \Upsilon(t-x)-\Upsilon(t)$ we can estimate the last integral as

$$|\mathbf{E}\int_0^n I^2_{t-}(f)g^2(t)\mathrm{d}m_t| \leq 2\varrho_2^2 n\|\Upsilon\|_1\,.$$

From this and by the symmetry arguments we obtain that

$$|\mathbf{E}\int_0^n \widetilde{I}_{t-}(f)\mathrm{d}\widetilde{I}_t(g)| + |\mathbf{E}\int_0^n \widetilde{I}_{t-}(g)\mathrm{d}\widetilde{I}_t(f)| \leq 4\varrho_2^4 n\|\Upsilon\|_1\,. \tag{3.50}$$

Note now that

$$\left[\widetilde{I}(f),\widetilde{I}(g)\right]_n = \left\langle \widetilde{I}^c(f),\widetilde{I}^c(g)\right\rangle_n + \mathbf{D}_n(f,g)\,, \tag{3.51}$$

where

$$\mathbf{D}_n(f,g) = \sum_{0\leq t\leq n}\Delta\widetilde{I}^d_t(f)\Delta\widetilde{I}^d_t(g)\,.$$

It should be noted that the continuous and the discrete parts of the processes (3.48) can be represented as

$$\widetilde{I}^c_t(f) = 2\varrho_1\check{\varrho}\int_0^t I_s(f)f(s)\mathrm{d}w_s \quad \text{and} \quad \widetilde{I}^d_t(g) = 2\int_0^t I_{s-}(f)f(s)d\xi^d_s + \int_0^t f^2(s)\mathrm{d}\widetilde{m}_s\,.$$

So, in view of Lemma 2.1,

$$\begin{aligned}\mathbf{E} < \widetilde{I}^c(f), \widetilde{I}^c(g) >_n &= 4\rho_1^2 \breve{\varrho}^2 \int_0^n \mathbf{E}(I_t(f) I_t(g)) f(t) g(t) \mathrm{d}t \\ &= 4\rho_1^4 \breve{\varrho}^2 \int_0^n (f,g)_t \, f(t) g(t) \mathrm{d}t + 4\rho_1^2 \rho_2^2 \breve{\varrho}^2 \int_0^n (f, g\rho)_t f(t) g(t) \mathrm{d}t \\ &= 4\rho_1^2 \breve{\varrho}^2 \sigma_Q \, (f,g)_n^2 + 4\rho_1^2 \rho_2^2 \breve{\varrho}^2 \int_0^n (f, g\Upsilon)_t f(t) g(t) \mathrm{d}t \,, \end{aligned} \tag{3.52}$$

with $(f,g)_t = \int_0^t f(s) g(s) \mathrm{d}s$. Taking into account that $\|f\|_* \le 1$ and $\|g\|_* \le 1$, we can estimate the last integral as

$$\int_0^n (f, g\Upsilon)_t f(t) g(t) \mathrm{d}t \le n \|\Upsilon\|_1 \,.$$

Therefore,

$$\left| \mathbf{E} \left\langle \widetilde{I}^c(f), \widetilde{I}^c(g) \right\rangle_n \right| \le 4\sigma_Q^2 \left((f,g)_n^2 + n\check{\tau} \|\Upsilon\|_1 \right) . \tag{3.53}$$

To study the last term in (3.51) note that

$$\mathbf{D}_n(f,g) = \sum_{0 \le t \le n} \left(2 I_{t-}(f) f(t) \Delta \xi_t^d + f^2(t) \Delta \widetilde{m}_t \right) \left(2 I_{t-}(g) g(t) \Delta \xi_t^d + g^2(t) \Delta \widetilde{m}_t \right) .$$

Taking into account that for any $t > 0$

$$\Delta \xi_t^d \Delta \widetilde{m}_t = \varrho_3^3 (\Delta \check{L}_t)^3 + \varrho_2^3 (\Delta z_t)^3 \,,$$

we obtain that

$$\mathbf{E} \sum_{0 \le t \le n} I_{t-}(f) f(t) g^2(t) \Delta \xi_t^d \, \Delta \widetilde{m}_t = \left(\varrho_3^3 \Pi(x^3) \, + \varrho_2^3 \mathbf{E} \, Y_1^3 \right) \int_0^n \mathbf{E} \, I_t(f) f(t) g^2(t) \mathrm{d}t = 0 \,.$$

So, using the symmetry arguments, we find that

$$\mathbf{E} \mathbf{D}_n(f,g) = 4 \mathbf{E} \, \mathbf{D}_{1,n}(f,g) + \mathbf{E} \, \mathbf{D}_{2,n}(f,g) \,, \tag{3.54}$$

where

$$\mathbf{D}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta \xi_t^d)^2 \quad \text{and} \quad \mathbf{D}_{2,n}(f,g) = \sum_{0 \le t \le n} f^2(t) \, g^2(t) (\Delta \widetilde{m}_t)^2 \,.$$

Note that

$$\mathbf{D}_{1,n}(f,g) = \varrho_3^2 \check{\mathbf{D}}_{1,n}(f,g) + \varrho_2^2 \widetilde{\mathbf{D}}_{1,n}(f,g) \,,$$

where

$$\check{\mathbf{D}}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta \check{L}_t)^2$$

and

$$\widetilde{\mathbf{D}}_{1,n}(f,g) = \sum_{0 \le t \le n} I_{t-}(f) I_{t-}(g) f(t) g(t) (\Delta z_t)^2 \,.$$

Now, similarly to (3.52) and taking into account that $\Pi(x^2)=1$, we get

$$\begin{aligned}\mathbf{E}\,\check{\mathbf{D}}_{1,n}(f,g) &= \int_0^n f(t)g(t)\mathbf{E}\,I_t(f)I_t(g)\,\mathrm{d}t = \varrho_1^2\int_0^n f(t)g(t)\,(f,g)_t\,\mathrm{d}t \\ &\quad + \varrho_2^2\int_0^n f(t)g(t)\,(f,g\rho)_t\,\mathrm{d}t \\ &= \sigma_Q(f,g)_n^2 + \varrho_2^2\int_0^n f(t)g(t)\,(f,g\Upsilon)_t\,\mathrm{d}t\,.\end{aligned}$$

So,

$$|\mathbf{E}\,\check{\mathbf{D}}_{1,n}(f,g)| \le \sigma_Q\left((f,g)_n^2 + n\check{\tau}\|\Upsilon\|_1\right)\,. \tag{3.55}$$

Moreover, taking into account that $\mathbf{E}Y_1^2=1$ we get

$$\mathbf{E}\,\widetilde{\mathbf{D}}_{1,n}(f,g) = \mathbf{E}\sum_{k\ge1} I_{T_k-}(f)I_{T_k-}(g)f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}}\,.$$

So, in view of Lemma 2.3

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

where

$$\mathbf{D}_{1,n}^{'}(f,g) = \sum_{k\ge1}\sum_{l=1}^{k-1} f(T_l)\,g(T_l)f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}}\,.$$

Noting now that

$$\int_0^n (f,g)_t\,f(t)g(t)\rho(t)\mathrm{d}t = \frac{1}{2\check{\tau}}(f,g)_n^2 + \int_0^n (f,g)_t\,f(t)g(t)\Upsilon(t)\mathrm{d}t\,,$$

we obtain

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

Furthermore, the expectation of $\mathbf{D}_{1,n}^{'}(f,g)$ can be represented as

$$\begin{aligned}\mathbf{E}\,\mathbf{D}_{1,n}^{'}(f,g) &= \mathbf{E}\sum_{l\ge1} f(T_l)\,g(T_l)\sum_{k\ge l+1} f(T_k)g(T_k)\,\mathbf{1}_{\{T_k\le n\}} \\ &= \int_0^n f(x)g(x)\left(\int_0^{n-x} f(x+t)g(x+t)\rho(t)\mathrm{d}t\right)\rho(x)\mathrm{d}x \\ &= \frac{1}{2\check{\tau}}(f,g)_n^2 + \mathbf{D}_{1,n}^{''}(f,g)\,,\end{aligned}$$

where the last term in this equality can be represented as

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

This implies

$$|\mathbf{D}^{''}_{1,n}(f,g)| \le n(1+\frac{1}{\check{\tau}})(1+\|\Upsilon\|_1^2)\,.$$

Therefore,

$$|\mathbf{E}\,\widetilde{\mathbf{D}}_{1,n}(f,g)| \le \sigma_Q\left((f,g)_n^2 + n(1+\check{\tau})\|\Upsilon\|_1^2\right)\,. \tag{3.56}$$

Finally we obtain that

$$|\mathbf{E}\,\mathbf{D}_{1,n}(f,g)| \le \sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\|\Upsilon\|_1^2\right)\,. \tag{3.57}$$

As to the last term in (3.54) we can calculate directly

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

i.e.

$$\mathbf{E}\,\mathbf{D}_{2,n}(f,g) \le n\sigma_Q^2\left(\Pi(x^4)+|\rho|_*\right)(1+\check{\tau})^2\,.$$

From here we obtain that

$$|\mathbf{E}\,\mathbf{D}_n(f,g)| \le \sigma_Q^2(1+\check{\tau})^2\left(4(f,g)_n^2 + n\check{c}_1\right)\,, \tag{3.58}$$

where $\check{c}_1$ is given in (3.44). From this and (3.53) we find

$$\mathbf{E}[\widetilde{I}(f),\widetilde{I}(g)]_n \le 8\sigma_Q^2(1+\check{\tau})^2\left((f,g)_n^2 + n\check{c}_1\right)\,. \tag{3.59}$$

This bound and (3.50) implies (3.45). Hence we get the desired result. □

Using these properties we can obtain the following bound.

Proposition 3.3. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for all* $n \ge 1$,

$$\mathbf{L}_{2,Q} = \check{c}_2\,\sigma_Q^2\,, \tag{3.60}$$

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

Proof. Note that

$$\mathbf{E}\left(\sum_{j=2}^{n} x_j \widetilde{\xi}_{j,p}\right)^2 \le \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n} |x_j|\,|x_l|\,|\mathbf{E}\widetilde{I}_n(\Psi_{j,p})\widetilde{I}_n(\Psi_{l,p})|\,.$$

Using here Proposition 3.2 and taking into account that

$$(\Psi_{j,p}\,,\,\Psi_{l,p})_n = \int_0^n \Psi_{j,p}(t)\Psi_{l,p}(t)\mathrm{d}t = n\mathbf{1}_{\{j=l\}}\,,$$

we obtain the bound (3.60). Hence we obtain the desired result. □

Now we can study the estimate (3.17).

Proposition 3.4. *Assume that Conditions* $\mathbf{H}_1$*) and* $\mathbf{H}_4$*) hold true for the model* (3.1) *and that* S *is continuously differentiable. Then, for any* $n \ge 2$ *and* $p \ge 3$*,*

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \sigma_Q| \le \check{c}_3\left(\frac{\sqrt{n}}{\check{p}} + \frac{1}{\sqrt{\check{p}}}\right)(1+\|\dot{S}\|^2)(1+\sigma_Q)^2\,, \tag{3.61}$$

where $\check{c}_3 = 6\,(14 + 2|\rho|_* + 3\sqrt{1+\check{c}_1})\,(1+\check{\tau})$.

Proof. It is clear that Inequality (3.61) holds true for $l > \check{p}$. Let now $l \leqslant \check{p}$. Setting $x_j^{'} = \mathbf{1}_{\{[\sqrt{n}]\leqslant j\leqslant\check{p}\}}$ and subtituting (3.7) in (3.13) yields

$$\widehat{\sigma}_n = \frac{n}{\check{p}}\sum_{j=l}^{\check{p}}(\overline{\theta}_{j,p})^2 + \frac{2n}{\check{p}}\,M(x^{'}) + \frac{1}{\check{p}}\sum_{j=l}^{\check{p}}\xi_{j,p}^2\,, \tag{3.62}$$

where $M(x^{'})$ is defined in (3.21). Furthermore, putting $x_j^{''} = \check{p}^{-1/2}\mathbf{1}_{\{l\leqslant j\leqslant\check{p}\}}$, one can write the last term on the right hand side of (3.62) as

$$\frac{1}{\check{p}}\sum_{j=l}^{\check{p}}\xi_{j,p}^2 = \frac{1}{\sqrt{\check{p}}}\,B_{2,Q}(x^{''}) + \frac{1}{\check{p}}B_{1,Q}(x^{'}) + \frac{(\check{p}-l+1)\sigma_Q}{\check{p}}\,,$$

where the functions $B_{1,Q}$ and $B_{2,Q}$ are given in conditions $\mathbf{L}_1$) and $\mathbf{L}_2$). Using Proposition 3.1, Proposition 3.3 and Lemma 3.3 , we come to the following upper bound

$$\mathbf{E}_Q|\widehat{\sigma}_n - \sigma_Q| \le \frac{16\|\dot{S}\|^2 n}{lp} + \frac{2n}{p}\mathbf{E}_Q\,|\mathbf{M}(x^{'})| + \frac{\mathbf{L}_{1,Q}}{p} + \frac{\sqrt{\mathbf{L}_{2,Q}}}{\sqrt{p}} + \frac{\sigma_Q(l-1)}{p}.$$

In the same way as in (3.26), we obtain

$$\mathbf{E}_Q\,|M(x^{'})| \le \left(\frac{\varkappa_Q}{n}\sum_{j=l}^{p}\overline{\theta}_{j,p}^2\right)^{1/2} \le \frac{4(\varkappa_Q\|\dot{S}\|^2)^{1/2}}{l}.$$

Taking into account that $\kappa_Q \le (1+\check{\tau}|\rho|_*)\sigma_Q$ and using the bounds (3.38) and (3.60) we obtain the inequality (3.61). Hence we obtain the desired result. □

Remark 3.1. Propositions 3.1 and 3.3 are used to obtain the oracle inequalities given in Section 4.4 (see, for example, [17]).

3.4 Efficiency

Now we study the asymptotically efficiency properties for the procedure (3.17) with the coefficients (2.28), with respect to the robust risk (3.4) defined by the distribution family (2.8) – (2.9). To this end, we assume that the unknown function S in the model (3.1) belongs to the Sobolev ball

$$W_{\mathbf{r}}^{k} = \{f \in \mathcal{C}_{per}^{k}[0,1]\,, \sum_{j=0}^{k} \|f^{(j)}\|^2 \leq \mathbf{r}\}\,, \tag{3.63}$$

where $r > 0\,,\ k \geq 1$ are some parameters, $\mathcal{C}_{per}^{k}[0,1]$ 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 $0 \leq i \leq k$. The function class $W_{\mathbf{r}}^{k}$ can be written as an ellipsoid in $\mathbf{L}_2$, i.e.

$$W_{\mathbf{r}}^{k} = \{f \in \mathcal{C}_{per}^{k}[0,1] : \sum_{j=1}^{\infty} a_j\,\theta_j^2 \leq \mathbf{r}\}, \tag{3.64}$$

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

Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (3.4) is given by

$$\mathbf{r}_k^* = ((2k+1)\mathbf{r})^{1/(2k+1)} \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)}. \tag{3.65}$$

Note that this is the well-known Pinsker's constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]).

Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the σ-algebra $\sigma\{y_t\,,\ 0 \leq t \leq n\}$ generated by the process (3.1).

Theorem 3.4. *Under the conditions* (2.8) *and* (2.9),

$$\liminf_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n \in \Pi_n} \sup_{S \in W_{\mathbf{r}}^{k}} \mathcal{R}_n^*(\widehat{S}_n, S) \geq \mathbf{r}_k^*\,, \tag{3.66}$$

where $\upsilon_n = n/\varsigma^*$.

Note that if the parameters r and k are known, i.e. for the non-adaptive estimation case, then to obtain the efficient estimation for the "signal+white noise"model. Pinsker in [47] proposed to use the estimate $\widehat{S}_{\lambda_0}$ defined in (3.8) with the weights (2.28) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, l_0)\,, \tag{3.67}$$

where $l_0 = [\mathbf{r}/\varepsilon]\varepsilon$. For the model (3.1) – (2.2) we show the same result.

Proposition 3.5. *The estimator* $\widehat{S}_{\lambda_0}$ *satisfies the following asymptotic upper bound*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^{k}} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^*\,.$$

Proof. First, we note that in view of (3.8) one can represent the quadratic risk for the empiric norm $\|\cdot\|_p$ as

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

where $\overline{\Theta}_p = \sum_{j=1}^{p}\left(\theta_{j,p} - \lambda_0(j)\,\overline{\theta}_{j,p}\right)^2$. We put here $\lambda_0(j) = 0$ for $j > n$ if $p > n$. The first term can be estimated by the bound (3.38) as

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

where $\mathbf{L}_{1,n}^* = \sup_{Q\in\mathcal{Q}_n} \mathbf{L}_{1,Q}$. Therefore, taking into account that $\upsilon_n = n/\sigma^*$, we get

$$\sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q \|\widehat{S}_{\lambda_0} - S\|_p^2 \le \frac{1}{\upsilon_n}\sum_{j=1}^{n} \lambda_0^2(j) + \frac{\mathbf{L}_{1,n}^*}{n} + \overline{\Theta}_p\,.$$

Note that

$$\lim_{n\to\infty} \frac{1}{\upsilon_n^{1/(2k+1)}} \sum_{j=1}^{n} \lambda_0^2(j) = \frac{2(\tau_k\,\mathbf{r})^{1/(2k+1)}\,k^2}{(k+1)(2k+1)}\,. \tag{3.68}$$

Furthermore, by Inequality (3.24) for any $0 < \widetilde{\varepsilon} < 1$ we get

$$\overline{\Theta}_p \le (1+\widetilde{\varepsilon})\,\Theta_p + (1+\widetilde{\varepsilon}^{-1}) \sum_{j=1}^{p} h_{j,p}^2\,, \tag{3.69}$$

where $\Theta_p = \sum_{j=1}^{p}(1-\lambda_0(j))^2\,\theta_{j,p}^2$. In view of Definition (2.28), we can represent this term as

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

where $\iota_0 = j_*(\alpha_0)$, $\omega_0 = \omega_{\alpha_0} = (\tau_k l_0 \upsilon_n)^{1/(2k+1)}$ and $l_0 = [\mathbf{r}/\varepsilon]\,\varepsilon$. Applying Lemma 3.5 yields

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

Similarly, through Lemma 3.4 we have

$$\Theta_{2,p} \le (1+\widetilde{\varepsilon}) \sum_{j\ge[\omega_0]+1} \theta_j^2 + (1+\widetilde{\varepsilon}^{-1})\,\mathbf{r}\,p^{-2}\,.$$

Hence,

$$\Theta_p \le (1+\widetilde{\varepsilon})\,\Theta_{\iota_0}^* + (1+\widetilde{\varepsilon}^{-1})\left(4\pi^2 r\omega_0^3 + \mathbf{r}\right) p^{-2}\,,$$

where $\Theta_l^* = \sum_{j\geq l} (1-\lambda_0(j))^2\, \theta_j^2$. Moreover, note that

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

Moreover, $W_{\mathbf{r}}^k \subseteq W_{\mathbf{r}}^2$ for any $k\geq 2$. From here and Lemma 3.6 we get

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

Moreover, in view of Condition $\mathbf{H}_5$) we have

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

So,

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \overline{\Theta}_p \leq \limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \Theta_{\iota_0}^*\,.$$

To estimate the term $\Theta_{\iota_0}^*$ we set

$$\mathbf{U}_n = \upsilon_n^{2k/(2k+1)} \sup_{j\geq \iota_0} (1-\lambda_0(j))^2/a_j\,,$$

where the sequence $(a_j)_{j\geq 1}$ is defined in (3.64). This leads to the inequality

$$\sup_{S\in W_{\mathbf{r}}^1} \upsilon_n^{2k/(2k+1)}\, \Theta_{\iota_0}^* \leq \mathbf{U}_n \sum_{j\geq 1} a_j\, \theta_j^2 \leq \mathbf{U}_n\, \mathbf{r}\,.$$

Taking into account that $\lim_{n\to\infty} t_0 = \mathbf{r}$, we get

$$\limsup_{n\to\infty} \mathbf{U}_n \leq \pi^{-2k}\, (\tau_k\, \mathbf{r})^{-2k/(2k+1)}\,,$$

where the coefficient τ_k is given in (2.28). This implies immediately that

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_r^k} \overline{\Theta}_p \leq \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2k/(2k+1)}}\,. \tag{3.70}$$

Moreover, note that

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

So, applying (3.68) and (3.70), yields

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q\, \|\widehat{S}_{\lambda_0} - S\|_p^2 \leq R_k^*\,. \tag{3.71}$$

Furthermore, Lemma 3.2 yields that for any $\widetilde{\varepsilon} > 0$

$$\sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq (1+\widetilde{\varepsilon}) \sup_{S\in W_{\mathbf{r}}^k} \sup_{Q\in\mathcal{Q}_n} \mathbf{E}_Q\, \|\widehat{S}_{\lambda_0} - S\|_p^2 + (1+\widetilde{\varepsilon}^{-1})\mathbf{r}\, p^{-2}\,.$$

So, in view of Condition $\mathbf{H}_5$), we derive the desired inequality

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \le R_k^* \,.$$

Hence the conclusion follows. □

For the adaptive estimation we use the model selection procedure (3.17) with the parameter δ defined as a function of n satisfying

$$\lim_{n\longrightarrow\infty} \delta_n = 0 \quad \text{and} \quad \lim_{n\longrightarrow\infty} n^{\check{\delta}}\, \delta_n = 0 \tag{3.72}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$.

Theorem 3.5. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_5$) *hold. Then the robust risk defined in* (3.4) *through the distribution family* (2.8) – (2.9) *for the procedure* (3.17) *with the coefficients* (2.28) *and the parameter* $\delta = \delta_n$ *satisfying* (3.72) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} \upsilon_n^{2k/(2k+1)} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le \mathbf{r}_k^* \,. \tag{3.73}$$

Theorem 3.4 and Theorem 3.5 imply the following result.

Corollary 3.2. *Under the conditions of Theorem 3.5,*

$$\lim_{n\to\infty} \upsilon_n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\widehat{S}_n, S) = \mathbf{r}_k^* \,. \tag{3.74}$$

Remark 3.2. It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball $W_{\mathbf{r}}^k$ is $n^{2k/(2k+1)}$ (see, for example, [47], [46]). We see here that the efficient robust rate is $\upsilon_n^{2k/(2k+1)}$, i.e. if the distribution upper bound $\varsigma^* \to 0$ as $n \to \infty$ we obtain a faster rate with respect to $n^{2k/(2k+1)}$, and if $\varsigma^* \to \infty$ as $n \to \infty$ we obtain a slower rate. In the case when ς^* is constant the robust rate is the same as the classical non robuste convergence rate.

3.5 Simulations

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (3.17). In (3.1) we chose a 1-periodic function which for $0 \le t \le 1$ is defined as

$$S(t) = \begin{cases} |t - \frac{1}{2}| & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4}\,, \\ \\ \frac{1}{4} & \text{else.} \end{cases} \tag{3.75}$$

We simulate the model

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t \,,$$

n	$\overline{\mathbf{R}}$	$\overline{\mathbf{R}_*}$
20	0.0398	0.211
100	0.0091	0.0483
200	0.0067	0.0355
1000	0.0022	0.0116

Table 3.1: Empirical risks

where $\xi_t = 0.5\mathrm{d}wt + 0.5\mathrm{d}z_t$. Here z_t is the semi-Markov process defined in (1.6) with a Gaussian $\mathcal{N}(0,1)$ sequence $(Y_j)_{j\geq 1}$ and $(\tau_k)_{k\geq 1}$ used in (1.7) taken as $\tau_k \sim \chi^2_3$.

We use the model selection procedure (3.17) with the weights (2.28) in which $k^* = 100 + \sqrt{(}\ln(n))$, $t_i = i/\ln(n)$, $m = [\ln^2(n)]$ and $\delta = (3+\ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathbf{R}} = \frac{1}{p}\sum_{j=1}^{p} \hat{\mathbf{E}}\left(\hat{S}_n(t_j) - S(t_j)\right)^2 , \tag{3.76}$$

where the observation frequency $p = 100001$ and the expectations was taken as an average over $N = 10000$ replications, i.e.

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

We set the relative quadratic risk as

$$\overline{\mathbf{R}_*} = \overline{\mathbf{R}}/\|S\|_p^2 \quad \text{and} \quad \|S\|_p^2 = \frac{1}{p}\sum_{j=0}^{p} S^2(t_j)\,. \tag{3.77}$$

In our case $\|S\|_p^2 = 0.1883601$.

Table 3.1 gives the values for the sample risks (3.76) and (3.77) for different numbers of observations n.

The Figures 3.1–3.4 show the behavior of the regression function and its estimates by the model selection procedure (3.17) depending on the values of observation periods n. The black full line is the regression function (3.75) and the red dotted line is the associated estimator.

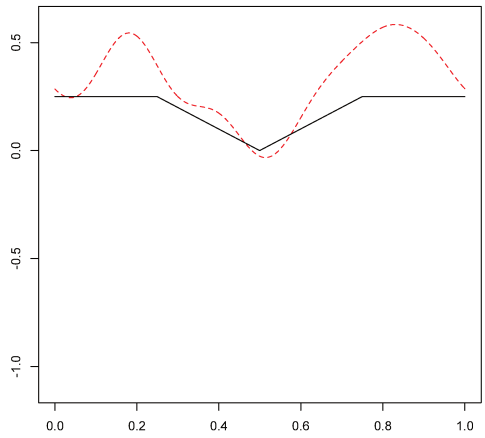

Figure 3.1: Estimator of S for n=20.

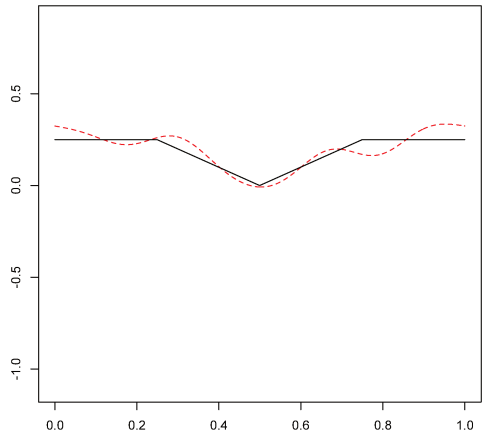

Figure 3.2: Estimator of S for n=100.

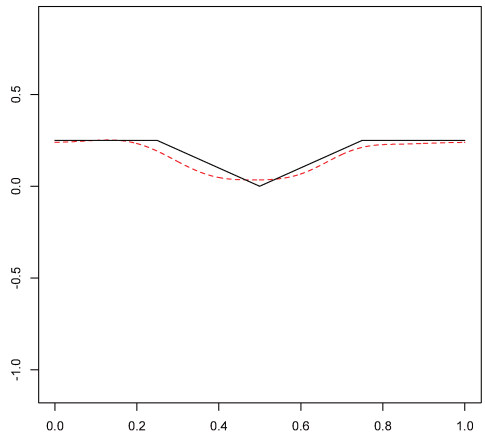

Figure 3.3: Estimator of S for n=200.

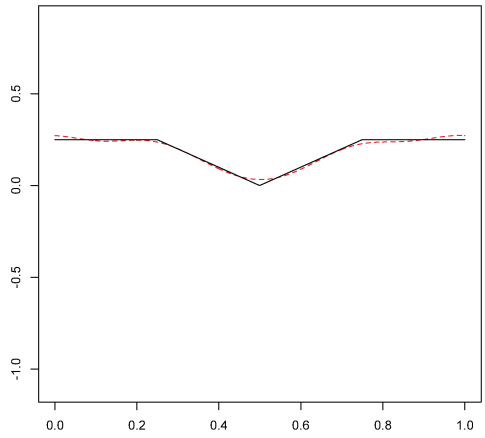

Figure 3.4: Estimator of S for n=1000.

Remark 3.3. From numerical simulations of the procedure (3.17) with various observations numbers n we may conclude that the quality of the proposed procedure is good for practical needs, i.e. for reasonable (non large) number of observations. We can also add that the quality of the estimation improves as the number of observations increases.

Now we give the algorithm of the model selection procedure given in Section 3.2

Algorithm 2 Model selection procedure

Require: n, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

p : satisfying Condition $\mathbf{H}_5$) given in (3.35)

ρ_1, ρ_2, ς^*: satisfying Conditions (2.8) and (2.9)

$k^* \geq 1$, ε: satisfying Condition (2.26)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$m = [1/\varepsilon^2]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

for $k \longleftarrow 1$ *to* n **do**

Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (2.28)

end for

end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varepsilon, \ldots, m\varepsilon\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,p} = \frac{1}{n} \int_0^n \Psi_{k,p}(t) \mathrm{d}\, y_t$.

$\widetilde{\theta}_{k,p} \longleftarrow \widehat{\theta}^2_{k,p} - \frac{1}{n}$.

The observation $(y_t)_{0 \leq t \leq n}$ are given in (2.1) with the noise process (2.2) and $(\Psi_{k,p})_{1 \leq k \leq p}$ is the basis given in (3.5)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,p}, ..., \widehat{\theta}_{n,p})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,p}, ..., \widetilde{\theta}_{n,p})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

for $j \longleftarrow [\varepsilon]$ *to* $[m\varepsilon]$ **do**

$J_n(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l) \widehat{\theta}^2_{j,p} - 2 \sum_{j=1}^n \lambda_{i,j}(l) \widetilde{\theta}_{j,p} + \delta\, P_n(\lambda)$.

where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_n is the penalty term given in (3.15)

end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_n(\lambda), \Lambda = \{\lambda_\alpha\,, \alpha \in \mathcal{A}\}$.

Appendix

Property of the penalty term

Lemma 3.1. *For any $n \geq 1$ and $\lambda \in \Lambda$,*

$$P_n^0(\lambda) \leq \mathcal{R}_Q(\widehat{S}_\lambda, S) + \frac{\mathbf{L}_{1,Q}}{n},$$

where the coefficient $P_n^0(\lambda)$ is defined in (3.21) *and the $\mathbf{L}_{1,Q}$ is defined in* (3.38).

Proof. By the definition of $\mathrm{Err}(\lambda)$ in (3.10) one has

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

In view of Proposition 3.1 we obtain that

$$\mathcal{R}_Q(\widehat{S}_\lambda, S) = \mathbf{E}_Q \,\mathrm{Err}(\lambda) \geq \frac{1}{n}\sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q\, \xi_{j,n}^2 \geq P_n^0(\lambda) - \frac{\mathbf{L}_{1,Q}}{n}\,.$$

Hence we otain Lemma 3.1.

Properties of the Fourier coefficients

Lemma 3.2. *Let f be an absolutely continuous function, $f : [0,1] \to \mathbb{R}$, with $\|\dot{f}\| < \infty$ and g be a simple function, $g : [0,1] \to \mathbb{R}$ of the form $g(t) = \sum_{j=1}^{p} c_j\, \chi_{(t_{j-1},t_j]}(t)$, where c_j are some constants. Then for any $\varepsilon > 0$, the function $\Delta = f - g$ satisfies the following inequalities*

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

Lemma 3.3. *Let the function $S(t)$ in* (3.1) *be absolutly continuous and have an absolutely integrable derivative. Then the coefficients $(\overline{\theta}_{j,p})_{1\leqslant j\leqslant p}$ defined in* (3.6) *satisfy the inequalities*

$$|\overline{\theta}_{1,p}| \leqslant \|S\|_1 \quad \textit{and} \quad \max_{2\leqslant j\leqslant p} j|\overline{\theta}_{j,p}| \leqslant 2\sqrt{2}\|\dot{S}\|_1\,. \tag{3.78}$$

Lemma 3.4. *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 $W_{\mathbf{r}}^1$ satisfy, for any $\widetilde{\varepsilon} > 0$, the following inequality*

$$\sum_{j=N}^{p} \theta_{j,p}^2 \;\leq\; (1+\widetilde{\varepsilon}) \sum_{j\geq N} \theta_j^2 \;+ (1+\widetilde{\varepsilon}^{-1})\,\mathbf{r}\,p^{-2}\,. \tag{3.79}$$

Lemma 3.5. *For any $p \geq 2$ and $\mathbf{r} > 0$, the coefficients $(\theta_{j,p})_{1\leq j\leq p}$ of functions S from the class $W_{\mathbf{r}}^1$ satisfy the following inequality*

$$\max_{1\leq j\leq p} \sup_{S\in W_{\mathbf{r}}^1} \left(|\theta_{j,p} - \theta_j| - 2\pi\sqrt{\mathbf{r}}\, j\, p^{-1} \right) \leq 0\,. \tag{3.80}$$

Lemma 3.6. *For any $p \geq 2$ and $\mathbf{r} > 0$ the correction coefficients $\left(h_{j,p}\right)_{1\leq j\leq p}$ for the functions S from the class $W^2_{\mathbf{r}}$ satisfy the following inequality*

$$\sup_{S\in W^2_{\mathbf{r}}} \sum_{j=1}^{p} h^2_{j,p} \leq 3\mathbf{r}\, p^{-2}\,. \tag{3.81}$$

Lemmas 3.2 – 3.6 are proven in [18].

Chapter 4

Non-parametric estimation for Lévy regression models

4.1 Introduction

Let us consider a regression model in continuous time with the Levy noise

$$\mathrm{d}\, y_t = S(t)\mathrm{d}\, t + \varepsilon\, \mathrm{d}\xi_t\,, \quad 0 \le t \le 1\,, \tag{4.1}$$

where S is an unknown function defined on $\mathbb{R}$ with values in $\mathbb{R}$, $(\xi_t)_{0\le t\le 1}$ is some unobserved noise and $\varepsilon > 0$ is the noise intensity. The problem is to estimate the function S on the basis of observations $(y_t)_{0\le t\le 1}$ when $\varepsilon \to 0$. In this chapter we consider the estimation problem in the adaptive setting, i.e. when the regularity of S is unknown and we assume that the noise $(\xi_t)_{0\le t\le 1}$ is a Lévy process with unknown distribution Q on the Skorokhod space $\mathcal{D}[0,1]$. We know only that this distribution belongs to some distribution family $\mathcal{Q}_\varepsilon$ specified below.

Note that if $(\xi_t)_{0\le t\le 1}$ is the Brownian motion, then we obtain the well known "signal+white noise" model (see, for example, [9], [47], [41]). It should be noted also that the model (4.1) is very popular in the statistical radio-physics. This is the estimation problem of the signal S, observed under the white noise, when the signal/noise ratio goes to infinity.

By making use of the robust estimation approach developed for nonparametric problems in [36, 17, 18] we set the robust risk as

$$\mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) = \sup_{Q\in\mathcal{Q}^*_\varepsilon} \mathcal{R}_Q(\widehat{S}_\varepsilon, S), \tag{4.2}$$

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

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

The goal of this chapter is to develop the sharp model selection method for estimating the unknown signal S. The interest in such statistical procedures can be explained by the fact that they provide adaptive solutions for the nonparametric estimation through the sharp non-asymptotic oracle inequalities which give non-asymptotic upper bound for the quadratic risk including the minimal risk over chosen family of estimators with some coefficient closed to

one (see, for example, [37] for discrete time and [18] for continuous time). The origin of the model selection method goes back to early seventies with the pioneering papers by Akaike [30] and Mallows [23] who suggested to use penalization in a log-likelihood type criterion. Barron, Birgé, Massart [31], Massart [45] and Kneip [40] developed a non-asymptotic model selection method which enables one to derive non-asymptotic oracle inequalities for the non-parametric regression models with Gaussian disturbances. Unfortunately, these methods cannot be applied to the non-Gaussian regression models, since the estimators of the Fourier coefficients in such cases are not independent random variables. For these reasons, in order to estimate the function in non-Gaussian regression models, we use the model selection method developed by [37, 38, 39] for non-Gaussian heteroscedastic regression models in discrete time.

When constructing the sharp model selection procedures, in this chapter, we will use the approach close to that of the papers [14], [15], [16], [18] developed for the estimation of a 1-periodic function in continuous time on a large time interval, i.e.

$$\mathrm{d}x_t = S(t)\mathrm{d}t + \mathrm{d}\eta_t\,, \quad 0 \le t \le n\,.$$

Note that, for any $0 < t < 1$, setting $y_t = n^{-1}\sum_{j=1}^{n}(x_{t+j} - x_j)$, we can represent this model as a model with small parameter of the form (4.1)

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \varepsilon\,\mathrm{d}\xi_t\,,$$

where $\varepsilon = n^{-1/2}$ and $\xi_t = n^{-1/2}\sum_{j=1}^{n}(\eta_{t+j} - \eta_t)$. The main difference between this model and the original one is that the jumps are small, i.e.

$$\Delta\xi_t = \mathrm{O}(n^{-1/2}) = \mathrm{O}(\varepsilon) \quad \text{as} \quad \varepsilon \to 0\,,$$

but we have not such property in the model (4.1). Therefore, unfortunately, we cannot use directly the method developed for the estimation problem on the large time interval to the model (4.1). So, the main goal of this paper is to develop a new sharp model selection method for the estimation problem of the function S as $\varepsilon \to 0$.

As an application of the sharp model selection method in this chapter we consider the problem of the detection of the number of signals for the model (4.1). In many areas of science and technology the problem arise how to select the number degrees of freedom for a statistical model that describes the phenomenons under study most adequately [30]. An important class of such problems is the detection problem of the number of signals with unknown parameters in the noise. For example, in the signal multi-path information transmission there is a detection problem of the number of rays in a multipath channel. This problem is often reduced to the detection of the number of signals. As a result, effective algorithms for the detection of the number of signals can significantly improve the noise immunity in the data transmission over a multipath channel [34, 42, 33, 48, 50, 49, 51]. In all this chapter the problem of the detection of the number of signals are considered only for observation with white noise. In this chapter we consider this problem for non-Gaussian noise with jumps given by (4.3).

4.2 Transformation of the observations

In this chapter the noise process $(\xi_t)_{0\le t\le 1}$ is defined by the following Lévy process

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \widetilde{\mu})_t \,. \tag{4.3}$$

Here, ϱ_1 and ϱ_2 are some constants, $(w_t)_{t\ge 0}$ is a standard Brownian motion, $\mu(\mathrm{d}s\,\mathrm{d}x)$ is the jump measure with the deterministic compensator $\widetilde{\mu}(\mathrm{d}s\,\mathrm{d}x) = \mathrm{d}s\Pi(\mathrm{d}x)$, $\Pi(\cdot)$ is some positive measure on $\mathbb{R}$ (see, for example [10, 6] for details).

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^4) < \infty \,. \tag{4.4}$$

Note that $\Pi(\mathbb{R})$ may be equal to $+\infty$. In the sequel we will denote by Q the distribution of the process $(\xi_t)_{0\le t\le 1}$ in the Skorokhod space $\mathbf{D}[0,1]$ and by $\mathcal{Q}^*_\varepsilon$ we denote all these distributions for which the parameters ϱ_1 and ϱ_2 satisfy the condition

$$\varkappa_Q = \varrho_1^2 + \varrho_2^2 \le \varsigma^*_\varepsilon \,, \tag{4.5}$$

where the bound ς^*_ε is such that for any $\check{\delta} > 0$

$$\liminf_{\varepsilon\to 0} \varepsilon^{-\check{\delta}} \varsigma^*_\varepsilon > 0 \quad \text{and} \quad \lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \varsigma^*_\varepsilon = 0 \,. \tag{4.6}$$

First of all, we need to eliminate the large jumps in the observations (4.1), i.e. we transform this model as

$$\check{y}_t = y_t - \sum_{0\le s\le t} \Delta y_s \, \mathbf{1}_{\{|\Delta y_s| > \overline{a}\}} \,. \tag{4.7}$$

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

$$\mathrm{d}\check{y}_t = S(t)\mathrm{d}t + \varepsilon \mathrm{d}\check{\xi}_t - \varepsilon\, \varrho_2\, \Pi(\overline{h}_\varepsilon)\, \mathrm{d}t \,, \tag{4.8}$$

where

$$\check{\xi}_t = \varrho_1 w_t + \varrho_2\, \check{z}_t \quad \text{and} \quad \check{z}_t = h_\varepsilon * (\mu - \widetilde{\mu})_t \,.$$

The functions $h_\varepsilon(x) = x\mathbf{1}_{\{|x|\le \widetilde{a}_\varepsilon\}}$ and $\overline{h}_\varepsilon(x) = x\mathbf{1}_{\{|x|>\widetilde{a}_\varepsilon\}}$ where the truncated threshold is defined by $\widetilde{a}_\varepsilon = \overline{a}/\varrho_2\varepsilon$.

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

$$\sup_{0\le j\le n} \sup_{0\le t\le 1} |\phi_j(t)| \le \phi^* < \infty \,, \tag{4.9}$$

where $n = n_\varepsilon = [1/\varepsilon^2]$ and $[x]$ denotes the integer part of x. For example, we can take the trigonometric basis defined in (1.15)

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

$$I_t(f) = \int_0^t f(s)\mathrm{d}\xi_s \quad \text{and} \quad \check{I}_t(f) = \int_0^t f(s)\mathrm{d}\check{\xi}_s \tag{4.10}$$

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

$$\mathbf{E}\, I_t^2(f) = \varkappa_Q \, \|f\|_t^2 \quad \text{and} \quad \mathbf{E}\, \check{I}_t^2(f) = \check{\varkappa}_Q \, \|f\|_t^2 \,, \tag{4.11}$$

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

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

.

To estimate the function S we use the following Fourier series

$$S(t) = \sum_{j\geq 1} \theta_j \, \phi_j(t) \,, \tag{4.12}$$

where

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

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

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

and, for $j \geq 2$,

$$\widehat{\theta}_{j,\varepsilon} = \int_0^1 \phi_j(t)\mathrm{d}\,\check{y}_t \,. \tag{4.13}$$

Taking into account here that for any j the integral $\int_0^1 \phi_j(t)\mathrm{d}t = 0$ we obtain from (4.8) that these Fourier coefficients can be represented as

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

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

$$\widehat{\theta}_{j,\varepsilon} = \theta_j + \varepsilon\,\check{\xi}_j \,. \tag{4.14}$$

Now, according to the model selection approach developed in [17] - [18] we need to define for any $u \in \mathbb{R}^n$ the following functions

$$B_{1,\varepsilon}(u) = \sum_{j=1}^n u_j \, \varsigma_j \quad \text{and} \quad B_{2,\varepsilon}(u) = \sum_{j=1}^n u_j \, \widetilde{\xi}_j \,, \tag{4.15}$$

where $\varsigma_j = \mathbf{E}\,(\check{\xi}_j)^2 - \check{\varkappa}_Q$ and $\widetilde{\xi}_j = (\check{\xi}_j)^2 - \mathbf{E}\,(\check{\xi}_j)^2$.

Proposition 4.1. *The following upper bound holds.*

$$\sup_{u\in[0,1]^n} \left|B_{1,\varepsilon}(u)\right| \le \varkappa_Q \,. \tag{4.16}$$

Proof. Taking into account that $\varsigma_1 = \varkappa_Q - \check{\varkappa}_Q \le \varkappa_Q$ and $\varsigma_j = 0$ for $j \ge 2$ we immediately have the upper bound (4.16). □

Now let us recall the Novikov inequalities, [44], also referred to as the Bichteler–Jacod inequalities, see ([32, 43]), providing bounds of the moments of the supremum of purely discontinuous local martingales for $p \ge 2$,

$$\mathbf{E}\sup_{t\le 1}|g*(\mu-\nu)_t|^p \le C_p^* \left(\mathbf{E}\left(|g|^2*\nu_1\right)^{p/2} + \mathbf{E}\left(|g|^p*\nu_1\right)\right) , \tag{4.17}$$

where C_p^* is some positive constant.
Now, for any $u \in \mathbb{R}^n$ we set

$$|u|^2 = \sum_{j=1}^n u_j^2 \quad \text{and} \quad \#(u) = \sum_{j=1}^n \mathbf{1}_{\{u_j\neq 0\}} \,. \tag{4.18}$$

Proposition 4.2. *For any fixed truncated model parameter $a > 0$ and for any vector $u \in \mathbb{R}^n$ with $|u| \le 1$, we have*

$$\mathbf{E}\left|B_{2,\varepsilon}^2(u)\right| \le U_Q + 6\varkappa_Q \left(\frac{a}{\varepsilon}\right)^2 \#(u)\,(\phi^*)^4 , \tag{4.19}$$

where $U_Q = 24\varkappa_Q^2 + 6\varrho_2^4\,\Pi(x^4)$.

Proof. First note that

$$B_{2,\varepsilon}^2(u) \le 2\,\widetilde{\xi}_1^2 + 2B_{2,\varepsilon}^2(u') \,, \tag{4.20}$$

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

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

To study the last term in the right hand side of the inequality (4.20) we set for any function f from $\mathbf{L}_2[0,1]$

$$\widetilde{I}_t(f) = \check{I}_t^2(f) - \mathbf{E}\,\check{I}_t^2(f) \,.$$

Note that for $j \ge 2$ we define the random variables $\widetilde{\xi}_j = \widetilde{I}_1(\phi_j)$. So,

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

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

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

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

$$\mathrm{d}\check{I}_t(f) = \varrho_1 \,\mathrm{d}w_t + \varrho_2 \,\mathrm{d}\check{z}_t\,,$$

we obtain that

$$\mathrm{d}\widetilde{I}_t(f) = 2\varrho_1\, \check{I}_t(f) f(t) \mathrm{d}w_t + 2\varrho_2\, \check{I}_{t-}(f) f(t) \mathrm{d}\check{z}_t + \varrho_2^2\, f^2(t)\, \mathrm{d}\check{m}_t\,.$$

So, setting

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

we obtain that

$$\mathrm{d}D_t = 2\varrho_1\, V_t\, \mathrm{d}w_t + 2\varrho_2\, V_{t-}\, \mathrm{d}\check{z}_t + \varrho_2^2\, \Psi_t\, \mathrm{d}\check{m}_t\,.$$

So, we obtain that

$$D_1^2 \le 12\varrho_1^2 \left(\int_0^1 V_t \,\mathrm{d}w_t \right)^2 + 12\varrho_2^2\, \check{M}_1^2 + 3\varrho_2^4 \left(\int_0^1 \Psi_{t-}\, \mathrm{d}\check{m}_t \right)^2, \tag{4.21}$$

where

$$\check{M}_t = \int_0^t V_{s-}(u)\, \mathrm{d}\check{z}_s.$$

Moreover, taking into account that for any f, g from $\mathbf{L}_2[0,1]$

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

we get

$$2\int_0^1 \mathbf{E}\, V_t^2 \,\mathrm{d}t = 2\sum_{i,j=2}^{n} u_i\, u_j \int_0^1 \phi_i(t)\phi_j(t)\, \mathbf{E}\, \check{I}_t(\phi_i)\, \check{I}_t(\phi_j)\, \mathrm{d}t \;\; = \check{\varkappa}_Q \sum_{i=2}^{n} u_i^2 \left(\int_0^1 \phi_i^2(t)\, \mathrm{d}t \right)^2.$$

Thus,

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

Now, to estimate the second term in the inequality (4.21) note that in view of the inequality (4.17) for any bounded function f and any $0 \le t \le 1$

$$\begin{aligned}
\mathbf{E}\, \check{I}_t^4(f) &\le 8\varrho_1^4\, \mathbf{E} \left(\int_0^t f(s) \mathrm{d}w_s \right)^4 + 8\varrho_2^4\, \mathbf{E} \left(\int_0^t f(s-) \mathrm{d}\check{z}_t \right)^4 \\
&\le 24\varrho_1^4 \int_0^1 f^2(t) \mathrm{d}t + C_4^* \left(\left(\Pi(h_\varepsilon^2) \int_0^1 f^2(t) \mathrm{d}t \right)^2 + \Pi(h_\varepsilon^4) \int_0^1 f^4(t) \mathrm{d}t \right),
\end{aligned}$$

i.e.

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

Now it is easy to see that through the Hölder's inequality the term V_t can be estimated as

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

From here and the inequality (4.17) it follows that

$$\sup_{0\le t\le 1} \mathbf{E}\, \check{M}_t^4 \le C_4^* \, \left(\left(\Pi(h_\varepsilon^2) \right)^2 + \Pi(h_\varepsilon^4) \right) \, \int_0^1 \mathbf{E}\, V_t^4 \, \mathrm{d}t < \infty$$

and, therefore,

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

This implies that

$$\mathbf{E} \int_0^1 \check{M}_{t-} \, \mathrm{d}\check{M}_t = 0 \, .$$

Thus, the Ito's formula implies

$$2\mathbf{E}\, \check{M}_1^2 = \mathbf{E} \sum_{0\le t\le 1} (\Delta \check{M}_t)^2 = 2\Pi(h_\varepsilon^2) \, \int_0^1 \mathbf{E}\, V_t^2 \mathrm{d}t \le \Pi(h_\varepsilon^2) \, \check{\varkappa}_Q \, .$$

In the same way we calculate

$$\begin{aligned} \varrho_2^2 \, \mathbf{E} \, \left(\int_0^1 \Psi_{t-} \mathrm{d}\check{m}_t \right)^2 &= \varrho_2^2 \mathbf{E} \sum_{0\le t\le 1} (\Delta \check{m}_t)^2 \, \Psi_{t-}^2 \\ &= \varrho_2^2 \Pi(h_\varepsilon^4) \, \int_0^1 \Psi_t^2 \mathrm{d}t \le (a/\varepsilon)^2 \, (\phi^*)^4 \, \#(u) \, . \end{aligned}$$

So, we obtain that

$$\mathbf{E}\, D_1^2 \le 6\varrho_1^2 \, \check{\varkappa}_Q + 3\varrho_2^2 \, \Pi(x^2) \left(2\check{\varkappa}_Q + (\phi^*)^4 \right) \le 6 \, \varkappa_Q^2 + 3\varrho_2^2 (\phi^*)^4 \, .$$

Similarly we obtain that

$$\mathbf{E}\widetilde{\xi}_1^2 \le 6\varkappa_Q^2 + 3\varrho_2^4 \, \Pi(x^4) \, .$$

This implies the upper bound (4.19). □

4.3 Model selection

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

$$\widehat{S}_\lambda(x) = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,\varepsilon} \phi_j(x) \, , \tag{4.22}$$

where $n = [1/\varepsilon^2]$, the weights $\lambda = (\lambda(j))_{1\le j\le n}$ belong to some finite set Λ from $[0,1]^n$, $\widehat{\theta}_{j,\varepsilon}$ is defined in (4.13) and ϕ_j in (4.9) . Now we set

$$\check{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = \max_{\lambda\in\Lambda} \sum_{j=1}^{n} \mathbf{1}_{\{\lambda_j>0\}}\,, \tag{4.23}$$

where $\#(\Lambda)$ is the cardinal number of Λ. In the sequel we assume that $|\Lambda|_* \ge 1$. Now we chose the truncating parameter $\overline{a}_\varepsilon$ as

$$\overline{a}_\varepsilon = \frac{\varepsilon}{|\Lambda|_*}\,. \tag{4.24}$$

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

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

which in our case is equal to

$$\mathrm{Err}_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widehat{\theta}_{j,\varepsilon}\theta_j + \sum_{j=1}^{\infty} \theta_j^2. \tag{4.25}$$

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

$$\widetilde{\theta}_{j,\varepsilon} = \widehat{\theta}^2_{j,\varepsilon} - \varepsilon^2\widehat{\varkappa}_\varepsilon\,, \tag{4.26}$$

where $\widehat{\varkappa}_\varepsilon$ is a some estimate for the variance parameter $\check{\varkappa}_Q$ from (4.11). If it is known we set $\widehat{\varkappa}_\varepsilon = \check{\varkappa}_Q$ if not this estimator will be prescribed later.
Finally, to choose the weights we will minimize the following cost function

$$J_\varepsilon(\lambda) = \sum_{j=1}^{n} \lambda^2(j)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^{n} \lambda(j)\widetilde{\theta}_{j,\varepsilon} + \delta\,\widehat{P}_\varepsilon(\lambda)\,, \tag{4.27}$$

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

$$\widehat{P}_\varepsilon(\lambda) = \varepsilon^2\widehat{\varkappa}_\varepsilon|\lambda|^2 \quad \text{and} \quad |\lambda|^2 = \sum_{j=1}^{n} \lambda_j^2\,. \tag{4.28}$$

Note that, if the $\varkappa_Q$ is known, then the penalty is defined as

$$P_\varepsilon(\lambda) = \varepsilon^2\,\check{\varkappa}_Q|\lambda|^2\,. \tag{4.29}$$

We define the model selection procedure as

$$\widehat{S}_* = \widehat{S}_{\hat{\lambda}}\,, \tag{4.30}$$

where

$$\widehat{\lambda} = \mathrm{argmin}_{\lambda\in\Lambda} J_\varepsilon(\lambda). \tag{4.31}$$

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

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

$$\mathcal{A} = \{1,\dots,k^*\}\times\{l_1,\dots,l_m\}\,, \tag{4.32}$$

where

$$l_i = i\,\varpi \quad \text{and} \quad m = [1/\varpi]\,.$$

We assume that both the parameters $k^*\ge 1$ and $0<\varpi<1$ are functions of ε, i.e.$k^* = k^*_\varepsilon$ and $\varpi = \varpi_\varepsilon$, such that

$$\begin{cases} \lim_{\varepsilon\to 0}\, k^*_\varepsilon = +\infty\,, \quad \lim_{\varepsilon\to 0}\, \dfrac{k^*_\varepsilon}{|\ln\varepsilon|} = 0\,, \\ \\ \lim_{\varepsilon\to 0}\varpi_\varepsilon = 0 \quad \text{and} \quad \lim_{\varepsilon\to 0}\,\varepsilon^{-\delta}\varpi_\varepsilon\, = +\infty, \end{cases} \tag{4.33}$$

for any $\delta>0$. One can take, for example, for $0<\varepsilon<1$

$$\varpi_\varepsilon = \frac{1}{|\ln\varepsilon|} \quad \text{and} \quad k^*_\varepsilon = k^*_0 + \sqrt{|\ln\varepsilon|}\,, \tag{4.34}$$

where $k^*_0\ge 0$ is some fixed constant and the threshold ς^*_ε is introduced in (4.5). For each $\alpha=(\beta,l)\in\mathcal{A}$, we introduce the weight sequence

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

where $p=[\varepsilon^{-2}]$,

$$\lambda_\alpha(j) = \mathbf{1}_{\{1\le j<j_*\}} + \Bigl(1-(j/\omega_\alpha)^\beta\Bigr)\,\mathbf{1}_{\{j_*\le j\le\omega_\alpha\}}\,. \tag{4.35}$$

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

$$\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon. \quad \text{and} \quad \mathrm{d}_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\,. \tag{4.36}$$

Now we define the set Λ as

$$\Lambda\ =\ \{\lambda_\alpha\,,\ \alpha\in\mathcal{A}\}\,. \tag{4.37}$$

Note, that these weight coefficients are used in [17, 18] for continuous time regression models to show the asymptotic efficiency.

In the sequel we need to estimate the variance parameter $\varkappa_Q$ from (4.11). To this end we set for any $0<\varepsilon\le 1/\sqrt{3}$

$$\widehat{\varkappa}_\varepsilon = \sum_{j=[1/\varepsilon]+1}^{n}\widehat{\mathbf{T}}^2_{j,\varepsilon}\,, \quad n=[1/\varepsilon^2]\,, \tag{4.38}$$

where $\widehat{\mathbf{T}}_{j,\varepsilon}$ are the estimators of the Fourrier coefficients with respect to the trigonometric basis (1.15) , i.e.

$$\widehat{\mathbf{T}}_{j,\varepsilon} = \int_0^1 \mathrm{Tr}_j(t)\mathrm{d}\check{y}_t\,. \tag{4.39}$$

Remark 4.1. Note that similar sharp oracle inequalities were obtained before in the papers [37] and [17] for the nonparametric regression models in the discrete and continuous time respectively. In this chapter we obtain these inequalities for the model selection procedures based on any arbitrary orthogonal basis function. We use the trigonometric function only to estimate the noise intensity $\check{\varkappa}_Q$.

4.4 Oracle inequality

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

$$\Psi_{Q,\varepsilon} = 8\check{\varkappa}_Q(1+\check{\iota}) + \frac{4U_{1,Q}\check{\iota}}{\check{\varkappa}_Q}\,, \tag{4.40}$$

where

$$U_{1,Q} = 24\check{\varkappa}_Q^2 + 6\varrho_2^4\,\Pi(x^4) + 6\check{\varkappa}_Q\,(\phi^*)^4\,.$$

We start with the sharp oracle inequalities.

Theorem 4.1. *Assume that for the model* (4.1) *the condition* (4.4) *holds. Then, for any* $0<\delta<1/6$*, the estimator of* S *given in* (4.30) *satisfies the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*,S) \le \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda,S) + \varepsilon^2\frac{\Psi_{Q,\varepsilon}+12|\Lambda|_*\,\mathbf{E}_S|\widehat{\varkappa}_\varepsilon-\check{\varkappa}_Q|}{\delta}\,. \tag{4.41}$$

Proof. First, note that we can rewrite the empirical squared error in (4.25) as follows

$$\mathrm{Err}_\varepsilon(\lambda) = J_\varepsilon(\lambda) + 2\sum_{j=1}^n \lambda(j)\check{\theta}_{j,\varepsilon} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda), \tag{4.42}$$

where $\check{\theta}_{j,\varepsilon} = \widetilde{\theta}_{j,\varepsilon} - \theta_j\widehat{\theta}_{j,\varepsilon}$. Now using the definition of $\widetilde{\theta}_{j,\varepsilon}$ in (4.26) we obtain that

$$\check{\theta}_{j,\varepsilon} = \varepsilon\theta_j\xi_j + \varepsilon^2\widetilde{\xi}_{j,\varepsilon} + \varepsilon^2\varsigma_{j,\varepsilon} + \varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,,$$

where $\varsigma_{j,\varepsilon} = \mathbf{E}(\check{\xi}_{j,\varepsilon})^2 - \check{\varkappa}_Q$ and $\widetilde{\xi}_j = (\check{\xi}_j)^2 - \mathbf{E}(\check{\xi}_j)^2$. Setting

$$M_\varepsilon(\lambda) = \varepsilon\sum_{j=1}^n \lambda(j)\theta_j\check{\xi}_j \quad\text{and}\quad L(\lambda) = \sum_{j=1}^n \lambda(j)\,, \tag{4.43}$$

we can rewrite (4.42) as

$$\begin{aligned}\mathrm{Err}_\varepsilon(\lambda) &= J_\varepsilon(\lambda) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\,L(\lambda) + 2M_\varepsilon(\lambda) + 2\varepsilon^2 B_{1,\varepsilon}(\lambda)\\ &\quad + 2\varepsilon\sqrt{P_\varepsilon(\lambda)}\frac{B_{2,\varepsilon}(u_\lambda)}{\sqrt{\check{\varkappa}_Q}} + \|S\|^2 - \delta\widehat{P}_\varepsilon(\lambda),\end{aligned} \tag{4.44}$$

where $u_\lambda = \lambda/|\lambda|$, the exact penalization is defined in (4.29) and the functions $B_{1,\varepsilon}(\cdot)$ and $B_{2,\varepsilon}(\cdot)$ are defined in (4.15). It should be noted that for the truncated parameter (4.24) the bound (4.19) implies

$$\sup_{\lambda\in\Lambda} \mathbf{E}_Q \left| B^2_{2,\varepsilon}(u_\lambda) \right| \le U_Q + 6\check{\varkappa}_Q \left(\frac{a}{\varepsilon}\right)^2 |\Lambda|_* \, (\phi^*)^4 = U_{1,Q}\,, \tag{4.45}$$

where $U_{1,Q} = U_Q + 6\check{\varkappa}_Q \, (\phi^*)^4$.

Let $\lambda_0 = (\lambda_0(j))_{1\le j\le n}$ be a fixed sequence in Λ and $\widehat{\lambda}$ be as in (4.31). Substituting λ_0 and $\widehat{\lambda}$ in Equation (4.44), we obtain

$$\begin{aligned}
\mathrm{Err}_\varepsilon(\widehat{\lambda}) - \mathrm{Err}_\varepsilon(\lambda_0) = {} & J(\widehat{\lambda}) - J(\lambda_0) + 2\varepsilon^2(\check{\varkappa}_Q - \widehat{\varkappa}_\varepsilon)\, L(\varpi) \\
& + 2\varepsilon^2 B_{1,\varepsilon}(\varpi) + 2M_\varepsilon(\varpi) \\
& + 2\varepsilon\sqrt{P_\varepsilon(\widehat{\lambda})}\frac{B_{2,\varepsilon}(\widehat{u})}{\sqrt{\check{\varkappa}_Q}} - 2\varepsilon\sqrt{P_\varepsilon(\lambda_0)}\frac{B_{2,\varepsilon}(u_0)}{\sqrt{\check{\varkappa}_Q}} \\
& - \delta\widehat{P}_\varepsilon(\widehat{\lambda}) + \delta\widehat{P}_\varepsilon(\lambda_0),
\end{aligned} \tag{4.46}$$

where $\varpi = \widehat{\lambda} - \lambda_0$, $\widehat{u} = u_{\widehat{\lambda}}$ and $u_0 = u_{\lambda_0}$. Note that by (4.23)

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

The inequality

$$2|ab| \le \delta a^2 + \delta^{-1} b^2 \tag{4.47}$$

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

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

From the bound (4.16) it follows that for $0 < \delta < 1$

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

where $B^*_{2,\varepsilon} = \sup_{\lambda\in\Lambda} B^2_{2,n}(u_\lambda)$. It should be noted that through (4.45) we can estimate this term as

$$\mathbf{E}_Q\, B^*_{2,\varepsilon} \le \sum_{\lambda\in\Lambda} \mathbf{E}_Q\, B^2_{2,\varepsilon}(u_\lambda) \le \imath U_{1,Q}\,. \tag{4.48}$$

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

$$\begin{aligned}
\mathrm{Err}_\varepsilon(\widehat{\lambda}) \le {} & \mathrm{Err}_\varepsilon(\lambda_0) + 2M_\varepsilon(\varpi) + 2\varepsilon^2\frac{B^*_{2,\varepsilon}}{\delta\check{\varkappa}_Q} + 2\varepsilon^2\,\check{\varkappa}_Q \\
& + \frac{6\varepsilon^2|\Lambda|_*}{n}|\widehat{\varkappa} - \check{\varkappa}_Q| + 2\delta P_\varepsilon(\lambda_0).
\end{aligned} \tag{4.49}$$

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

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

Moreover, note that

$$M_\varepsilon^2(\upsilon) \le 2\upsilon^2(1)\,\xi_1^2 + 2\,\check{I}_1(\Phi)\,,$$

where $\Phi(t) = \sum_{j=2}^{n} \upsilon(j)\theta_j\phi_j(t)$. Therefore, thanks to (4.11) we obtain that for any non-random $\upsilon \in \mathbb{R}^n$

$$\mathbf{E}M_\varepsilon^2(\upsilon) \le 2\check{\varkappa}_Q\varepsilon^2\sum_{j=1}^{n}\upsilon^2(j)\theta_j^2 = 2\check{\varkappa}_Q\varepsilon^2||S_\upsilon||^2\,. \tag{4.50}$$

To estimate this function for a random vector we set

$$M_\varepsilon^* = \sup_{\upsilon\in\Lambda_1}\frac{M^2(\upsilon)}{\varepsilon^2||S_\upsilon||^2} \quad \text{and} \quad \Lambda_1 = \Lambda - \lambda_0\,.$$

So, through the inequality (4.47)

$$2|M_\varepsilon(\upsilon)| \le \delta||S_\upsilon||^2 + \varepsilon^2\frac{M_\varepsilon^*}{\delta}\,. \tag{4.51}$$

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

$$\mathbf{E}\,M_\varepsilon^* \le \sum_{\upsilon\in\Lambda_1}\frac{\mathbf{E}\,M_\varepsilon^2(\upsilon)}{\varepsilon^2||S_\upsilon||^2} \le 2\sum_{\upsilon\in\Lambda_1}\check{\varkappa}_Q = 2\check{\varkappa}_Q\,\check{\iota}\,, \tag{4.52}$$

where $\nu = \#(\Lambda)$. Moreover, note that, for any $\upsilon \in \Lambda_1$,

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

where $\upsilon^2 = (\upsilon^2(j))_{1\le j\le n}$. Taking into account that, for any $x \in \Lambda_1$, the components $|\upsilon(j)| \le 1$, we can estimate the last term as in (4.50), i.e.

$$\mathbf{E}\,M_\varepsilon^2(\upsilon^2) \le 2\varepsilon^2\check{\varkappa}_Q\,||S_\upsilon||^2\,.$$

Similarly, setting

$$M_{1,\varepsilon}^* = \sup_{\upsilon\varepsilon\Lambda_1}\frac{M_\varepsilon^2(\upsilon^2)}{\varepsilon^2||S_\upsilon||^2},$$

we obtain

$$\mathbf{E}_Q\,M_{1,\varepsilon}^* \le 2\check{\varkappa}a_Q\,\check{\iota}\,. \tag{4.53}$$

In the same way we find that

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

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

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

So, from (4.51) we get

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

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

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

Using this bound in (4.49) we obtain

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

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

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

Using here the bounds (4.48), (4.52), (4.53) we obtain that

$$\mathcal{R}(\widehat{S}_*, S) \le \frac{1+\delta}{1-3\delta}\mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{8\varepsilon^2\check{\varkappa}_Q\check{\iota}}{\delta} + \frac{4\varepsilon^2 U_{1,Q}\check{\iota}}{\delta\check{\varkappa}_Q}$$
$$+ 4\varepsilon^2\,\check{\varkappa}_Q + 12\varepsilon^2\,|\Lambda|_*\mathbf{E}_Q\,|\widehat{\varkappa} - \check{\varkappa}_Q| + \frac{2\delta}{1-3\delta}\,P_\varepsilon(\lambda_0)\,.$$

Now, Lemma 4.1 implies directly the inequality (4.1). Hence we get the desired result. □

Corollary 4.1. *Assume that for the model* (4.1) *the condition* (4.4) *holds. If the variance parameter* $\check{\varkappa}_Q$ *is known, then for any* $0 < \delta < 1/6$*, the estimator of* S *given in* (4.30)*, with the truncated parameter* $\overline{a} = \varepsilon/\sqrt{|\Lambda|_*}$ *satisfying the following oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta}\min_{\lambda\in\Lambda}\mathcal{R}_Q(\widehat{S}_\lambda, S) + \varepsilon^2\frac{\Psi_{Q,\varepsilon}}{\delta}\,. \tag{4.54}$$

We need to study the estimate (4.38).

Proposition 4.3. *Assume that in the model* (4.1) *the unknown function* S *is continuously differentiable. Then, for any* $0 < \varepsilon \le 1/\sqrt{3}$

$$\mathbf{E}_Q \left| \widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q \right| \le \varepsilon \Upsilon_Q(S) + \frac{\sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} , \tag{4.55}$$

where $\Upsilon_Q(S) = 4(\|\dot{S}\| + 1)^2 \left(1 + \sqrt{\check{\varkappa}_Q} + 2\check{\varkappa}_Q + \sqrt{U_Q}\right)$.

Proof. We use here the same method as in [14]. First, note that from the definitions (4.14) and (4.39) we obtain

$$\widehat{\mathbf{T}}_{j,\varepsilon} = \mathbf{T}_j + \varepsilon \check{\eta}_j , \tag{4.56}$$

where

$$\mathbf{T}_j = \int_0^1 S(t)\, Tr_j(t) \mathrm{d}t \quad \text{and} \quad \check{\eta}_j = \int_0^1 \mathrm{Tr}_j(t)\, \mathrm{d}\check{\xi}_t .$$

So, we have

$$\widehat{\varkappa}_\varepsilon = \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j^2 + 2\check{M}_\varepsilon + \varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} (\check{\eta}_j)^2 , \tag{4.57}$$

where $\check{M}_\varepsilon = \varepsilon \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j \check{\eta}_j$. Note that for continiously differentiable functions (see, for example, Lemma A.6 in [14]) the Fourrier coefficients $(\mathbf{T}_j)$ for any $n \ge 1$ satisfy the following inequality

$$\sum_{j=[1/\varepsilon]+1}^{\infty} \mathbf{T}_j^2 \le 4\varepsilon \left(\int_0^1 |\dot{S}(t)| \mathrm{d}t \right)^2 \le 4\varepsilon \|\dot{S}\|^2 . \tag{4.58}$$

The term $\check{M}_\varepsilon$ can be estimated in the same way as in (4.50), i.e.

$$\mathbf{E}_Q \check{M}_\varepsilon^2 \le \check{\varkappa}_Q \varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} \mathbf{T}_j^2 \le 4\varepsilon^3 \check{\varkappa}_Q \|\dot{S}\|^2 .$$

Moreover, taking into account that for $j \ge 2$ the expectation $\mathbf{E}\,(\check{\eta}_j)^2 = \check{\varkappa}_Q$ we can represent the last term in (4.57) as

$$\varepsilon^2 \sum_{j=[1/\varepsilon]+1}^{n} (\check{\eta}_j)^2 = \varepsilon^2\, \check{\varkappa}_Q (n - [1/\varepsilon]) + \varepsilon\, B_{2,\varepsilon}(x') ,$$

where the function $B_{2,\varepsilon}(x')$ is defined in (4.15) and $x'_j = \varepsilon \mathbf{1}_{\{1/\varepsilon < j \le 1/\varepsilon^2\}}$. We remind that $n = [1/\varepsilon^2]$. Therefore, in view of Proposition 4.2 we obtain

$$\mathbf{E}_Q \left| \varepsilon^2 \sum_{j=[\sqrt{1/\varepsilon}]+1}^{n} \check{\eta}_j^2 - \check{\varkappa}_Q \right| \le 2\varepsilon\, \check{\varkappa}_Q + \varepsilon \sqrt{U_Q} + \frac{\sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} .$$

So, we obtain the bound (4.55). □

It is clear that in the case when $|\Lambda|_* \le 1/\varepsilon$ we obtain that

$$\mathbf{E}_Q \, |\widehat{\varkappa}_\varepsilon - \check{\varkappa}_Q| \le \frac{\Upsilon_Q(S) + \sqrt{6\check{\varkappa}_Q}}{|\Lambda|_*} \, . \tag{4.59}$$

Now using this proposition we can obtain the following inequality.

Theorem 4.2. *Assume that for the model* (4.1) *the condition* (4.4) *holds and the unknown function S is continuously differentiable. Then the procedure* (4.30) *with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfies the following oracle inequality*

$$\begin{aligned} \mathcal{R}_Q(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}_Q(\widehat{S}_\lambda, S) \\ + \varepsilon^2 \frac{\Psi_{Q,\varepsilon} + (\|\dot{S}\| + 1)^2 g_{1,Q} + g_{2,Q}}{\delta} \, , \end{aligned} \tag{4.60}$$

where

$$g_{1,Q} = 48 \left(1 + \sqrt{\check{\varkappa}_Q} + 2\check{\varkappa}_Q + \sqrt{U_Q}\right) \quad \textit{and} \quad g_{2,Q} = 12\sqrt{6\check{\varkappa}_Q} \, .$$

Now we study the robust risk defined in (4.2) for the procedure (4.30).

We assume also that the upper bound for the basis functions in (4.9) may be dependent on $n \ge 1$, i.e. $\phi_* = \phi_*(n)$, such that for any $\check{\epsilon} > 0$

$$\lim_{n\to\infty} \frac{\phi_*(n)}{n^{\check{\epsilon}}} = 0 \, . \tag{4.61}$$

Theorem 4.3. *Assume that for the model* (4.1) *Condition* (4.4) *holds and the unknown function S is continuously differentiable. Then the robust risk of the procedure* (4.30) *with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfy the following oracle inequality*

$$\mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le \frac{1+3\delta}{1-3\delta} \min_{\lambda\in\Lambda} \mathcal{R}^*_\varepsilon(\widehat{S}_\lambda, S) + \varepsilon^2 \, \frac{\mathbf{U}^*_\varepsilon(S)}{\delta} \, , \tag{4.62}$$

where the term $\mathbf{U}^*_\varepsilon(S) > 0$ is such that under the conditions (4.61) *and* (4.33) *for any $r > 0$ and $\check{\delta} > 0$*

$$\lim_{\varepsilon\to 0} \varepsilon^{\check{\delta}} \sup_{\|\dot{S}\|\le r} \mathbf{U}^*_\varepsilon(S) \, = 0 \, . \tag{4.63}$$

4.5 Adaptive robust efficiency

Now we study the asymptotically efficiency properties for the procedure (4.30) with the coefficients (4.35) with respect to the robust risks (4.12) defined by the distribution family (4.5) – (4.6). To this end we assume that the unknown function (4.12) belongs to the following ellipsoid in $\mathbf{L}_2$,

$$W^k_{\mathbf{r}} = \{S \in \mathbf{L}_2[0,1] \, : \, \sum_{j=1}^{\infty} a_j \, \theta_j^2 \, \le \mathbf{r}\}, \tag{4.64}$$

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

It is easy to see that in the case when the functions $(\phi_j)_{j\ge1}$ are trigonometric (1.15), then this set coincides with the Sobolev ball

$$W^k_{\mathbf{r}} = \{f \in \mathbf{C}^k_{per}[0,1] \, : \, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le \mathbf{r}\}\,, \tag{4.65}$$

where $\mathbf{r} > 0$ and $k \ge 1$ are some parameters, $\mathbf{C}^k_{per}[0,1]$ 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 $0 \le i \le k$. Similarly to [17, 18] we will show here that the asymptotic sharp lower bound for the robust risk (4.12) is given by

$$l_*(\mathbf{r}) = \, ((2k+1)\mathbf{r})^{1/(2k+1)} \, \left(\frac{k}{(k+1)\pi}\right)^{2k/(2k+1)} . \tag{4.66}$$

Note that this is the well-known Pinsker's constant obtained for the non-adaptive filtration problem in "signal + small white noise" model (see, for example, [47]).

Let $\mathcal{S}_\varepsilon$ be the set of all estimators $\widehat{S}_\varepsilon$ measurable with respect to the σ-algebra $\sigma\{y_t \, , \, 0 \le t \le 1\}$ generated by the process (4.1).

Theorem 4.4. *For the distribution family* (4.5) – (4.6)*, the robust risks admit the following lower bound*

$$\liminf_{\varepsilon\to0} \, \upsilon_\varepsilon^{2k/(2k+1)} \inf_{\widehat{S}_\varepsilon\in\mathcal{S}_\varepsilon} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) \ge \, l_*(\mathbf{r})\,, \tag{4.67}$$

where $\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon$.

We set the parameter δ in (4.27) as a function of ε, i.e. $\delta = \delta_\varepsilon$ is such that

$$\lim_{\varepsilon\to0} \, \delta_\varepsilon = 0 \quad \text{and} \quad \lim_{\varepsilon\to0} \, \varepsilon^{-\check{\delta}} \, \delta_\varepsilon = 0 \tag{4.68}$$

for any $\check{\delta} > 0$. For example, we can take $\delta_\varepsilon = (6 + |\ln\varepsilon|)^{-1}$.

Theorem 4.5. *Assume that Conditions* (4.33) *hold. Then the model selection procedure admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to0} \, \upsilon_\varepsilon^{2k/(2k+1)} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_*, S) \le l_*(\mathbf{r})\,. \tag{4.69}$$

Theorem 4.4 and Theorem 4.5 imply the following result

Corollary 4.2. *Under the conditions of Theorem 4.5, we have*

$$\lim_{\varepsilon\to 0} \upsilon_\varepsilon^{2k/(2k+1)} \inf_{\widehat{S}_\varepsilon\in\mathcal{S}_\varepsilon} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) = l_*(\mathbf{r})\,. \tag{4.70}$$

Remark 4.2. It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball W^k_r is $\varepsilon^{-4k/(2k+1)}$ (see, for example, [47]). We see here that the efficient robust rate is $\upsilon_\varepsilon^{2k/(2k+1)}$, i.e. if the distribution upper bound $\varsigma^*_\varepsilon \to 0$ as $n\to\infty$ we obtain a faster rate with respect to $\varepsilon^{-4k/(2k+1)}$, and if $\varsigma^*_\varepsilon \to \infty$ as $\varepsilon\to 0$ we obtain a slower rate. In the case when ς^*_ε is constant the robust rate is the same as the classical non robust convergence rate.

4.5.1 Lower bound

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

$$\sup_{S\in W^k_r} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon, S) \geq \sup_{S\in W^k_r} \mathcal{R}_Q(\widehat{S}_\varepsilon, S)\,. \tag{4.71}$$

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

$$d = d_\varepsilon = \left[\frac{k+1}{k}\upsilon_\varepsilon^{1/(2k+1)}\, l_*(\mathbf{r}_0)\right] \quad \text{and} \quad \mathbf{r}_0 = (1-\check{\gamma})\mathbf{r}\,. \tag{4.72}$$

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

$$S_z(x) = \sum_{j=1}^{d} z_j\,\phi_j(x)\,. \tag{4.73}$$

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

$$\kappa = (\kappa_j)_{1\leq j\leq d} \quad \text{and} \quad \kappa_j = s_j\,\eta_j\,, \tag{4.74}$$

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

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

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

$$\widetilde{\mathcal{R}}_Q(\widehat{S}) = \int_{\mathbb{R}^d} \mathcal{R}_Q(\widehat{S}, S_z)\,\mu_\kappa(\mathrm{d}z)\,. \tag{4.75}$$

Furthermore, for any function f, we denote by $\mathbf{p}(f)$ its projection in $\mathbf{L}_2[0,1]$ onto $W^k_{\mathbf{r}}$, i.e.

$$\|f-\mathbf{p}(f)\| = \inf_{h\in W^k_{\mathbf{r}}} \|f-h\|\,.$$

Since $W^k_{\mathbf{r}}$ is a convex and closed set in $\mathbf{L}_2[0,1]$, this projector exists and is unique for any function $f \in \mathbf{L}_2[0,1]$ and, moreover,

$$\|f-h\|^2 \geq \|\mathbf{p}(f)-h\|^2 \quad \text{for any} \quad h \in W^k_{\mathbf{r}}\,.$$

So, setting $\widehat{\mathbf{p}} = \mathbf{p}(\widehat{S})$, we obtain that

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

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

$$\sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}_Q(\widehat{S},S) \;\geq\; \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon \tag{4.76}$$

and

$$\Delta_\varepsilon = \int_{\{z\in\mathbb{R}^d\,:\,S_z\notin W^k_{\mathbf{r}}\}} (\mathbf{r}+\|S_z\|^2)\,\mu_\kappa(\mathrm{d}z)\,.$$

Therefore, in view of (4.71)

$$\sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon(\widehat{S}_\varepsilon,S) \geq \sup_{Q\in\mathcal{Q}^*_\varepsilon} \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) - 2\,\Delta_\varepsilon\,. \tag{4.77}$$

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

$$\lim_{\varepsilon\to 0} \varepsilon^{-\check{\delta}}\,\Delta_\varepsilon = 0\,. \tag{4.78}$$

Now it is easy to see that

$$\|\widehat{\mathbf{p}}-S_z\|^2 \geq \sum_{j=1}^d (\widehat{z}_j - z_j)^2\,,$$

where $\widehat{z}_j = \int_0^1 \widehat{\mathbf{p}}(t)\,\phi_j(t)\mathrm{d}t$. So, in view of Lemma 4.2 and reminding that $\upsilon_\varepsilon = \varepsilon^{-2}/\varsigma^*_\varepsilon$ we obtain

$$\begin{aligned}\sup_{Q\in\mathcal{Q}^*_\varepsilon} \widetilde{\mathcal{R}}_Q(\widehat{\mathbf{p}}) \;&\geq\; \sup_{0<\varrho_1^2\leq\varsigma^*_\varepsilon} \sum_{j=1}^d \frac{1}{\varepsilon^{-2}\,\varrho_1^{-2}+\upsilon_\varepsilon\,(s^*_j)^{-1}} \\ &= \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d \frac{s^*_j}{s^*_j+1} = \frac{1}{\upsilon_\varepsilon}\sum_{j=1}^d\left(1-\frac{j^k}{d^k_\varepsilon}\right).\end{aligned}$$

Therefore, using now Definition (4.72), Inequality (4.77) and the limit (4.78), we obtain that

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

Taking here the limit as $\check{\gamma}\to 0$ implies the desired result . □

4.5.2 Upper bound

Known smoothness

First we suppose that the parameters $k \geq 1$, $\mathbf{r} > 0$ in (4.65) and ς_ε^* in (4.5) are known. Let the family of admissible weighted least square estimates $(\widehat{S}_\lambda)_{\lambda\in\Lambda}$ be given by (4.37). Consider the pair

$$\check{\alpha} = (k, \check{r}) \quad \text{and} \quad \check{r} = \varpi\,[\mathbf{r}/\varpi]\,,$$

where $\check{\varsigma}_\varepsilon = \varpi\varsigma_\varepsilon^*$ and ϖ satisfy the conditions in (4.33). Denote the corresponding estimate as

$$\check{S} = \widehat{S}_{\check{\lambda}} \quad \text{and} \quad \check{\lambda} = \lambda_{\check{\alpha}}\,. \tag{4.79}$$

Note that for sufficiently small ε the pair $\check{\alpha}$ belongs to the set (4.32).

Theorem 4.6. *The estimator $\check{S}$ admits the following asymptotic upper bound*

$$\limsup_{\varepsilon\to0}\, \upsilon_\varepsilon^{2k/(2k+1)} \sup_{S\in W^k_{\mathbf{r}}} \mathcal{R}^*_\varepsilon\left(\check{S}, S\right) \leq l_*(\mathbf{r})\,. \tag{4.80}$$

Proof. Substituting (4.14) and taking into account the definition (4.79) one gets

$$\|\check{S} - S\|^2 = \sum_{j=1}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 - 2\check{M}_\varepsilon + \varepsilon^2 \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,\check{\xi}_j^2\,,$$

where $\check{M}_\varepsilon = \varepsilon \sum_{j=1}^{\infty} (1-\check{\lambda}(j))\,\check{\lambda}(j)\,\theta_j\,\check{\xi}_j$. Note now that for any $Q \in \mathcal{Q}^*_\varepsilon$ the expectation $\mathbf{E}_{Q,S}\,\check{M}_\varepsilon = 0$ and, in view of the upper bound (4.12),

$$\sup_{Q\in\mathcal{Q}^*_\varepsilon} \mathbf{E}_{Q,S} \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,\check{\xi}_j^2 \leq \varsigma_\varepsilon^* \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,.$$

Therefore,

$$\mathcal{R}^*_\varepsilon(\check{S}, S) \leq \sum_{j=\check{\iota}}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 + \frac{1}{\upsilon_\varepsilon} \sum_{j=1}^{\infty} \check{\lambda}^2(j)\,, \tag{4.81}$$

where $\check{j}_* = j_*(\check{\alpha})$. Setting

$$u_\varepsilon = (\upsilon_\varepsilon)^{2k/(2k+1)} \sup_{j\geq\check{j}_*} (1-\check{\lambda}(j))^2/a_j\,,$$

we obtain that for each $S \in W^k_{\mathbf{r}}$

$$\Upsilon_{1,\varepsilon}(S) = (\upsilon_\varepsilon)^{2k/(2k+1)} \sum_{j=\check{j}_*}^{\infty} (1-\check{\lambda}(j))^2\,\theta_j^2 \leq u_\varepsilon \sum_{j=\check{j}_*}^{\infty} a_j\,\theta_j^2 \leq u_\varepsilon\,\mathbf{r}\,.$$

Tazking into account that $\check{r} \to \mathbf{r}$, we obtain that

$$\limsup_{\varepsilon\to0} \sup_{S\in W^k_{\mathbf{r}}} \Upsilon_{1,\varepsilon}(S) \leq \frac{\mathbf{r}^{1/(2k+1)}}{\pi^{2k}(\mathrm{d}_k)^{2k/(2k+1)}} := \Upsilon_1^*\,,$$

where the coefficient τ_k is given in (4.35). To estimate the last term in the right hand of (4.81), we set

$$\Upsilon_{2,\varepsilon} = \frac{1}{(\upsilon_\varepsilon)^{1/(2k+1)}} \sum_{j=1}^{+\infty} \check{\lambda}^2(j)\,.$$

It is easy to check that

$$\limsup_{\varepsilon\to 0}\, \Upsilon_{2,\varepsilon} \le \frac{2(\mathbf{r}\mathrm{d}_k)^{1/(2k+1)}\, k^2}{(k+1)(2k+1)} := \Upsilon_2^*\,.$$

Therefore, taking into account that, by the definition of the Pinsker's constant in (4.66), $\Upsilon_1^* + \Upsilon_2^* = l_*(\mathbf{r})$, we arrive at the inequality

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

Hence we obtain the desired result. □

Unknown smoothness

Combining Theorem 4.6 and Theorem 4.3 yields Theorem 4.5. □

4.6 Detection of the number of signals

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

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

where $(n_t)_{t\ge 0}$ is some noise, usually this "white noise". The energetic parameters $(\theta_j)_{j\ge 1}$ and the number of the signals q are unknown and the signals $(\phi_j)_{j\ge 1}$ are known orthonormal functions, i.e.

$$\int_0^1 \phi_i(t)\,\phi_j(t)\,\mathrm{d}t = \mathbf{1}_{\{i\ne j\}}\,.$$

The problem is to estimate q, when the signal/noise ratio goes to infinity. To describe this problem in the framework of the mathematical model we use the following stochastic differential equation

$$\mathrm{d}y_t = \left(\sum_{j=1}^{q} \theta_j \phi_j(t)\right) \mathrm{d}t + \varepsilon \mathrm{d}w_t\,, \tag{4.82}$$

where $(w_t)_{t\geq 0}$ is the standard Brownian motion and the parameter ε goes to zero. This means tha the signal/noise ratio goes to infinity. The logarithm of the likelihood ratio for the model (4.82) can be represented as

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

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

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

Therefore, the maximum likelihood estimate for $\widehat{q}=q^*$. So, if $q^*=\infty$ we obtain that $\widehat{q}=\infty$. So, this estimator does not work. For these reasons we propose to study the estimation problem for q for the process (4.82) in the nonparametric setting and to apply the model selection procedure (4.30). To this end we consider the model (4.1) with the unknown function S defined as

$$S(t)=\sum_{j=1}^{q}\theta_j\,\phi_j(t)\,. \tag{4.83}$$

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

$$\widehat{S}_d(x)=\sum_{j=1}^{d}\widehat{\theta}_{j,\varepsilon}\phi_j(x)\,. \tag{4.84}$$

This estimate can be obtained from (4.22) with the weights $\lambda_d(j)=\chi\{j\leq d\}$. The number of estimators ι is some function of ε, i.e. $\iota=\iota_\varepsilon$, such that

$$\lim_{\varepsilon\to 0}\iota_\varepsilon=\infty\quad\text{and}\quad\lim_{\varepsilon\to 0}\varepsilon^{\check{\delta}}\iota_\varepsilon=0 \tag{4.85}$$

for any $\check{\delta}>0$. As a risk for the signals number we use

$$\mathbf{D}_\varepsilon(d,q)=\mathcal{R}^*_\varepsilon(\widehat{S}_d,S)\,, \tag{4.86}$$

where the risk $\mathcal{R}^*_\varepsilon(\widehat{S},S)$ is defined in (4.2) and d is some integer number (maybe random) from the set $\{1,\ldots,\iota\}$. In this case the cost function (4.27) has the following form.

$$J_\varepsilon(d)=\sum_{j=1}^{d}\widehat{\theta}^2_{j,\varepsilon}-2\sum_{j=1}^{d}\widetilde{\theta}_{j,\varepsilon}+\delta\,\widehat{P}_\varepsilon(\lambda)\,. \tag{4.87}$$

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

$$\widehat{q}_\varepsilon=\mathrm{argmin}_{1\leq d\leq\iota}J_\varepsilon(d)\,. \tag{4.88}$$

Note that Theorem 4.3 implies that the robust risks of the procedure (4.30) with $|\Lambda|_* \le 1/\varepsilon$, for any $0 < \delta < 1/6$, satisfy the following oracle inequality

$$\mathbf{D}_\varepsilon(\widehat{q}_\varepsilon\,,\,q) \le \frac{1+3\delta}{1-3\delta}\min_{1\le d\le \iota}\mathbf{D}_\varepsilon(d,q) + \varepsilon^2\,\frac{\mathbf{U}^*_\varepsilon(S)}{\delta}\,, \tag{4.89}$$

where the last term satisfies the property (3.37).

4.7 Simulation

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure (4.30). In (4.1) we chose

$$S(t) = \sum_{j=1}^{10}\frac{j}{j+1}\,\phi_j(t)\,, \tag{4.90}$$

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

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

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

We calculated the empirical quadratic risk defined as

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

and the relative quadratic risk

$$\overline{\mathbf{R}}_* = \overline{\mathbf{R}}/\|S\|_p^2\,.$$

The expectation is taken as an average over $N = 10000$ replications, i.e.

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

We used the cost function with

$$\delta = \frac{1}{(3+|\ln\varepsilon|)^2}\,.$$

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

Table 4.1: Empirical risks

In the following graphics the dashed line is the model selection procedure (4.30), the continuous line is the function (4.90) and the bold line is the corresponding observations (4.1).

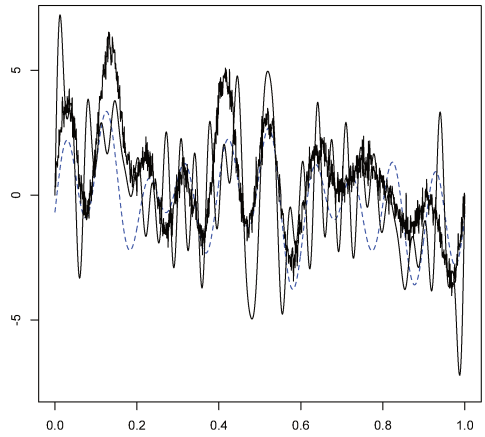

Figure 4.1: Estimator of S for $\varepsilon = 1/\sqrt{20}$

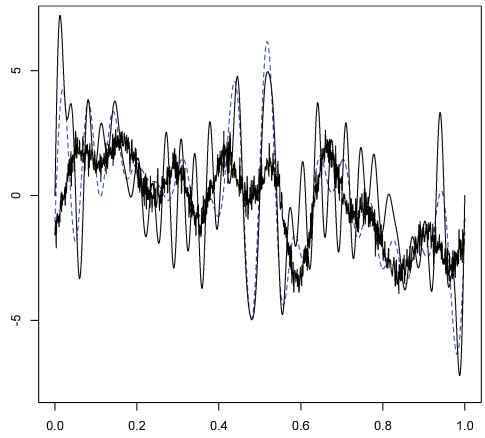

Figure 4.2: Estimator of S for $\varepsilon = 1/\sqrt{100}$

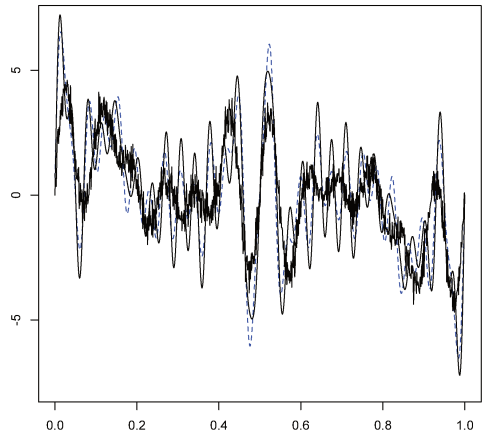

Figure 4.3: Estimator of S for $\varepsilon = 1/\sqrt{200}$

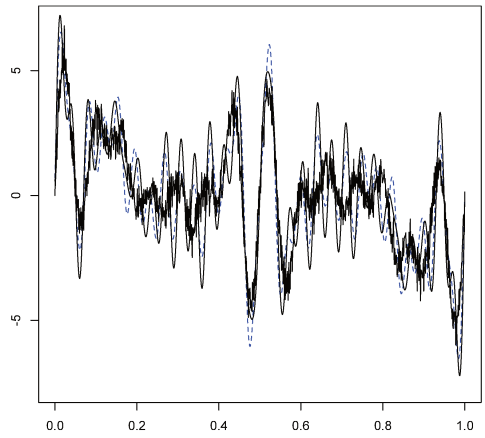

Figure 4.4: Estimator of S for $\varepsilon = 1/\sqrt{1000}$

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

Table 4.2: Estimation of the number of signals

To estimate the number of signals q we use two procedures. The first $\widehat{q}_1$ is (4.89) with $\nu = [\ln \varepsilon^{-2}]$. The second $\widehat{q}_2$ is defined through the shrinkage approach for the model selection procedure (4.90),

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

Remark 4.3. It should be noted that the LSE procedure (4.89) is more appropriate than the shrinkage method for such a number detection problem.

Now we give the algorithm of the model selection procedure given in Section 4.3

Algorithm 3 Model selection procedure

Require: $\varepsilon > 0$, $0 \leq \breve{\varrho} \leq 1$ and $\delta > 0$

ρ_1, ρ_2, ς_ε^*: satisfying Conditions (4.5) and (4.6)

$k^* \geq 1$, ϖ: satisfying Condition (4.33)

Output: The optimal weight vector $\hat{\lambda}$

{**Step 1**} Computation of the weights

$n = [1/\varepsilon^2]$, $m = [1/\varpi]$

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varpi]$ *to* $[m\varpi]$ **do**

    for $k \longleftarrow 1$ *to* n **do**

      Compute the wheight coefficients $\lambda_{i,j}(k)$ using the formula (4.35)

    end for

  end for

end for

return: the vectors $\lambda = (\lambda_\alpha(1), ..., \lambda_\alpha(n)), \alpha \in \mathcal{A} = \{1, \ldots, k^*\} \times \{\varpi, \ldots, m\varpi\}$

{**Step 2**} Computation of the Fourrier coefficients

for $k \longleftarrow 1$ *to* n **do**

$\widehat{\theta}_{k,\varepsilon} = \int_0^1 \phi_k(t) \mathrm{d}\, \breve{y}_t$.

$\widetilde{\theta}_{k,\varepsilon} \longleftarrow \widehat{\theta}^2_{k,\varepsilon} - \varepsilon^2$.

The observation $(\breve{y}_t)_{0 \leq t \leq n}$ are given in (4.7) and $(\phi_k)_{k \geq 1}$ is the basis given in (4.9)

end for

return: the vectors $\widehat{\theta} = (\widehat{\theta}_{1,\varepsilon}, ..., \widehat{\theta}_{n,\varepsilon})$ and $\widetilde{\theta} = (\widetilde{\theta}_{1,\varepsilon}, ..., \widetilde{\theta}_{n,\varepsilon})$

{**Step 3**} The cost function

for $i \longleftarrow 1$ *to* $[k^*]$ **do**

  for $j \longleftarrow [\varpi]$ *to* $[m\varpi]$ **do**

    $J_\varepsilon(\lambda) \longleftarrow \sum_{l=1}^n \lambda^2_{i,j}(l)\widehat{\theta}^2_{j,\varepsilon} - 2\sum_{j=1}^n \lambda_{i,j}(l)\widetilde{\theta}_{j,\varepsilon} + \delta\, P_\varepsilon(\lambda)$.

    where the vectors $\lambda = (\lambda_{i,j}(1), ..., \lambda_{i,j}(n))$ are computed in **Step1**, the vectors $\widehat{\theta}$ and $\widetilde{\theta}$ are given in **Step2** and P_ε is the penalty term given in (4.29)

  end for

end for

return: $\widehat{\lambda} = \mathrm{argmin}_{\lambda \in \Lambda} J_\varepsilon(\lambda), \Lambda = \{\lambda_\alpha\,, \alpha \in \mathcal{A}\}$.

Appendix

Property of the penalty term

Lemma 4.1. *Assume that Proposition 4.1 holds. Then for any $n \geq 1$ and $\lambda \in \Lambda$,*

$$P_{\varepsilon}(\lambda) \leq \mathcal{R}(\widehat{S}_{\lambda}, S) + \varepsilon^2 \check{\varkappa}_Q \,,$$

where the coefficient $P_{\varepsilon}(\lambda)$ is defined in (4.43).

Proof. By the definition of $\mathrm{Err}_{\varepsilon}(\lambda)$ one has

$$\mathrm{Err}_{\varepsilon}(\lambda) = \sum_{j=1}^{n} \left((\lambda(j) - 1)\theta_j + \varepsilon^2 \lambda(j) \check{\xi}_j \right)^2 .$$

Through Proposition 4.1 it is easy to see that

$$\mathbf{E}_Q \, \mathrm{Err}_{\varepsilon}(\lambda) \geq \varepsilon^2 \sum_{j=1}^{n} \lambda^2(j) \mathbf{E}_Q \, (\check{\xi}_j)^2 \geq P_{\varepsilon}(\lambda) - \varepsilon^2 \check{\varkappa}_Q \,.$$

Hence we obtain the desired result. □

Proof of the limit equality (4.78)

First, setting $\zeta_{\varepsilon} = \sum_{j=1}^{d} \kappa_j^2 \, a_j$, we obtain that

$$\left\{ S_{\kappa} \notin W_{\mathbf{r}}^{k} \right\} = \left\{ \zeta_{\varepsilon} > \mathbf{r} \right\} .$$

Moreover, note that one can check directly that

$$\lim_{\varepsilon \to 0} \mathbf{E} \, \zeta_{\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{v_{\varepsilon}} \sum_{j=1}^{d} s_j^* \, a_j = \check{\mathbf{r}} = (1 - \check{\gamma}) \mathbf{r} \,.$$

So, for sufficiently small ε we obtain that

$$\left\{ S_{\kappa} \notin W_{\mathbf{r}}^{k} \right\} \subset \left\{ \widetilde{\zeta}_{\varepsilon} > \mathbf{r}_1 \right\} ,$$

where $\mathbf{r}_1 = \mathbf{r}\check{\gamma}/2$, $\widetilde{\zeta}_{\varepsilon} = \zeta_{\varepsilon} - \mathbf{E} \, \zeta_{\varepsilon} = v_{\varepsilon}^{-1} \sum_{j=1}^{d} s_j^* a_j \widetilde{\eta}_j$ and $\widetilde{\eta}_j = \eta_j^2 - 1$ Through the correlation inequality (see, Proposition A.1 in [35]) we can get that for any $p \geq 2$

$$\mathbf{E} \, \widetilde{\zeta}_{\varepsilon}^{p} \leq (2p)^{p/2} \mathbf{E} |\widetilde{\eta}_1|^p \, v_{\varepsilon}^{-p} \left(\sum_{j=1}^{d} (s_j^*)^2 a_j^2 \right)^{p/2} = \mathrm{O}(\, v_{\varepsilon}^{-\frac{p}{4k+2}}) \,,$$

as $\varepsilon \to 0$. Therefore, for any $\check{\delta} > 0$ using the Chebychev inequality for $p > (4k+2)\check{\delta}$ we obtain that

$$v_{\varepsilon}^{\check{\delta}} \mathbf{P}(\widetilde{\zeta}_{\varepsilon} > \mathbf{r}_1) \to 0 \quad \text{as} \quad \varepsilon \to 0 \,.$$

Hence we obtain the equality (4.78). □

The absolute continuity of distributions for Lévy processes

In this section we study the absolute continuity for Lévy processes defined as

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le T\,, \tag{4.91}$$

where S is any arbitrary non-random square integrated function, i.e. from $\mathbf{L}_2[0,T]$ and $(\xi_t)_{0\le t\le T}$ is a Lévy process of the form (4.3) with nonzero constants ϱ_1 and ϱ_2. We denote by $\mathbf{P}_y$ and $\mathbf{P}_\xi$ the distributions of the processes $(y_t)_{0\le t\le 1}$ and $(\xi_t)_{0\le t\le 1}$ on the Skorokhod space $\mathbf{D}[0,T]$. Now for any $0 \le t \le T$ and $(x_t)_{0\le t\le T}$ from $\mathbf{D}[0,T]$ we set

$$\Upsilon_t(x) = \exp\left\{\int_0^t \frac{S(u)}{\varrho_1^2}\,\mathrm{d}x_u^c - \int_0^t \frac{S^2(u)}{2\varrho_1^2}\,\mathrm{d}u\right\}\,, \tag{4.92}$$

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

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

and, for any $t > 0$ and any measurable Γ from $\mathbb{R}\setminus\{0\}$,

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

Now we study the measures $\mathbf{P}_y$ and $\mathbf{P}_\xi$ in $\mathbf{D}[0,T]$.

Proposition 4.4. *For any $T > 0$, the measure $\mathbf{P}_y \ll \mathbf{P}_\xi$ in $\mathbf{D}[0,T]$ and the Radon-Nikodym derivative is*

$$\frac{\mathrm{d}\mathbf{P}_y}{\mathrm{d}\mathbf{P}_\xi}(\xi) = \Upsilon_T(\xi)\,.$$

Proof. Note that to show this proposition it suffices to check that for any $0 = t_0 < \ldots < t_n = T$ any $b_j \in \mathbb{R}$ for $1 \le j \le n$

$$\mathbf{E}\,\exp\left\{i\sum_{l=1}^n b_j(y_{t_j} - y_{t_{j-1}})\right\} = \mathbf{E}\,\exp\left\{i\sum_{l=1}^n b_j(\xi_{t_j} - \xi_{t_{j-1}})\right\}\Upsilon_T(\xi)\,.$$

Taking into account that the processes $(y_t)_{0\le t\le T}$ and $(\xi_t)_{0\le t\le T}$ have independent homogeneous increments, one needs to check only that for any $b \in \mathbb{R}$ and $0 \le s < t \le T$

$$\mathbf{E}\,\exp\{i\,b(y_t - y_s)\} = \mathbf{E}\,\exp\{i\,b(\xi_t - \xi_s)\}\,\frac{\Upsilon_t(\xi)}{\Upsilon_s(\xi)}\,. \tag{4.93}$$

To check this equality, note that the process

$$\Upsilon_t(\xi) = \exp\left\{\int_0^t \frac{S(u)}{\varrho_1}\,\mathrm{d}w_u - \int_0^t \frac{S^2(u)}{2\varrho_1^2}\,\mathrm{d}u\right\}$$

is a Gaussian martingale. From here we directly obtain Equation (4.93). □

The van Trees inequality for Lévy processes

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

$$\mathrm{d}y_t = S(t,\theta)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le 1\,, \tag{4.94}$$

where $S(t,\theta) = \sum_{i=1}^{d} \theta_i\,\phi_i(t)$ with the unknown parameters $\theta = (\theta_1,\dots,\theta_d)'$ and the process $(\xi_t)_{0\le t\le 1}$ is defined in (4.3). Note now that according to Proposition 4.4 the distribution $\mathbf{P}_\theta$ of the process (4.94) is absolutely continuous with respect to the $\mathbf{P}_\xi$ on $\mathbf{D}[0,1]$ and the corresponding Radon-Nikodym derivative is

$$f(x,\theta) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mathbf{P}_\xi}(x) = \exp\left\{\int_0^1 \frac{S(t,\theta)}{\varrho_1^2}\,\mathrm{d}x_t^c - \int_0^1 \frac{S^2(t,\theta)}{2\varrho_1^2}\,\mathrm{d}t\right\}, \tag{4.95}$$

where $x = (x_t)_{0\le t\le T}$ is an arbitrary function from $\mathbf{D}[0,1]$.

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

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

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

$$\lim_{|\theta_j|\to\infty} \mathbf{g}(\theta)\,\varphi_j(\theta_j) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} |\mathbf{g}_j'(\theta)|\,\Phi(\theta)\,\mathrm{d}\theta < \infty\,, \tag{4.96}$$

where

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

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

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

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

Lemma 4.2. *For any $\mathcal{F}^y = \sigma\{y_t, 0 \le t \le 1\}$-measurable square integrable function $\widehat{\mathbf{g}}$ and for any $1 \le j \le d$, the following inequality holds*

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

where

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

Proof. First of all, note that the density (4.95) of the process ξ is bounded with respect to $\theta_j \in \mathbb{R}$ and, for any $1 \le j \le d$,

$$\limsup_{|\theta_j|\to\infty} f(\xi,\theta) \,=\, 0\,. \qquad \text{a.s.}$$

Now, we set

$$\widetilde{\Phi}_j = \widetilde{\Phi}_j(x,\theta) = \frac{\partial\,(f(x,\theta)\Phi(\theta))/\partial\theta_j}{f(x,\theta)\Phi(\theta)}\,.$$

Taking into account the condition (4.96) and integrating by parts yield

$$\widetilde{\mathbf{E}}\left((\widehat{\mathbf{g}}-\mathbf{g}(\theta))\widetilde{\Phi}_j\right) = \int_{\mathcal{X}\times\mathbb{R}^d} (\widehat{\mathbf{g}}(x)-\mathbf{g}(\theta))\frac{\partial}{\partial\theta_j}\,(f(x,\theta)\Phi(\theta))\,\mathrm{d}\theta\,\mathbf{P}_\xi(\mathrm{d}x)$$

$$= \int_{\mathcal{X}\times\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \mathbf{g}_j'(\theta)\,f(x,\theta)\Phi(\theta)\mathrm{d}\theta_j\right)\left(\prod_{i\neq j}\mathrm{d}\theta_i\right)\,\mathbf{P}_\xi(\mathrm{d}x) = \Lambda_j\,.$$

Now by the Cauchy's inequality we obtain the following lower bound for the quadratic risk

$$\widetilde{\mathbf{E}}(\widehat{\mathbf{g}}-\mathbf{g}(\theta))^2 \ge \frac{\Lambda_j^2}{\widetilde{\mathbf{E}}\Psi_j^2}\,.$$

To study the denominator in the left handside of this inequality note that, in view of the reprentation (4.95),

$$\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j} = \frac{1}{\varrho_1}\int_0^1 \psi_j(t)\,\mathrm{d}w_t\,.$$

Therefore, for each $\theta\in\mathbb{R}^d$,

$$\mathbf{E}_\theta\,\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j} \,=0$$

and

$$\mathbf{E}_\theta\left(\frac{1}{f(y,\theta)}\frac{\partial\,f(y,\theta)}{\partial\theta_j}\right)^2 = \frac{1}{\varrho_1^2}\int_0^1 \psi_j^2(t)\mathrm{d}t = \frac{1}{\varrho_1^2}\|\psi\|^2\,.$$

Taking into account that

$$\widetilde{\Phi}_j = \frac{1}{f(x,\theta)}\frac{\partial\,f(x,\theta)}{\partial\theta_j} + \frac{1}{\Phi(\theta)}\frac{\partial\,\Phi(\theta))}{\partial\theta_j}\,,$$

we get

$$\widetilde{\mathbf{E}}\Psi_j^2 = \frac{n}{\varrho_1^2}\,\|\psi\|^2 + \,I_j\,.$$

Hence we got the desired result. □

Chapter 5

Renewal theory

5.1 Renewal density

This section is concerned with results related to the renewal measure (1.8).

Theorem 5.1. *(Goldie's theorem)*
Let η be a probability law on $\mathbb{R}$ with finite second moment and positive first moment m, such that $\tilde{\eta}(\beta) < \infty$ for some $\beta > 0$. Suppose that η is spread out, so for some n_0 we have

$$\eta^{(n_0)} = (1-\delta)\phi_0 + \delta\phi_1,$$

where $\delta \in [0,1[$ is constant and ϕ_0, ϕ_1 are probability measures with ϕ_0 absolutely continuous. Suppose that β has been taken so small that $\delta\tilde{\phi}_1(\beta) < 1$. Suppose that $\hat{\eta}(\theta) \neq 1$ on the line $\Im\theta = -\beta$. Then the renewal measure $\nu := \sum_{j=0}^{\infty} \eta^{(n)}$ may be written $\nu = \nu_0 + \nu_1$, where ν_1 is a finite measure such that $\tilde{\nu_1}(\beta) < \infty$, and ν_0 is absolutlely continuous with a continuous bounded density $p(.)$ such that

$$p(t) = \frac{1}{m} - \frac{1}{2\pi}\int_C e^{-i\theta t}\frac{d\theta}{1-\hat{\eta}(\theta)} + o(e^{-\beta t}),\ t \longrightarrow \infty.$$

Here C is a simple closed contour in the domain $\mathbf{D} := \{\theta :\ -\beta < \Im\theta < 0\}$, enclosing all the zeroes of $1-\hat{\eta}$ in $\mathbf{D}$, $\tilde{\eta}(\theta) = \int_{\mathbb{R}} e^{\theta t}\eta(dt)$ and $\hat{\eta}(\theta) = \int_{\mathbb{R}} e^{i\theta t}\eta(dt)$.
The proof of this Theorem is given in [8].

Now we need to adapt this result to our framwork. We start with the following lemma.

Lemma 5.1. *Let τ be a positive random variable with a density g, such that $\mathbf{E}e^{\beta\tau} < \infty$ for some $\beta > 0$. Then there exists a constant β_1, $0 < \beta_1 \leq \beta$ for which,*

$$\mathbf{E}e^{(\beta_1 + i\omega)\tau} \neq 1 \qquad \forall\omega \in \mathbb{R}\,.$$

Proof. We will show this lemma by the contradiction, i.e. we assume that there exist some sequence of positive numbers going to zero $(\gamma_k)_{k\geq 1}$ and a sequence $(w_k)_{k\geq 1}$ such that

$$\mathbf{E}e^{(\gamma_k + i\omega_k)\tau} = 1 \tag{5.1}$$

for any $k \geq 1$. Firstly, assume that $\limsup_{k\to\infty} w_k = +\infty$. Note that in this case, for any $N \geq 1$,

$$\left|\int_0^N e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t\right| \leq \left|\int_0^N \cos(w_k t)\, g(t)\mathrm{d}t\right| + \left|\int_0^N (e^{\gamma_k t}-1)\cos(w_k t)\, g(t)\mathrm{d}t\right|,$$

i.e., in view of Lemma 5.2, for any fixed $N \geq 1$

$$\limsup_{k\to\infty} \int_0^N e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t = 0\,.$$

Since for some $\beta > 0$ the integral $\int_0^{+\infty} e^{\beta t}\, g(t)\mathrm{d}t < \infty$, we get

$$\lim_{k\to\infty} \int_0^{+\infty} e^{\gamma_k t}\cos(w_k t)\, g(t)\mathrm{d}t = 0\,.$$

Let now assume that $\limsup_{k\to\infty} w_k = \omega_\infty \neq 0$ and $0 < |\omega_\infty| < \infty$. In this case there exists a sequence $(l_k)_{k\geq 1}$ such that $\lim_{k\to\infty} w_{l_k} = \omega_\infty$, i.e.

$$1 = \limsup_{k\to\infty} \mathbf{E} e^{\gamma_{l_k}\tau}\cos(\tau w_{l_k}) = \mathbf{E}\,\cos(\tau w_\infty)\,.$$

It is clear that, for random variables having density, the last equality is possible if and only if $w_\infty = 0$. In this case, i.e. when $\limsup_{k\to\infty} w_{l_k} = 0$, the equation (5.1) implies

$$\limsup_{k\to\infty} \mathbf{E}\, e^{\gamma_{l_k}\tau}\frac{\sin(\tau w_{l_k})}{w_{l_k}} = \mathbf{E}\,\tau = 0\,.$$

But, under our conditions, $\mathbf{E}\tau > 0$. These contradictions imply the desired result. □

Proposition 5.1. *Let τ be a positive random variable with the distribution η having a density g which satisfies Conditions $\mathbf{H}_1$)–$\mathbf{H}_4$). Then the renewal measure* (1.8) *is absolutely continuous with density ρ, for which*

$$\rho(x) = \frac{1}{\check{\tau}} + \Upsilon(x)\,, \tag{5.2}$$

where $\check{\tau} = \mathbf{E}\tau_1$ and $\Upsilon(\cdot)$ is some function defined on $\mathbb{R}_+$ with values in $\mathbb{R}$ such that

$$\sup_{x\geq 0} x^\gamma |\Upsilon(x)| < \infty \quad \textit{for all} \quad \gamma > 0\,.$$

Proof. First, note that we can represent the renewal measure $\check{\eta}$ as $\check{\eta} = \eta * \eta_0$ and $\eta_0 = \sum_{j=0}^\infty \eta^{(j)}$. It is clear that in this case the density ρ of $\check{\eta}$ can be written as

$$\rho(x) = \int_0^x g(x-y)\sum_{n\geq 0} g^{(n)}(y)\mathrm{d}y\,. \tag{5.3}$$

Now we use the arguments proposed in the proof of Lemma 9.5 from [8]. For any $0 < \epsilon < 1$ we set

$$\rho_\epsilon(x) = \int_0^x g(x-y) \left(\sum_{n\geq 0} (1-\epsilon)^n \, g^{(n)}(y) - \frac{(1-\epsilon)}{\check{\tau}} \, g_0(y) \right) \mathrm{d}y - g(x) \,, \tag{5.4}$$

where $g_0(y) = e^{-\epsilon y/\check{\tau}} 1_{\{y>0\}}$. It is easy to deduce that for any $x \in \mathbb{R}$

$$\lim_{\epsilon \to 0} \rho_\epsilon(x) \, = \, \rho(x) - \frac{1}{\check{\tau}} \int_0^x g(z) \, \mathrm{d}z - g(x) \,. \tag{5.5}$$

Moreover, in view of the condition $\mathbf{H}_1$) we obtain that the function $\rho_\epsilon(x)$ satisfies the condition **D**) from Section 5.2. So, through Proposition 5.3 we get

$$\rho_\epsilon(x+) + \rho_\epsilon(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{\rho}_\epsilon(\theta) \, \mathrm{d}\theta \,,$$

where $\widehat{\rho}_\epsilon(\theta) = \int_{\mathbb{R}} e^{i\theta x} \rho_\epsilon(x) \mathrm{d}x$. Note that

$$|\widehat{g}(\theta)| = \left| \int_{\mathbb{R}} e^{i\theta x} g(x) \mathrm{d}x \right| \leq \int_{\mathbb{R}} g(x) \mathrm{d}x = 1 \,,$$

i.e. for any $0 < \epsilon < 1$ we have $|(1-\epsilon)\widehat{g}(\theta)| < 1$ and therefore

$$\sum_{n=0}^{\infty} (1-\epsilon)^n (\widehat{g}(\theta))^n = \frac{1}{1-(1-\epsilon)\widehat{g}(\theta)} \,.$$

From this and, taking into account that

$$\widehat{g}_0(\theta) = \int_{\mathbb{R}} e^{i\theta x} g_0(x) \mathrm{d}x = \frac{\check{\tau}}{\epsilon - i\check{\tau}\theta} \,,$$

we obtain

$$\begin{aligned} \widehat{\rho}_\epsilon(\theta) &= \widehat{g}(\theta) \sum_{n=0}^{\infty} (1-\epsilon)^n (\widehat{g}(\theta))^n - \left(\frac{1-\epsilon}{\check{\tau}} \right) \widehat{g}(\theta) \widehat{g}_0(\theta) - \widehat{g}(\theta) \\ &= \widehat{g}(\theta) G_\epsilon(\theta) \quad \text{and} \quad G_\epsilon(\theta) = \frac{1}{1-(1-\epsilon)\widehat{g}(\theta)} - \frac{1-i\check{\tau}\theta}{\epsilon - i\check{\tau}\theta} \,, \end{aligned}$$

i.e.

$$\rho_\epsilon(x-) + \rho_\epsilon(x+) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{g}(\theta) G_\epsilon(\theta) \, \mathrm{d}\theta \,. \tag{5.6}$$

One can check directly that

$$\sup_{0<\epsilon<1, \theta \in \mathbb{R}} |G_\epsilon(\theta)| \, < \, \infty \,.$$

Therefore, using the condition $\mathbf{H}_3$) and the Lebesgue's dominated convergence theorem, we can pass to limit as $\epsilon \to 0$ in (5.6), i.e., we obtain that

$$\rho(x+) + \rho(x-) - \frac{2}{\check{\tau}} \int_0^x g(z) \, \mathrm{d}z - g(x+) - g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \, \widehat{g}(\theta) G_0(\theta) \, \mathrm{d}\theta \,,$$

where

$$G_0(\theta) = \frac{1}{1-\widehat{g}(\theta)} + \frac{1-i\check{\tau}\theta}{i\check{\tau}\theta}\,.$$

Using here again Proposition 5.3 we deduce that

$$\rho(x+) + \rho(x-) = \frac{2}{\check{\tau}} \int_0^x g(z)\,\mathrm{d}z + \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta \tag{5.7}$$

and

$$\check{G}(\theta) = \frac{1}{1-\widehat{g}(\theta)} + \frac{1}{i\check{\tau}\theta}\,.$$

Note now that we can represent the density (5.3) as

$$\rho(x) = g * \sum_{n\geq 0} g^{(n)} = \sum_{n\geq 1} g^{(n)}(x) = g(x) + \sum_{n\geq 2} g^{(n)}(x) =: g(x) + \rho_c(x)$$

and the function $\rho_c(x)$ is continuous for all $x \in \mathbb{R}$. This means that

$$\widetilde{\rho}(x) = \frac{\rho(x+) + \rho(x-)}{2} - \rho(x) = \frac{g(x+) + g(x-)}{2} - g(x)$$

and, therefore, the condition $\mathbf{H}_2$) implies that, for any $\gamma > 0$,

$$\sup_{x\geq 0} x^{\gamma}\,|\widetilde{\rho}(x)| < \infty.$$

Now we can rewrite (5.7) as

$$\rho(x) = \frac{1}{\check{\tau}} \int_0^x g(z)\,\mathrm{d}z + \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta - \widetilde{\rho}(x). \tag{5.8}$$

Taking into account that $\mathbf{E}\,e^{\beta\tau} < \infty$ for some $\beta > 0$ we can obtain that

$$\sup_{x\geq 0} x^{\gamma} \int_x^{+\infty} g(z)\,\mathrm{d}z < \infty\,.$$

To study the second term in (5.8) we will use Proposition 5.2. Indeed, Condition $\mathbf{H}_3$) implies the first limit equality in (5.10). The second one follows directly from Lemma 5.2. Therefore, in view of Proposition 5.2, there exists some $\beta^* > 0$ such that, for any $0 \leq \beta_0 \leq \beta^*$,

$$\int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta)\check{G}(\theta)\,\mathrm{d}\theta = e^{-\beta_0 x} \int_{\mathbb{R}} e^{-ix\theta}\,\widehat{g}(\theta - i\beta_0)\check{G}(\theta - i\beta_0)\,\mathrm{d}\theta\,.$$

Note that, due to Lemma 5.1, the function $1-\widehat{g}(\theta)$ has no zeros on the line $\{z \in \mathbb{C} : \mathrm{Im}(z) = -\beta_1\}$. Moreover, one can check directly that $\theta = 0$ is an isolated zero. So, this means that for any $N > 1$ there can be only finitely many zeros in $\{z \in \mathbb{C} : -\beta_1 < \mathrm{Im}(z) < 0\,,\, |\mathrm{Re}(z)| < N\}$ of the function $1 - \widehat{g}(\theta)$. Moreover, note that in view of lemma 5.2 for any $r > 0$

$$\lim_{\mathrm{Re}(\theta)\to\infty, |\mathrm{Im}(\theta)|\leq r} \widehat{g}(\theta) = 0\,.$$

This means that there exists $N > 0$ such that the function $1-\widehat{g}(\theta) \neq 0$ for $\theta \in \{z \in \mathbb{C} : -\beta_1 < \mathrm{Im}(z) < 0\,, |\mathrm{Re}(z)| \geq N\}$. So, there can be only finitely many zeros of the function $1 - \widehat{g}(\theta)$ in $\{z \in \mathbb{C} : -\beta_1 < Im(z) < 0\}$ for some fixed $0 < \beta_1 < \beta$. Therefore, there exists some $\beta_0 > 0$ for which the function $1 - \widehat{g}(\theta)$ has no zeros in $\{z \in \mathbb{C} : -\beta_0 < Im(z) < 0\}$, i.e. the function $\check{G}(\theta)$ will be bounded in this set and we obtain that

$$\sup_{x\geq 0} e^{\beta_0 x} \left| \int_{\mathbb{R}} e^{-ix\theta}\, \widehat{g}(\theta)\check{G}(\theta)\, \mathrm{d}\theta \right| < \infty\,.$$

Thus the conclusion follows. □

Using this proposition we can study the renewal process $(N_t)_{t\geq 0}$ introduced in (1.7).

Corollary 5.1. *Assume that Conditions* $\mathbf{H}_1$)–$\mathbf{H}_4$) *hold true. Then, for any* $t > 0$,

$$\mathbf{E}\, N_t \leq |\rho|_*\, t \quad \textit{and} \quad \mathbf{E}\, N_t^2 \leq\ |\rho|_*\, t +\ |\rho|_*^2\, t^2\,. \tag{5.9}$$

Proof. First, by means of Proposition 5.1, note that we get

$$\mathbf{E}\, N_t = \mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} = \int_0^t \rho(v)\, \mathrm{d}v \leq |\rho|_*\, t\,.$$

Regarding the last bound in (5.9), we use the same reasoning as in the previous inequality, i.e., we obtain

$$\begin{aligned}\mathbf{E}\, N_t^2 &= \mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} + 2\mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}} \sum_{j=k+1} \mathbf{1}_{\{T_j \leq t\}} \\ &= \mathbf{E}\, N_t + 2\mathbf{E} \sum_{k\geq 1} \mathbf{1}_{\{T_k \leq t\}}\, \Theta(T_k) = \mathbf{E}\, N_t + \int_0^t \Theta(v)\, \rho(v)\, \mathrm{d}v\,,\end{aligned}$$

where, for $0 \leq v \leq t$, we defined the function $\Theta(v) = \mathbf{E}\, N_{t-v} \leq |\rho|_*(t - v)$. □

5.2 Properties of the Fourier transform

Theorem 5.2. *Cauchy (1825)*

Let U be a simply connected open subset of $\mathbb{C}$, let $g : U \to \mathbb{C}$ be a holomorphic function, and let γ be a rectifiable path in U whose start point is equal to its end point. Then

$$\oint_\gamma g(z)\mathrm{d}z = 0\,.$$

Proposition 5.2. *Let $g : \mathbb{C} \to \mathbb{C}$ be a holomorphic function in $U = \{z \in \mathbb{C} : -\beta_1 < Imz < \beta_2\}$ for some $\beta_1 > 0$ and $\beta_2 > 0$. Assume that, for any $-\beta_1 \leq t \leq 0$,*

$$\int_{\mathbb{R}} |g(\theta + it)|\, \mathrm{d}\theta < \infty \quad \textit{and} \quad \lim_{|\theta|\to\infty} g(\theta + it)\ = 0\,. \tag{5.10}$$

Then, for any $x \in \mathbb{R}$ and for any $0 < \beta < \beta_1$,

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta)\,\mathrm{d}\theta = e^{-\beta x} \int_{\mathbb{R}} e^{i\theta x} g(\theta - i\beta)\,\mathrm{d}\theta. \tag{5.11}$$

Proof. First note that the conditions of this theorem imply that

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta)\,\mathrm{d}\theta = \lim_{N\to\infty} \int_{-N}^{N} e^{i\theta x} g(\theta)\,\mathrm{d}\theta\,.$$

We fix now $0 < \beta < \beta_1$ and we set for any $N \geq 1$

$$\gamma = \{z \in \mathbb{C} : -N \leq \mathrm{Re}z \leq N\,,\, \mathrm{Im}z = 0\} \cup \{z \in \mathbb{C} : -N \leq \mathrm{Im}z \leq N\,,\, \mathrm{Re}z = N\}$$

$$\cup \{z \in \mathbb{C} : -N \leq \mathrm{Re}z \leq N\,,\, \mathrm{Im}z = -\beta\} \cup \{z \in \mathbb{C} : -\beta \leq \mathrm{Im}z \leq 0\,,\, \mathrm{Re}z = -N\}\,.$$

Now, in view of Theorem 5.2, we obtain that for any $N \geq 1$

$$\oint_{\gamma} e^{izx}\, g(z)\mathrm{d}z = \int_{-N}^{N} e^{i\theta x} g(\theta)\,\mathrm{d}\theta + \int_{0}^{-\beta} e^{i(N+it)x} g(N+it)\,\mathrm{d}t$$

$$+ \int_{N}^{-N} e^{i(-i\beta+\theta)x} g(-i\beta+\theta)\mathrm{d}\theta + \int_{-\beta}^{0} e^{i(-N+it)x} g(-N+it)\mathrm{d}t = 0\,. \tag{5.12}$$

Conditions (5.10) provide that

$$\lim_{N\to\infty} \int_{0}^{-\beta} e^{i(N+it)x} g(N+it)\,\mathrm{d}t = \lim_{N\to\infty} \int_{-\beta}^{0} e^{i(-N+it)x} g(-N+it)\,\mathrm{d}t = 0\,.$$

Therefore, letting $N \to \infty$ in (5.12) we obtain (5.11). Hence we get the desired result. □

The following technical lemma is also needed.

Lemma 5.2. *Let $g : [a,b] \to \mathbb{R}$ be a function from $\mathbf{L}_1[a,b]$. Then, for any fixed $-\infty \leq a < b \leq +\infty$,*

$$\lim_{N\to\infty} \int_{a}^{b} g(x)\,\sin(Nx)\mathrm{d}x = 0 \quad \textit{and} \quad \lim_{N\to\infty} \int_{a}^{b} g(x)\,\cos(Nx)\mathrm{d}x = 0\,. \tag{5.13}$$

Proof. Let first $-\infty < a < b < +\infty$. Assume that g is continuously differentiable, i.e. $g \in \mathbf{C}^1[a,b]$. Then integrating by parts gives us

$$\int_{a}^{b} g(x)\,\sin(Nx)\,\mathrm{d}x = \frac{1}{N}\left(g(b)\,\sin(Nb)\; - g(a)\,\sin(Na)\; - \int_{a}^{b} g^{'}(x)\,\cos(Nx)\,\mathrm{d}x\right).$$

So, from this we obtain that

$$\left|\int_{a}^{b} g(x)\,\sin(Nx)\,\mathrm{d}x\right| \leq \frac{|g(a)| + |g(a)| + (b-a)\max_{a\leq x\leq b}|g^{'}(x)|}{N}\,.$$

This implies the first limit in (5.13) for this case. The second one is obtained similarly. Let now g be any absolutely integrated function on $[a,b]$, i.e. $g \in \mathbf{L}_1[a,b]$. In this case there exists a sequence $g_n \in \mathbf{C}^1[a,b]$ such that

$$\lim_{n\to\infty} \int_a^b |g(x) - g_n(x)| \mathrm{d}x = 0\,.$$

Therefore, taking into account that for any $n \geq 1$

$$\lim_{N\to\infty} \int_a^b g_n(x)\, \sin(Nx) \mathrm{d}x = 0\,,$$

we obtain that

$$\limsup_{n\to\infty} \left| \int_a^b g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \int_a^b |g(x) - g_n(x)| \mathrm{d}x\,.$$

So, letting in this inequality $n \to \infty$ we obtain the first limit in (5.13) and, similarly, we obtain the second one. Let now $b = +\infty$ and $a = -\infty$. In this case we obtain that for any $-\infty < a < b < +\infty$

$$\left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| + \int_b^{+\infty} |g(x)\,| \mathrm{d}x + \int_{-\infty}^{a} |g(x)\,| \mathrm{d}x\,.$$

Using here the previous results we obtain that for any $-\infty < a < b < +\infty$

$$\limsup_{N\to\infty} \left| \int_{-\infty}^{+\infty} g(x)\, \sin(Nx) \mathrm{d}x \right| \leq \int_b^{+\infty} |g(x)\,| \mathrm{d}x + \int_{-\infty}^{a} |g(x)\,| \mathrm{d}x\,.$$

Passing here to limit as $b \to +\infty$ and $a \to -\infty$ we obtain the first limit in (5.13). Similarly, we can obtain the second one. □

Let us now study the inverse Fourier transform. To this end, we need the following local Dini condition.

D) *Assume that, for some fixed $x \in \mathbb{R}$, there exist the finite limits*

$$g(x-) = \lim_{z\to x-} g(z) \quad \textit{and} \quad g(x+) = \lim_{z\to x+} g(z)$$

and there exists $\delta = \delta(x) > 0$ for which

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

Proposition 5.3. *Let $g : \mathbb{R} \to \mathbb{R}$ be a function from $\mathbf{L}_1(\mathbb{R})$. If, for some $x \in \mathbb{R}$, this function satisfies the condition* **D**, *then*

$$g(x+) + g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-i\theta x} \widehat{g}(\theta)\, \mathrm{d}\theta\,, \tag{5.14}$$

where

$$\widehat{g}(\theta) = \int_{\mathbb{R}} e^{i\theta t}\, g(t)\, \mathrm{d}t\,.$$

Proof. First, for any fixed $N > 0$ we set

$$J_N(x) = \frac{1}{2\pi} \int_{-N}^{N} e^{-i\theta x} \widehat{g}(\theta)\, \mathrm{d}\theta = \frac{1}{\pi} \int_{\mathbb{R}} g(z) \int_0^N \cos(\theta(z-x))\, \mathrm{d}\theta \mathrm{d}z\,,$$

i.e.,

$$J_N(x) = \frac{1}{\pi} \int_{\mathbb{R}} g(z)\, \frac{\sin(N(z-x))}{z-x}\, \mathrm{d}z = \frac{1}{\pi} \int_0^{\infty} (g(x+t) + g(x-t))\, \frac{\sin(Nt)}{t}\, \mathrm{d}t\,.$$

Taking into account that for any $N > 0$ the integral

$$\frac{2}{\pi} \int_0^{\infty} \frac{\sin(Nt)}{t}\, \mathrm{d}t = 1 \tag{5.15}$$

and denoting $B(x) = (g(x+) + g(x-))/2$, we obtain that

$$J_N(x) - B(x) = \frac{1}{\pi} \int_0^{\infty} \frac{\omega(x,t)\sin(Nt)}{t}\, \mathrm{d}t \quad \text{and} \quad \omega(x,t) = g(x+t) + g(x-t) - 2B(x)\,.$$

Now we represent the last integral as

$$\int_0^{\infty} \frac{\omega(x,t)\sin(Nt)}{t}\, \mathrm{d}t = I_{1,N} + I_{2,N} - 2B(x) I_{3,N}\,,$$

where

$$I_{1,N} = \int_0^{\delta} \frac{\omega(x,t)}{t}\, \sin(Nt)\mathrm{d}t\,, \quad I_{2,N} = \int_{\delta}^{\infty} G(t)\, \sin(Nt)\mathrm{d}t\,, \quad I_{3,N} = \int_{\delta}^{\infty} \frac{\sin(Nt)}{t}\, \mathrm{d}t$$

and $G(t) = (g(x+t) + g(x-t))/t$. Condition **D** and Lemma 5.2 imply directly the convergence $I_{1,N} \to 0$ as $N \to \infty$. Now note that, since $g \in \mathbf{L}_1(\mathbb{R})$, then the function G is absolutely integrated. Therefore, in view of Lemma 5.2, $I_{2,N} \to 0$ as $N \to \infty$. As to the last integral we use the property (5.15), i.e., the changing of the variables gives

$$I_{3,N} = \int_{\delta N}^{\infty} \frac{\sin t}{t}\, \mathrm{d}t \to 0 \quad \text{as} \quad N \to \infty\,.$$

Hence we have the desired result. □

Conclusion

The main purpose of this work is the non-parametric estimation for regression models in continuous time. First, we consider the problem of estimation an unknown fonction S on the basis of continuous observations, we define the noise in this model through a semi-Markov process which keeps the dependence for any duration n. So, we are in the case of dependent observations for which the dependence does not disapear for a sufficient large duration of observation. Second, we consider the same model when the estimation is based on discrete data and we obtain the sufficient conditions on the frequency observations under wich the robust effecient is shown. In the third model we consider a Lévy non-parametric regression with noise intensity and we estimate the unknown function S in the case where the noise level goes to 0 and the Lévy measure can go to infinity. In all of these models, we propose an adaptive model selection for the robust risk.

Bibliography

[1] V. S. Barbu, S. Beltaief and S. M. Pergamenchtchikov. Robust adaptive efficient estimation for semi - Markov nonparametric regression models. - Preprint, **15 April** (2016), *https://arxiv.org/pdf/1604.04516.pdf* submitted in *Statistical inference for stochastic processes*

[2] V. S. Barbu, S. Beltaief and S. M. Pergamenchtchikov. Robust adaptive efficient estimation for a semi - Markov continuous time regression from discrete data. - Preprint, **24 Mai** (2017),

[3] S. Beltaief, O. Chernoyaro and S. M. Pergamenchtchikov. Model selection for the robust efficient estimation in the Levy nonparametric regression. - Preprint, **23 Nov** (2016), *https://arxiv.org/pdf/1611.07378.pdf*

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

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

[6] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall, 2004.

[7] K. Bichteler and J. Jacod. *Calcul de Malliavin pour les diffusions avec sauts: existence d'une densité dans le cas unidimensionnel.* Séminaire de Probabilité, XVII, Lecture Notes in Math., **986**, Springer, Berlin, 132–157, 1983.

[8] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, **1** (1), 126–166, 1991.

[9] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation: Asymptotic Theory.* Springer, Berlin-New York, 1981.

[10] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes.* 2nd edition, Springer, Berlin, 2002.

[11] 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.

[12] 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.

[13] 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**, 215–235, 2003.

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

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

[16] 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**, 1083–1111, 2010.

[17] 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.

[18] V. V. Konev and S. M. Pergamenshchikov. Robust model selection for a semi-martingale continuous time regression from discrete data. *Stochastic Processes and their Applications*, **125**, 294–326, 2015.

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

[20] N. Limnios and G. Oprisan. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston, 2001.

[21] R. Sh. Liptser and A. N. Shiryaev. *Theory of Martingales*. Springer, 1986.

[22] T. Mikosch. *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes*. Springer, 2004.

[23] C. Mallows. Some comments on C_p. *Technometrics* **15**,661–675, 1973.

[24] C. Marinelli and M. Röckner. On maximal inequalities for purely discontinuous martingales in infinite dimensions. *Séminaire de Probabilités*, Lect. Notes Math., **XLVI**, 293–315, 2014.

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

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

[27] M. S. Pinsker. Optimal filtration of square integrable signals in Gaussian white noise. *Problems of Transmission Information*, **17**, 120–133, 1981.

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

[29] C. Mallows. Some comments on C_p. *Technometrics* **15**, 661–675, 1973.

[30] Akaike H. A new look at the statistical model identification. *IEEE Trans. on Automatic Control* **19**, 716–723, 1974.

[31] A. Barron, L. Birgé and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301–415, 1999.

[32] K. Bichteler and J. Jacod *Calcul de Malliavin pour les diffusions avec sauts: existence dune densité dans le cas unidimensionnel*. Séminaire de Probabilité, XVII, Lecture Notes in Math., **986**, Springer, Berlin, 132–157, 1983.

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

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

[35] L. Galtchouk, and S. Pergamenshchikov. Uniform concentration inequality for ergodic diffusion processes observed at discrete times. *Stochastic Processes and their Applications*, **123**, 91109, 2013.

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

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

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

[39] L. I. Galtchouk and S. M. Pergamenshchikov. Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection, 2009, *http://hal.archives-ouvertes.fr/hal-00326910/fr/.*

[40] A. Kneip. Ordered linear smoothers. *Annals of Statistcs* **22**, 835–866, 1994.

[41] Yu. A. Kutoyants. *Parameter Estimation for Stochastic Processes.* Heldeman-Verlag, Berlin, 1984.

[42] B. V. Manelis. Algorithms for tracking and demodulation of the mobile communication signal in the conditions of the Indiscriminating multipath propagation, *Radiotekhnika*, 4, 16–21, 2007 (in Russian).

[43] C. Marinelli and M. Röckner. On maximal inequalities for purely discontinuous martingales in infinite dimensions. *Séminaire de Probabilités*, Lect. Notes Math., **XLVI** , 293–315, 2014.

[44] A. A. Novikov. On discontinuous martingales. *Theory Probab. Appl.*, **20**, 1, 11–26, 1975.

[45] P. Massart. *A non-asymptotic theory for model selection.* European Congress of Mathematics, Eur. Math. Soc., Zrich, 2005.

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

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

[48] A. P. Trifonov and A. V. Kharin. Estimating the number of signals with unknown amplitudes . *Nelineinyi Mir*, **11** (12), 853 – 866, 2013 (in Russian).

[49] A. P. Trifonov, A. V. Kharin, O.V. Chernoyarov and K. S. Kalashnikov. Determining the number of radio signals with unknown phases. *International Journal on Communications Antenna and Propagation*, **5** (6), 367–374, 2015.

[50] A. P. Trifonov and Yu. S. Shinakov. *Joint distinction of signals and estimation of their parameters on the background noise*, Radio i Svyaz', Moscow, 1986 (in Russian).

[51] A. P. Trifonov and A. V. Kharin. Estimation of the number of radio signals with unknown amplitudes and phases. *Radiophysics and Quantum Electronics*, **58** (1), 57–70, 2015.

[52] A. P. Trifonov and A. V. Kharin . Estimation of the number of orthogonal signals with unknown non-energy parameters. *Radioelectronics and Communications Systems*, **58** (8), 362 – 370, 2015.

[53] A. P. Trifonov, A. V. Kharin and O. V. Chernoyarov. Estimation of the number of orthogonal signals with unknown parameters. *Journal of Communications Technology and Electronics*, **61** (9), 1026 –1033, 2016.

[54] L. I. Volkov, , M. S. Nemirovskii and Yu. S. Shinakov. *Digital Radio Communication Systems.* Eko-Trendz, Moscow, 2005 (in Russian).

[55] M. Wax and M. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech Signal Process*, **33** (2), 387 – 392, 1985.