



HAL
open science

Efficient adaptive nonparametric estimation in heteroscedastic regression models

Leonid Galtchouk, Sergei Pergamenshchikov

► **To cite this version:**

Leonid Galtchouk, Sergei Pergamenshchikov. Efficient adaptive nonparametric estimation in heteroscedastic regression models. 2005. hal-00129707

HAL Id: hal-00129707

<https://hal.science/hal-00129707>

Preprint submitted on 8 Feb 2007

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

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

Efficient adaptive nonparametric estimation in heteroscedastic regression models *

L. Galtchouk¹, S. Pergamenshchikov²

¹IRMA, Département de Mathématiques,
Université de Strasbourg, 7 rue René Descartes,
F67084, Strasbourg Cedex France
e-mail: galtchou@math.u-strasbg.fr

²Laboratoire de Mathématiques Raphael Salem,
UMR CNRS 6085, Avenue de l'Université, BP. 12,
UFR Sciences, Université de Rouen,
F76801, Saint Etienne du Rouvray, Cedex, France.
e-mail: Serge.Pergamenchtchikov@univ-rouen.fr

Abstract

An adaptive nonparametric estimation procedure is constructed for the estimation problem of heteroscedastic regression. A non-asymptotic upper bound for the quadratic risk (the Oracle inequality) is obtained. Asymptotic efficiency of this procedure is proved, i.e. Pinsker's constant is found in the asymptotical lower bound for the risk. It is shown that the asymptotical quadratic risk for the constructed procedure coincides with this constant.

Key words: adaptive estimation, asymptotic bounds, efficient estimation, heteroscedastic regression, nonparametric estimation, non-asymptotic estimation, Oracle inequality, Pinsker's constant.

AMS (1991) Subject Classification : primary 62G08; secondary 62G05, 62G20

*The second author is partially supported by the RFFI-Grant 04-01-00855.

1 Introduction

Suppose we are given observations $(y_j)_{1 \leq j \leq n}$ which obey the heteroscedastic regression equation

$$y_j = S(x_j) + \sigma_j(S) \xi_j, \quad (1.1)$$

where design points $x_j = j/n$, $S(\cdot)$ is an unknown function to be estimated, $(\xi_j)_{1 \leq j \leq n}$ is the sequence of i.i.d. random variables, $(\sigma_j(S))_{1 \leq j \leq n}$ are unknown scale functionals depending on unknown regression function S and the design points. Our goal is to estimate S in the mean quadratic sense in the both non-asymptotic and asymptotic setups when the smoothness of S is unknown.

Note that heteroscedastic regressions with this type of scale functionals has been encountered in consumer budget studies utilizing observations on individuals with diverse incomes and in analyses of the investment behavior of firms of different sizes (see, for example, Goldfeld and Quandt (1972)). Such kind of models was considered by Gunst and Mason (1980), Efroimovich (1999), Akritas and Van Keilegom (2001) as well.

The problem of minimax estimation of nonparametric regression function in homoscedastic case (i.e. $\sigma_j \equiv \sigma$) in the asymptotic setup has been studied in a number of papers. The optimal rate for L_2 -losses and regression functions from a L_2 -Sobolev or a Hölder space was studied by Ibragimov and Hasminskii (1982), Speckman (1985), Donoho et al.(1995); the case of regression functions from Triebel and Besov space was investigated by Donoho and Johnston (1998); the optimal rate for model (1.1) in L_q -losses and adaptive estimate for Sobolev spacial regression was considered by Nemirovskii (2000). Efficient linear estimators for L_2 -risk over linear estimators and some other risks was given by Donoho and Liu (1991), Donoho (1994).

The notion of asymptotic optimality is usually associated with convergence rate of the minimax risk (see for example Ibragimov and Hasminskii (1981), Stone (1982))). An important question in the development of non-parametric estimation is to study the exact asymptotic behaviour of minimax risk and to find an efficient estimator, i.e. an estimator which achieves this asymptotics.

The optimal constant and efficient estimators for L_2 -losses was obtained by Nussbaum (1985), Golubev (1992), Golubev and Nussbaum (1993), for sup-norm losses by Korostelev (1993). For the absolute error loss and the estimation of regression function S at a fixed point, the optimal constant and efficient estimators was found by Galtchouk and Pergamenschikov (2005).

A non-asymptotic approach for nonparametric estimation problem in the model (1.1) with $\sigma_j \equiv \sigma$ was studied in a few papers. A non-asymptotic upper bound for quadratic risk over thresholding estimators is given by Kalifa and

Mallat (2003). Barron et al. (1999), Massart (2004) have constructed an adaptive procedure of model selection based on least squares estimators and have obtained a non-asymptotic upper bound for quadratic risk which is best in the principal term for given class of estimators. This type of upper bounds is called the Oracle inequalities.

It should be noted that to obtain efficient estimators (in asymptotic setup), in contrast with least squares estimators Nussbaum (1985) and later Golubev and Nussbaum (1993) have used some special class of weighted least squares estimators, where the weights are depending on regularity of unknown function and was chosen specially to obtain the optimal constant in quadratic risk (so-called the Pinsker constant).

By reason of asymptotic efficiency, we make use of Golubev-Nussbaum estimators to obtain non-asymptotic upper bound for quadratic risk in this paper. The main distinction of our approach from that of Barron-Birge-Massart is the following one. We choose a family of estimators such that the every estimator is optimal when the regularity of unknown function is fixed and a basis is given. In contrast with our approach, in the selection model theory one chooses a family of basis (models) for which a least squares estimator will be constructed.

In this paper an adaptive procedure is proposed which is based on weighted least squares estimators and the non-asymptotic upper bound for the quadratic risk is obtained which is best in the principal term for the chosen family of estimators. Moreover, at the same time the asymptotic properties of the proposed procedure are studied as $n \rightarrow \infty$. It turns out that this procedure is asymptotically efficient, that is the asymptotic quadratic risk coincides with the lower bound for all estimators, i.e. with the Pinsker constant.

The paper is organized as follows. In the next section we formulate the problem and give main results. In section 3 we construct an adaptive estimation procedure based on Nussbaum's estimators and we obtain a non-asymptotic upper bound for the quadratic risk of this procedure. In section 4 we prove the sharp asymptotic upper bound for the quadratic risk of this estimator. In section 5 the sharp asymptotic lower bound for the minimax risk is obtained. An appendix contains some technical results.

2 Problems and Results

First we consider the model (1.1) in which the sequence $(\xi_j)_{1 \leq j \leq n}$ is i.i.d. with

$$\mathbf{E} \xi_1 = 0, \quad \mathbf{E} \xi_1^2 = 1 \quad \text{and} \quad \mathbf{E} \xi_1^4 = \xi^* < \infty. \quad (2.1)$$

We introduce the Sobolev class W_r^q as

$$W_r^q = \{S \in \mathbf{C}^{q-1}[0, 1] : \sum_{j=0}^q \|S^{(j)}\|^2 \leq r\}, \quad (2.2)$$

where $S^{(j)}$ is the j -th derivative of S , r is some positive constant, $q \geq 1$ is an integer,

$$\|S\|^2 = \int_0^1 S^2(x) dx.$$

For any estimate \hat{S}_n of S , based on observations $(y_j)_{1 \leq j \leq n}$, we use the following quadratic risk

$$\mathcal{R}_n(\hat{S}_n, S) = \mathbf{E}_S \|\hat{S}_n - S\|_n^2, \quad (2.3)$$

where

$$\|S\|_n^2 = \frac{1}{n} \sum_{j=1}^n S^2(x_j).$$

In the next section we construct an adaptive procedure (see below (3.8)) for which we obtain a non-asymptotic upper bound for the quadratic risk. To construct this procedure, we make use of the estimator family introduced by Nussbaum (1985), see below (3.4)-(3.5), and the approach proposed by Golubev and Nussbaum (1993) for the homoscedastic case, i.e. $\sigma_l \equiv \sigma$. The key idea in the construction of this procedure is the following one. We replace the unknown variance σ^2 in the Golubev-Nussbaum estimation procedure by some estimator for $n^{-1} \sum_{j=1}^n \sigma_j^2$. For this procedure we obtain the non-asymptotic upper bound for the risk (2.3) which is best for the chosen family of estimators (see Theorem 3.3). This type of upper bound is called *Oracle inequality*.

To study asymptotic properties, we need to impose some additional conditions on the sequence (σ_j) in (1.1).

H₁) $\sigma_j = g(x_j, S)$ for some unknown function $g : [0, 1] \times \mathbf{L}_1[0, 1] \rightarrow \mathbb{R}_+$, which is square integrable with respect to x such that

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^q} \left| \frac{1}{n} \sum_{j=1}^n g^2(x_j, S) - \varsigma(S) \right| = 0, \quad (2.4)$$

where $\varsigma(S) := \int_0^1 g^2(x, S) dx$. Moreover there exist some reals $0 < g_* \leq g^* < \infty$, for which

$$g_* \leq \inf_{0 \leq x \leq 1} \inf_{S \in W_q^r} g(x, S) \leq \sup_{0 \leq x \leq 1} \sup_{S \in W_q^r} g(x, S) \leq g^*. \quad (2.5)$$

H₂) The function $g(x, S)$ is differentiable in the Frechet sense with respect to S in $\mathbf{L}_1[0, 1]$ uniformly over $0 \leq x \leq 1$, i.e. for any S, S_0 from $\mathbf{L}_1[0, 1]$

$$g(x, S) = g(x, S_0) + L_{x, S_0}(S - S_0) + \Upsilon(x, S_0, S),$$

where the linear operator L_{x, S_0} (the Frechet derivative) is bounded uniformly over $0 \leq x \leq 1$ in $\mathbf{L}_1[0, 1]$, i.e. for any S_0 from $\mathbf{L}_1[0, 1]$ there exists some positive constant $C^* = C^*(S_0)$ such that

$$\sup_{0 \leq x \leq 1} \sup_{S \in \mathbf{L}_1[0, 1], S \neq 0} \frac{|L_{x, S_0}(S)|}{\|S\|_1} \leq C^* \quad (2.6)$$

and the residual term $\Upsilon(x, S_0, S)$ satisfies the following property

$$\lim_{\|S\|_1 \rightarrow 0} \sup_{0 \leq x \leq 1} \frac{|\Upsilon(x, S_0, S)|}{\|S\|_1} = 0,$$

where $\|S\|_1 = \int_0^1 |S(x)| dx$.

H₃) The function $g_0(x) = g(x, 0)$ is continuous on the interval $[0, 1]$.

Now we formulate the main asymptotic results. We set

$$\gamma_q(S) = C_q^* r^{1/(2q+1)} (\zeta(S))^{2q/(2q+1)}, \quad (2.7)$$

with

$$C_q^* = (2q + 1)^{1/(2q+1)} \left(\frac{q}{\pi(q+1)} \right)^{2q/(2q+1)}.$$

It is well known (see, for example [19]) that the optimal rate is $\varphi_n = n^{2q/(2q+1)}$ in the case when $S \in W_r^q$.

Theorem 2.1. Assume that in the model (1.1)-(2.1) the sequence (σ_j) fulfils the condition **H₁**). Then the estimator S_n^* defined by (3.8) with $\varepsilon = 1/\ln(n+1)$ satisfies the inequality

$$\limsup_{n \rightarrow \infty} \varphi_n \sup_{S \in W_r^q} \frac{1}{\gamma_q(S)} \mathcal{R}_n(S_n^*, S) \leq 1 \quad (2.8)$$

Moreover we will show that the estimator S_n^* is efficient in the following sense.

Theorem 2.2. *Assume that in the model (1.1) the sequence (ξ_j) is i.i.d. $\sim \mathcal{N}(0, 1)$ and the sequence (σ_j) fulfils the conditions $\mathbf{H}_1)$ – $\mathbf{H}_3)$. Then the risk (2.3) admets the following asymptotic lower bound*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n} \varphi_n \sup_{S \in W_r^q} \frac{1}{\gamma_q(S)} \mathcal{R}_n(\hat{S}_n, S) \geq 1. \quad (2.9)$$

Remark 2.1. *Inequalities (2.8) and (2.9) imply that $\gamma_q(S)$ is the so-called Pinsker's constant, i.e. the sharp asymptotic lower bound for the quadratic risk.*

3 Non-asymptotic estimation

In this section we construct the estimation procedure for the model (1.1)–(2.1). We use the trigonometric basis $\{\phi_j, j \geq 1\}$ in $L_2[0, 1]$ with

$$\phi_1(x) = 1, \quad \phi_j(x) = \sqrt{2} \text{Tr}_j(2\pi[j/2]x), \quad j \geq 2, \quad (3.1)$$

where $[a]$ is the integer part of a real a ,

$$\text{Tr}_j(x) = \begin{cases} \cos x & \text{for even } j, \\ \sin x & \text{for odd } j. \end{cases}$$

It is easy to see that for this basis

$$(\phi_i, \phi_j)_n = \frac{1}{n} \sum_{l=1}^n \phi_i(x_l) \phi_j(x_l) = 0$$

for $i \neq j$ and $1 \leq i, j \leq n$. Moreover, $\|\phi_j\|_n^2 = 1$ for $j \leq n-1$ and $\|\phi_n\|_n^2 = \nu_n$ with

$$\nu_n = \begin{cases} 2 & \text{for even } n, \\ 1 & \text{for odd } n. \end{cases} \quad (3.2)$$

Now by making use of the discrete Fourier transformation we reduce the model (1.1) to the model

$$\hat{\theta}_{j,n} = \theta_{j,n} + \frac{1}{\sqrt{n}} \xi_{j,n} \quad (3.3)$$

with

$$\hat{\theta}_{j,n} = (y, \phi_j)_n, \quad \theta_{j,n} = (S, \phi_j)_n, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} \sum_{l=1}^n \sigma_l \xi_l \phi_j(x_l).$$

For any given $\varepsilon > 0$, let us define a 2-dimensional parameter $\tau = (\beta, t)$ with values into the set $\mathcal{A}_\varepsilon = \{1, \dots, \beta_*\} \times \{t_1, \dots, t_m\}$, where $\beta_* = \lceil 1/\sqrt{\varepsilon} \rceil$, $t_i = i\varepsilon$ for $1 \leq i \leq m = \lceil 1/\varepsilon^2 \rceil$. Notice that the number of elements in the set \mathcal{A}_ε is $v_\varepsilon = \text{card}(\mathcal{A}_\varepsilon) = \beta_* \cdot \lceil 1/\varepsilon^2 \rceil$.

For any $\tau \in \mathcal{A}_\varepsilon$ we set

$$\rho_\tau(j) = \begin{cases} 1 & \text{for } 1 \leq j \leq k_0(\tau) \\ \Psi_\beta(j w_\tau^{-1}) & \text{for } j > k_0(\tau), \end{cases} \quad (3.4)$$

where

$$\begin{aligned} \Psi_\beta(z) &= (1 - z^\beta) \mathbf{1}_{(|z| \leq 1)}, \quad w_\tau = n^{1/(2\beta+1)} c_\tau, \\ c_\tau &= \pi^{-2\beta/(2\beta+1)} (t/\iota_\beta)^{1/(2\beta+1)}, \quad \iota_\beta = \frac{\beta}{(\beta+1)(2\beta+1)}, \\ k_0(\tau) &= w_\tau / \ln(n+2). \end{aligned}$$

For every $\tau \in \mathcal{A}_\varepsilon$, we define the following estimator

$$\hat{S}_\tau(x) = \sum_{j=1}^n \rho_\tau(j) \frac{1}{\|\phi_j\|_n^2} \hat{\theta}_{j,n} \phi_j(x). \quad (3.5)$$

In the sequel we use the following weighted Fourier transformation of S :

$$S_\tau(x) = \sum_{j=1}^n \rho_\tau(j) \frac{1}{\|\phi_j\|_n^2} \theta_{j,n} \phi_j(x).$$

To estimate the unknown function S by estimators (3.4), we should find τ from \mathcal{A}_ε which minimizes the loss function

$$\|\hat{S}_\tau - S\|_n^2 = \sum_{j=1}^n \rho_\tau^2(j) \frac{\hat{\theta}_{j,n}^2}{\|\phi_j\|_n^2} - 2 \sum_{j=1}^n \rho_\tau(j) \frac{\hat{\theta}_{j,n} \theta_{j,n}}{\|\phi_j\|_n^2} + \sum_{j=1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2} \quad (3.6)$$

or equivalently, minimizes the function

$$\sum_{j=1}^n \rho_\tau^2(j) \frac{\hat{\theta}_{j,n}^2}{\|\phi_j\|_n^2} - 2 \sum_{j=1}^n \rho_\tau(j) \frac{\hat{\theta}_{j,n} \theta_{j,n}}{\|\phi_j\|_n^2}.$$

Since the coefficients $\theta_{j,n}$ are unknown, we will find τ which minimizes the function

$$J_n(\tau) = \sum_{j=1}^n \rho_\tau^2(j) \frac{\hat{\theta}_{j,n}^2}{\|\phi_j\|_n^2} - 2 \sum_{j=1}^n \rho_\tau(j) \frac{\hat{\theta}_{j,n}^2 - \hat{\zeta}_n/n}{\|\phi_j\|_n^2},$$

where $\hat{\varsigma}_n$ is some estimator of

$$\varsigma_n = \frac{1}{n} \sum_{l=1}^n \sigma_l^2. \quad (3.7)$$

The estimator $\hat{\varsigma}_n$ will be constructed below. Denote by $\hat{\tau} = (\hat{\beta}, \hat{t})$ the argmin of the function $J_n(\tau)$, i.e.

$$J_n(\hat{\tau}) = \min_{\tau \in \mathcal{A}_\varepsilon} J_n(\tau).$$

We denote

$$S_{n,\varepsilon}^* = \hat{S}_{\hat{\tau}}. \quad (3.8)$$

In this section we study this estimator in the non-asymptotic setup.

Theorem 3.1. *Assume that in the model (1.1)–(2.1) the function S belongs to $\cup_{\beta \geq 1, 0 < r \leq r_*} W_r^\beta$, where r_* is an unknown constant. Then, for any $n \geq 1$ and $\varepsilon > 0$, the estimate $S_{n,\varepsilon}^*$ satisfies the following inequality*

$$\begin{aligned} \mathbf{E}_S \|S_{n,\varepsilon}^* - S\|_n^2 &\leq \frac{1}{1 - 4\rho\sigma_*} \inf_{\tau \in \mathcal{A}_\varepsilon} \mathbf{E}_S \|\hat{S}_\tau - S\|_n^2 \\ &\quad + \frac{2c^* v_\varepsilon B_1(\varrho)}{n^{2/3}} \mathbf{E}_S |\hat{\varsigma}_n - \varsigma_n| + B_n(\varrho, \varepsilon), \end{aligned} \quad (3.9)$$

where $0 < \varrho < 1/4\sigma_*$ is an arbitrary constant and

$$\begin{aligned} B_n(\varrho, \varepsilon) &= \frac{2\sqrt{2(\xi^* - 1)} \sigma_* c^* B_1(\varrho)}{n^{7/6}} v_\varepsilon + \frac{7\sigma_* 2^{\beta_*} B_1(\varrho) + B_2(\varrho)}{n} v_\varepsilon, \\ B_1(\varrho) &= 2 \frac{1 - 2\rho\sigma_* \nu_n}{1 - 4\rho\sigma_* \nu_n}, \quad B_2(\varrho) = 2 \frac{(1 + \xi^*)(1 - 2\rho\sigma_* \nu_n)}{(1 - 4\rho\sigma_* \nu_n)\varrho}, \\ \sigma_* &= \max_{1 \leq l \leq n} \sigma_l^2, \quad c^* = \sup_{\tau \in \mathcal{A}_\varepsilon} c_\tau. \end{aligned} \quad (3.10)$$

To estimate ς_n , we make use of the following estimator

$$\hat{\varsigma}_n = \sum_{j=l_n+1}^n \frac{\hat{\theta}_{j,n}^2}{\|\phi_j\|_n^2}, \quad n \geq 3, \quad (3.11)$$

with $l_n = [n^{1/3} + 1]$. This estimator satisfies the following inequality.

Proposition 3.2.

$$\sup_{\beta \geq 1, r \leq r_*} \sup_{S \in W_r^\beta} \mathbf{E}_S |\hat{\varsigma}_n - \varsigma_n| \leq \frac{1}{\sqrt{n}} T_n^*,$$

where

$$T_n^* = \sqrt{2(\xi^* - 1)\nu_n} \sigma_* + \frac{2r_* + \sigma_*}{n^{1/6}} + \frac{2\sqrt{2r_*\sigma_*\nu_n}}{n^{1/3}} + \frac{5\sigma_*}{\sqrt{n}}.$$

The proof of this proposition is given in Appendix.

Theorem 3.1 and Proposition 3.2 imply immediately the following result.

Theorem 3.3. *Assume that in the model (1.1)–(2.1) the function S belongs to $\cup_{\beta \geq 1, 0 < r \leq r_*} W_r^\beta$, where r_* is an unknown constant. Then, for any $n \geq 1$ and $\varepsilon > 0$, the estimate $S_{n,\varepsilon}^*$ satisfies the following inequality*

$$\mathbf{E}_S \|S_{n,\varepsilon}^* - S\|_n^2 \leq \frac{1}{1 - 4\rho\sigma_*} \inf_{\tau \in \mathcal{A}_\varepsilon} \mathbf{E}_S \|\hat{S}_\tau - S\|_n^2 + \mathcal{D}_n(\varrho, \varepsilon), \quad (3.12)$$

where

$$\mathcal{D}_n(\varrho, \varepsilon) = \frac{2c^* v_\varepsilon B_1(\varrho)}{n^{7/6}} T_n^* + B_n(\varrho, \varepsilon)$$

and $0 < \varrho < 1/4\sigma_*$ is an arbitrary constant.

Remark 3.1. *Note that the principal term in the right-hand side of (3.12) is best in the class of estimators $(\hat{S}_\tau, \tau \in \mathcal{A}_\varepsilon)$. Usually inequalities of such type are called the Oracle inequalities.*

Proof of Theorem 3.1. Fisrt, by (3.6)–(3.7) we get the following equality

$$\|\hat{S}_\tau - S\|_n^2 = J_n(\tau) + 2 \sum_{j=1}^n \frac{\rho_\tau(j)}{\|\phi_j\|_n^2} (\hat{\theta}_{j,n}^2 - \frac{\hat{\varsigma}}{n} - \hat{\theta}_{j,n} \theta_{j,n}) + \sum_{j=1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2}.$$

Further, from the model (3.3) we obtain that

$$\hat{\theta}_{j,n}^2 - \frac{\hat{\varsigma}_n}{n} - \hat{\theta}_{j,n} \theta_{j,n} = \theta_{j,n} \frac{\xi_{j,n}}{\sqrt{n}} + \frac{\tilde{\xi}_{j,n}}{n} + \frac{\tilde{m}_{j,n}}{n} + \frac{1}{n} (\varsigma_n - \hat{\varsigma}_n),$$

where

$$\begin{aligned} \tilde{\xi}_{j,n} &= \xi_{j,n}^2 - m_{j,n}, \quad m_{j,n} = \mathbf{E} \xi_{j,n}^2 = \frac{1}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(x_l), \\ \tilde{m}_{j,n} &= m_{j,n} - \varsigma_n = \frac{1}{n} \sum_{l=1}^n \sigma_l^2 \bar{\phi}_j(x_l), \\ \bar{\phi}_j(x_l) &= \phi_j^2(x_l) - 1. \end{aligned} \quad (3.13)$$

Therefore, for any $\tau \in \mathcal{A}_\varepsilon$,

$$\begin{aligned} \|\hat{S}_\tau - S\|_n^2 &= J_n(\tau) + \frac{2}{\sqrt{n}} \sum_{j=1}^n \rho_\tau(j) \frac{\theta_{j,n}}{\|\phi_j\|_n^2} \xi_{j,n} \\ &\quad + 2N_n(\tau) + \sum_{j=1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2}, \end{aligned} \quad (3.14)$$

where $N_n(\tau) = \sum_{i=1}^3 N_{i,n}(\tau)$ with

$$\begin{aligned} N_{1,n}(\tau) &= n^{-1} \sum_{j=1}^n \frac{\rho_\tau(j)}{\|\phi_j\|_n^2} \tilde{\xi}_{j,n}, \\ N_{2,n}(\tau) &= n^{-1} \sum_{j=1}^n \frac{\rho_\tau(j)}{\|\phi_j\|_n^2} \tilde{m}_{j,n}, \\ N_{3,n}(\tau) &= n^{-1} \sum_{j=1}^n \frac{\rho_\tau(j)}{\|\phi_j\|_n^2} (\varsigma_n - \hat{\varsigma}_n). \end{aligned}$$

Let τ_1 be some fixed value of parameter from \mathcal{A}_ε . Then by (3.14) we find

$$\|\hat{S}_{\hat{\tau}} - S\|_n^2 - \|\hat{S}_{\tau_1} - S\|_n^2 = J_n(\hat{\tau}) - J_n(\tau_1) + 2Z_{1,n}(\hat{\tau}) + 2V_n(\hat{\tau}), \quad (3.15)$$

where

$$\begin{aligned} Z_{1,n}(\tau) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{\rho_\tau(j) - \rho_{\tau_1}(j)}{\|\phi_j\|_n^2} \theta_{j,n} \xi_{j,n}, \\ V_n(\tau) &= N_n(\tau) - N_n(\tau_1). \end{aligned}$$

Lemma 6.4 implies that

$$\mathbf{E} Z_{1,n}^4(\tau) \leq \xi^* d_{1,\tau}^4 \quad (3.16)$$

where

$$\begin{aligned} d_{1,\tau}^2 &= \mathbf{E} Z_{1,n}^2(\tau) \leq \frac{\sigma_* \nu_n}{n} \sum_{j=1}^n \frac{(\rho_\tau(j) - \rho_{\tau_1}(j))^2}{\|\phi_j\|_n^4} \theta_{j,n}^2 \\ &\leq \frac{\sigma_* \nu_n}{n} \|S_\tau - S_{\tau_1}\|_n^2, \end{aligned}$$

ν_n is defined by (3.2).

Taking into account in (3.15) that $J_n(\hat{\tau}) - J_n(\tau_1) \leq 0$, we obtain

$$\|\hat{S}_{\hat{\tau}} - S\|_n^2 - \|\hat{S}_{\tau_1} - S\|_n^2 \leq 2Z_{1,n}(\hat{\tau}) + 2V_n^*,$$

where $V_n^* = \max_{\tau \in \mathcal{A}_\varepsilon} |V_n(\tau)|$.

We set, for $x > 0$, $\varrho > 0$ and fixed $\tau \in \mathcal{A}_\varepsilon$,

$$\tilde{Z}_{1,n}(\tau, x) = \frac{2 Z_{1,n}(\tau) \varrho n}{(\varrho n)^2 d_{1,\tau}^2 + x}$$

and $\Gamma_{1,n}(x) = \left\{ \sup_{\tau} |\tilde{Z}_{1,n}(\tau, x)| \leq 1 \right\}$. Notice now that $(\varrho n)^2 d_{1,\tau}^2 + x \geq 2d_{1,\tau} \sqrt{x} \varrho n$. Therefore by Chebyshev's inequality and (3.16), we obtain, for any $x > 0$ and $\tau \in \mathcal{A}_\varepsilon$,

$$\mathbf{P} \left(|\tilde{Z}_{1,n}(\tau, x)| > 1 \right) \leq \mathbf{P} \left(|Z_{1,n}(\tau)| > \sqrt{x} d_{1,\tau} \right) \leq e^*(x),$$

where $e^*(x) = \mathbf{1}_{\{x \leq 1\}} + \xi^* x^{-2} \mathbf{1}_{\{x > 1\}}$. Hence

$$\mathbf{P} \left(\Gamma_{1,n}^c(x) \right) \leq \sum_{\tau \neq \tau_1} \mathbf{P} \left(|Z_{1,n}(\tau, x)| > \sqrt{x} d_{1,\tau} \right) \leq v_\varepsilon e^*(x). \quad (3.17)$$

Thus on the set $\Gamma_{1,n}(x)$

$$\begin{aligned} \|\hat{S}_{\hat{\tau}} - S\|_n^2 &\leq \|\hat{S}_{\tau_1} - S\|_n^2 + \varrho n d_{1,\hat{\tau}}^2 + \frac{x}{\varrho n} + 2V_n^* \\ &\leq \|\hat{S}_{\tau_1} - S\|_n^2 + \varrho \sigma_* \nu_n \|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 + \frac{x}{\varrho n} + 2V_n^*. \end{aligned} \quad (3.18)$$

Let us estimate now $\|S_{\hat{\tau}} - S_{\tau_1}\|_n^2$. We have

$$\begin{aligned} \|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 &= \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 + (\|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 - \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2) \\ &= \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 + \sum_{j=1}^n \frac{(\rho_{\hat{\tau}}(j) - \rho_{\tau_1}(j))^2}{\|\phi_j\|_n^2} (\theta_{j,n}^2 - \hat{\theta}_{j,n}^2) \\ &\leq \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 - 2 Z_{2,n}(\hat{\tau}), \end{aligned}$$

where

$$Z_{2,n}(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{(\rho_\tau(j) - \rho_{\tau_1}(j))^2}{\|\phi_j\|_n^2} \theta_{j,n} \xi_{j,n}$$

with

$$\begin{aligned} d_{2,\tau}^2 &= \mathbf{E} Z_{2,n}^2(\tau) \leq \frac{\sigma_* \nu_n}{n} \sum_{j=1}^n \frac{(\rho_\tau(j) - \rho_{\tau_1}(j))^4}{\|\phi_j\|_n^4} \theta_{j,n}^2 \\ &\leq \frac{2\sigma_* \nu_n}{n} \|S_\tau - S_{\tau_1}\|_n^2. \end{aligned}$$

By the same way as before, for $x > 0$, we set

$$\Gamma_{2,n}(x) = \{\sup_{\tau} |\tilde{Z}_{2,n}(\tau)| \leq 1\},$$

where

$$\tilde{Z}_{2,n}(\tau, x) = \frac{2Z_{2,n}(\tau)\varrho n}{(\varrho n)^2 d_{2,\tau}^2 + x}.$$

Similarly to (3.17), one shows that

$$\mathbf{P}(\Gamma_{2,n}^c(x)) \leq v_{\varepsilon} e^*(x).$$

On the set $\Gamma_{2,n}(x)$ we obtain that

$$\begin{aligned} \|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 &\leq \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 + \varrho n d_{2,\hat{\tau}}^2 + \frac{x}{\varrho n} \\ &\leq \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 + 2\varrho\sigma_*\nu_n \|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 + \frac{x}{\varrho n}. \end{aligned}$$

Hence

$$\|S_{\hat{\tau}} - S_{\tau_1}\|_n^2 \leq v^*(\varrho) \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 + \frac{v^*(\varrho)}{\varrho n} x, \quad (3.19)$$

where $v^*(\varrho) = (1 - 2\varrho\sigma_*\nu_n)^{-1}$. Thus on the set $\Gamma_n(x) = \Gamma_{1,n}(x) \cap \Gamma_{2,n}(x)$, in view of (3.18),(3.19), we get

$$\begin{aligned} \|\hat{S}_{\hat{\tau}} - S\|_n^2 &\leq \|\hat{S}_{\tau_1} - S\|_n^2 + \varrho\sigma_*\nu_n v^*(\varrho) \|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2 \\ &\quad + \frac{1 + \varrho\sigma_*\nu_n v^*(\varrho)}{\varrho n} x + 2V_n^*. \end{aligned}$$

Estimating here the term $\|\hat{S}_{\hat{\tau}} - \hat{S}_{\tau_1}\|_n^2$ by $2\|\hat{S}_{\hat{\tau}} - S\|_n^2 + 2\|S - \hat{S}_{\tau_1}\|_n^2$ yields the inequality

$$\begin{aligned} \|\hat{S}_{\hat{\tau}} - S\|_n^2 &\leq \frac{1}{1 - 4\varrho\sigma_*\nu_n} \|\hat{S}_{\tau_1} - S\|_n^2 + 2\frac{1 - 2\varrho\sigma_*\nu_n}{1 - 4\varrho\sigma_*\nu_n} V_n^* \\ &\quad + \frac{1 - \varrho\sigma_*\nu_n}{(1 - 4\varrho\sigma_*\nu_n)\varrho n} x. \end{aligned}$$

Applying Lemma 6.5 with $M(x) = 2e^*(x)v_{\varepsilon}$ gives

$$\begin{aligned} \mathbf{E}_S \|\hat{S}_{\hat{\tau}} - S\|_n^2 &\leq \frac{1}{1 - 4\varrho\sigma_*\nu_n} \mathbf{E}_S \|\hat{S}_{\tau_1} - S\|_n^2 \\ &\quad + B_1(\varrho) \mathbf{E}_S V_n^* + v_{\varepsilon} B_2(\varrho) \frac{1}{n}, \end{aligned} \quad (3.20)$$

where $B_1(\varrho)$ and $B_2(\varrho)$ are defined by (3.10). In Appendix 6.2 we show that

$$\mathbf{E}_S V_n^* \leq v_\varepsilon \left(\frac{2\sqrt{2(\zeta^* - 1)\sigma_* c^*}}{n^{7/6}} + \frac{7\sigma_* 2^{\beta_*}}{n} + \frac{2c^*}{n^{2/3}} \mathbf{E}_S |\hat{\varsigma}_n - \varsigma_n| \right). \quad (3.21)$$

Therefore the inequality (3.9) follows immediately from (3.20) and (3.21). \square

4 Asymptotic upper bound

In this section we prove Theorem 2.1.

We start with the estimation problem (1.1) under the condition that $S \in W_r^q$ with the known parameters q , r and $\varsigma(S)$ defined in (2.4). In this case we use the estimator $\tilde{S}_n = \hat{S}_{\tau_\varepsilon}$ defined in (3.5) with $\tau_\varepsilon = (q, r_\varepsilon)$ from \mathcal{A}_ε ,

$$r_\varepsilon = \inf\{i \geq 1 : i\varepsilon \geq r/\varsigma(S)\}$$

and $\varepsilon = \varepsilon_n = 1/\ln(n+1)$.

Theorem 4.1. *Under the condition \mathbf{H}_1)*

$$\limsup_{n \rightarrow \infty} \varphi_n \sup_{S \in W_r^q} \frac{1}{\gamma_q(S)} \mathcal{R}(\tilde{S}_n, S) \leq 1, \quad (4.1)$$

where $\varphi_n = n^{2q/(2q+1)}$.

Proof. First, note that from (3.5), taking into account that $\|\phi_j\|_n \geq 1$, we obtain that

$$\mathbf{E}_S \|\tilde{S}_n - S\|_n^2 \leq \sum_{j=1}^n (1 - \tilde{\rho}_j)^2 \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2} + \frac{1}{n} \sum_{j=1}^n \tilde{\rho}_j^2 m_{j,n},$$

where $\tilde{\rho}_j = \rho_{\tau_\varepsilon}(j)$ and $m_{j,n}$ is defined in (3.13). Moreover, putting $\tilde{k}_0 = [k_0(\tau_\varepsilon)]$ and $\tilde{k}_1 = [\omega_{\tau_\varepsilon} \cdot \ln(n+1)]$, we estimate the risk $\mathcal{R}(\tilde{S}_n, S)$ as

$$\mathcal{R}(\tilde{S}_n, S) \leq \sum_{j=\tilde{k}_0}^{\tilde{k}_1-1} (1 - \tilde{\rho}_j)^2 \theta_{j,n}^2 + \varsigma_n \frac{1}{n} \sum_{j=1}^n \tilde{\rho}_j^2 + \Delta_1(n) + \Delta_2(n)$$

with

$$\Delta_1(n) = \sum_{j=\tilde{k}_1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2}$$

and

$$\Delta_2(n) = \frac{1}{n} \sum_{j=1}^n \tilde{\rho}_j^2 \tilde{m}_{j,n} = \frac{1}{n^2} \sum_{d=1}^n \sigma_d^2 \sum_{j=1}^n \tilde{\rho}_j^2 \bar{\phi}_j(x_d).$$

The last inequality implies that, for any fixed $0 < \delta < 1$,

$$\begin{aligned} \mathcal{R}(\tilde{S}_n, S) &\leq (1 + \delta) \sum_{j=\tilde{k}_0}^{\tilde{k}_1-1} (1 - \tilde{\rho}_j)^2 \theta_j^2 + \varsigma_n \frac{1}{n} \sum_{j=1}^n \tilde{\rho}_j^2 \\ &\quad + \Delta_1(n) + \Delta_2(n) + (1 + 1/\delta) \Delta_3(n), \end{aligned}$$

where $\Delta_3(n) = \sum_{j=\tilde{k}_0}^{\tilde{k}_1-1} (\theta_{j,n} - \theta_j)^2$,

$$\theta_j = \int_0^1 S(x) \phi_j(x) dx. \quad (4.2)$$

Lemmas 6.1–6.3 imply immediately that

$$\lim_{n \rightarrow \infty} \sup_{S \in W_r^q} \varphi_n \sum_{l=1}^3 |\Delta_l(n)| = 0.$$

Therefore denoting $\varphi_n \Delta_l(n)$ as $o(1)$ yields

$$\begin{aligned} \varphi_n \mathcal{R}(\tilde{S}_n, S) &\leq (1 + \delta) \varphi_n \sum_{j=\tilde{k}_0}^{\tilde{k}_1-1} (1 - \tilde{\rho}_j)^2 \theta_j^2 \\ &\quad + \varsigma_n n^{-1/(2q+1)} \sum_{j=1}^n \tilde{\rho}_j^2 + o(1). \end{aligned}$$

Further, for any $S \in W_r^q$ denoting $a_j = \sum_{i=0}^q (2\pi[j/2])^{2i}$, we have that

$$\sum_{j \geq 1} a_j \theta_j^2 \leq r. \quad (4.3)$$

Therefore, in view of (3.4)

$$\limsup_{n \rightarrow \infty} \sup_{j \geq \tilde{k}_0} \varphi_n \frac{(1 - \tilde{\rho}_j)^2}{a_j} \leq \frac{1}{\pi^{2q} c_0^{2q}},$$

where $c_0 = c_{\tau_0}$ with $\tau_0 = (q, r/\varsigma(S))$. From here, by taking into account the conditions (2.4)–(2.5), we obtain, for sufficiently large n , the following upper bound for the quadratic risk

$$\begin{aligned} \varphi_n \mathcal{R}(\tilde{S}_n, S) &\leq (1 + 2\delta) \left(\frac{r}{\pi^{2q} c_0^{2q}} + \varsigma(S) c_0 \int_0^1 \Psi_q^2(z) dz \right) + o(1) \\ &= (1 + 2\delta) \gamma_q(S) + o(1). \end{aligned}$$

This implies the inequality (4.1). Hence Theorem 4.1. \square

Now Theorem 3.1 and Theorem 4.1 imply immediately Theorem 2.1. \square

5 Asymptotic lower bound

In this section we prove Theorem 2.2. For this we need the next auxiliary result.

Lemma 5.1. *For any $0 < \varepsilon < 1$ and any estimate \hat{S}_n of S ,*

$$\|\hat{S}_n - S\|_n^2 \geq (1 - \varepsilon)\|T(\hat{S})_n - S\|^2 - (\varepsilon^{-1} - 1)r/n^2,$$

where $T(\hat{S})_n(x) = \sum_{k=1}^n \hat{S}_n(x_k) \mathbf{1}_{(x_{k-1}, x_k]}(x)$.

Proof of this Lemma is given in Appendix 6.5.

From this Lemma we deduce that to prove (2.9), it suffices to show that

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n} \varphi_n \mathcal{R}_0(\hat{S}_n) \geq 1, \quad (6.1)$$

where

$$\mathcal{R}_0(\hat{S}_n) = \sup_{S \in W_r^q} \mathbf{E}_S \frac{\|\hat{S}_n - S\|^2}{\gamma_q(S)}.$$

For $\eta > 0$ and $x \in \mathbb{R}$, denote

$$I_\eta(x) = \eta^{-1} \int_{\mathbb{R}} \mathbf{1}_{(|u| \leq 1 - \eta)} G\left(\frac{u - x}{\eta}\right) du,$$

where $\mathbf{1}_A$ is the indicator of a set A , the kernel $G \in \mathbf{C}^\infty(\mathbb{R})$ is such that

$$G(-u) = G(u) \text{ for } |u| \leq 1, \quad G(u) = 0 \text{ for } |u| \geq 1, \quad \int_{-1}^1 G(u) du = 1.$$

It is easy to see that the function $I_\eta(x)$ possesses the properties : I_η is symmetric,

$$I_\eta(x) = \begin{cases} 1 & \text{for } |x| \leq 1 - 2\eta; \\ 0 & \text{for } |x| \geq 1; \end{cases}$$

and

$$\begin{aligned} \int_{\mathbb{R}} f(x) I_\eta(x) dx &\rightarrow \int_{-1}^1 f(x) dx, \quad \eta \rightarrow 0, \\ \int_{\mathbb{R}} f(x) I_\eta^2(x) dx &\rightarrow \int_{-1}^1 f(x) dx, \quad \eta \rightarrow 0, \end{aligned} \quad (6.2)$$

uniformly over all functions bounded by any fixed constant $c > 0$, i.e. $\sup_f \sup_{-1 \leq x \leq 1} |f(x)| \leq c$.

We choose the trigonometric basis in $L_2[-1, 1]$ as follows

$$\begin{aligned} e_1(x) &= 1/\sqrt{2}, \quad e_2(x) = \cos(\pi x), \quad e_3(x) = \sin(\pi x), \dots, \\ e_{2i}(x) &= \cos(i\pi x), \quad e_{2i+1}(x) = \sin(i\pi x), \dots \end{aligned} \quad (6.3)$$

For any array $z = \{z_{m,j}, 1 \leq m \leq M, 1 \leq j \leq N\}$ with $M = [1/(2h)] - 1$ and N a positive integer, we denote

$$S_z(x) = \sum_{m=1}^M \sum_{j=1}^N z_{m,j} D_{m,j}(x), \quad (6.4)$$

where

$$\begin{aligned} D_{m,j}(x) &= e_j(v_m(x)) I_\eta(v_m(x)), \quad v_m(x) = (x - a_m)/h, \\ a_m &= 2mh, \quad h = L_n(n)^{-1/(2q+1)}, \quad L_n = \ln(n+1). \end{aligned}$$

To construct a priori distribution on the family of arrays, we choose the following random array $\vartheta = \{\vartheta_{m,j}, 1 \leq m \leq M, 1 \leq j \leq N\}$ with

$$\vartheta_{m,j} = \delta_{m,j} \hat{\xi}_{m,j}, \quad \delta_{m,j} = \frac{g_0(a_m) \kappa_j}{\sqrt{nh}}, \quad (6.5)$$

where

$$\begin{aligned} \kappa_j^2 &= \left(\left(\frac{N}{j} \right)^q - 1 \right)_+, \quad N = [L_n N_\varepsilon^*], \quad g_0(\cdot) = g(\cdot, 0), \\ N_\varepsilon^* &= 2 \left(\frac{\tilde{r}_\varepsilon (q+1)(2q+1)}{q\pi^{2q}} \right)^{1/(2q+1)}, \quad \tilde{r}_\varepsilon = \frac{(1-\varepsilon)r}{\varsigma(0)}, \end{aligned} \quad (6.6)$$

ε is a small positive real, $0 < \varepsilon < 1$, $\varsigma(0) = \int_0^1 g^2(x, 0) dx$, $(\hat{\xi}_{m,j})$ are i.i.d. bounded random variables with a density $\rho_d(x)$ such that

$$\begin{aligned} \mathbf{E} \hat{\xi}_{m,j} &= 0, \quad \mathbf{E} \hat{\xi}_{m,j}^2 = 1, \quad |\hat{\xi}_{m,j}| \leq d, \\ i_d &= \int_{-d}^d \frac{(\dot{\rho}_d(x))^2}{\rho_d(x)} dx \rightarrow 1 \quad \text{as } d \rightarrow \infty. \end{aligned} \quad (6.7)$$

The construction of a such density ρ_d is given in Appendix 6.6.

For any estimator \hat{S}_n , we denote by \hat{S}_n^0 its projection on W_r^q , i.e. $\hat{S}_n^0 = \Pr_{W_r^q}(\hat{S}_n)$. Since W_r^q is convex set we get that $\|\hat{S}_n - S\|^2 \geq \|\hat{S}_n^0 - S\|^2$. Therefore, for sufficiently large n ,

$$\begin{aligned} \mathcal{R}_0(\hat{S}_n) &= \sup_{S \in W_r^q} \mathbf{E}_S \frac{\|\hat{S}_n - S\|^2}{\gamma_q(S)} \geq \sup_{S \in W_r^q} \mathbf{E}_S \frac{\|\hat{S}_n^0 - S\|^2}{\gamma_q(S)} \\ &\geq \int_{\{z: S_z \in W_r^q\}} \mathbf{E}_z \frac{\|\hat{S}_n - S_z\|^2}{\gamma_q(S_z)} \mu_\vartheta(dz) \\ &\geq \frac{1}{\gamma_q(\varepsilon)} \int_{\{z: S_z \in W_r^q\}} \mathbf{E}_z \|\hat{S}_n^0 - S_z\|^2 \mu_\vartheta(dz), \end{aligned}$$

where $\gamma_q^*(\varepsilon) = \sup_{\|S\| \leq \varepsilon} \gamma_q(S)$ (the last inequality is true since $\|S_\vartheta\| \rightarrow 0$ as $n \rightarrow \infty$). Here we denote by \mathbf{E}_z the expectation with respect to the distribution of the process (1.1) with $S = S_z$ and the measure μ_ϑ denotes the distribution of ϑ in \mathbb{R}^ι ($\iota = MN$) which is a priori distribution for the Bayes risk. The last inequality implies that

$$\mathcal{R}_0(\hat{S}_n^0) \geq \frac{1}{\gamma_q^*(\varepsilon)} \tilde{\mathcal{R}}_0(\hat{S}_n^0) - 2 \frac{1}{\gamma_q^*(\varepsilon)} \omega_n, \quad (6.8)$$

where

$$\begin{aligned} \tilde{\mathcal{R}}_0(\hat{S}_n^0) &= \int \mathbf{E}_z \|\hat{S}_n^0 - S_z\|^2 \mu_\vartheta(dz), \\ \omega_n &= \int_{\{z: S_z \notin W_r^q\}} (r + \|S_z\|^2) \mu_\vartheta(dz). \end{aligned}$$

To reduce the nonparametric problem to a parametric one, we replace the functions \hat{S}_n^0 and S by their Fourier series with respect to the basis $(\tilde{e}_{m,i})$, where

$$\tilde{e}_{m,i}(x) = \frac{1}{\sqrt{h}} e_i(v_m(x)) I(v_m(x)), \quad 1 \leq m \leq M, \quad 1 \leq i \leq N$$

and $I(x) = \mathbf{1}_{(|x| \leq 1)}$. We estimate the term $\|\hat{S}_n^0 - S_z\|^2$ as

$$\|\hat{S}_n^0 - S_z\|^2 \geq \sum_{m=1}^M \sum_{j=1}^N (\hat{\lambda}_{m,j} - \lambda_{m,j}(z))^2.$$

Here $\hat{\lambda}_{m,j}$ and $\lambda_{m,j}$ mean the Fourier coefficients for the functions \hat{S}_n^0 and S , respectively, i.e.

$$\hat{\lambda}_{m,j} = \int_0^1 \hat{S}_n^0(x) \tilde{e}_{m,j}(x) dx \quad \text{and} \quad \lambda_{m,j}(z) = \int_0^1 S_z(x) \tilde{e}_{m,j}(x) dx.$$

Due to the definition of the function $\lambda_{m,j}(z)$ we obtain that

$$\begin{aligned}\lambda_{m,j}(z) &= \sum_{l=1}^M \sum_{i=1}^N z_{l,i} \int_0^1 D_{l,i}(x) \tilde{e}_{m,j}(x) dx \\ &= \sum_{l=1}^M \sum_{i=1}^N \frac{z_{l,i}}{\sqrt{h}} \int_0^1 e_i(v_l(x)) e_j(v_m(x)) I_\eta(v_l(x)) I(v_m(x)) dx \\ &= \sqrt{h} \sum_{i=1}^N z_{m,i} \int_{-1}^1 e_i(u) e_j(u) I_\eta(u) du.\end{aligned}$$

Therefore

$$\Lambda_{m,j} = \frac{\partial \lambda_{m,j}}{\partial z_{m,j}} = \sqrt{h} \int_{-1}^1 e_j^2(v) I_\eta(v) dv = \sqrt{h} (e_j^2, I_\eta).$$

By the van Trees inequality (see Appendix 6.7) one gets

$$\begin{aligned}\tilde{\mathcal{R}}_0(S_n^0) &\geq \sum_{m=1}^M \sum_{j=1}^N \frac{\Lambda_{m,j}^2}{A_{m,j} + B_{m,j} + \mathcal{I}_{m,j}} \\ &= \sum_{m=1}^M \sum_{j=1}^N \frac{h(e_j^2, I_\eta)^2}{A_{m,j} + B_{m,j} + (\delta_{m,j})^{-2} i_d},\end{aligned}\quad (6.9)$$

where $\mathcal{I}_{m,j}$ is the Fisher information relative the random variable $\xi_{m,j}$,

$$\begin{aligned}A_{m,j} &= \sum_{k=1}^n \int_{\mathbb{R}^\ell} \frac{1}{g^2(x_k, S_z)} \left(\frac{\partial S_z(x_k)}{\partial z_{m,j}} \right)^2 \mu_\vartheta(dz), \\ B_{m,j} &= \sum_{k=1}^n \int_{\mathbb{R}^\ell} \left(\frac{\tilde{L}_{m,j}(x_k, S_z)}{g(x_k, S_z)} \right)^2 \mu_\vartheta(dz), \quad \tilde{L}_{m,j}(x, S_z) = L_{x, S_z} \left(\frac{\partial S_z}{\partial z_{m,j}} \right).\end{aligned}$$

First note that from (6.4) we obtain that

$$\begin{aligned}A_{m,j} &= \sum_{k=1}^n D_{m,j}^2(x_k) \int_{\mathbb{R}^\ell} \frac{1}{g^2(x_k, S_z)} \mu_\vartheta(dz) = (1 + o(1)) \sum_{k=1}^n \frac{1}{g_0^2(x_k)} D_{m,j}^2(x_k) \\ &= n h (1 + o(1)) \frac{1}{g_0^2(a_m)} \int_{-1}^1 e_j^2(x) I_\eta^2(x) dx = n h (1 + o(1)) \frac{1}{g_0^2(a_m)} (e_j^2, I_\eta^2).\end{aligned}$$

Now we study the behaviour of $B_{m,j}$. Due to the inequality (2.6) and the definition (6.4), we obtain that

$$|\tilde{L}_{m,j}(x, S_z)| \leq C^* \left\| \frac{\partial S_z}{\partial z_{m,j}} \right\| = C^* \|D_{m,j}\| \leq C^* h.$$

Therefore by the condition \mathbf{H}_3) we get

$$|B_{m,j}| \leq C^* \frac{1}{g_*^2} n h^2 .$$

Taking into account this inequality, the inequality (6.9) and the definition of $\delta_{m,j}$ in (6.5), we evaluate the Bayes risk as

$$\begin{aligned} \frac{1}{\gamma_q^*(\varepsilon)} \tilde{\mathcal{R}}_0(\hat{S}_n^0) &\geq \frac{1}{n \gamma_q^*(\varepsilon)} \sum_{m=1}^M g_0^2(a_m) \sum_{j=1}^N \frac{\alpha_j(\eta)}{(1 + o(1)) \beta_j(\eta) + (\kappa_j)^{-2} i_d} \\ &\geq \frac{(1 + o(1))}{2nh \gamma_q^*(\varepsilon)} \varsigma(0) \sum_{j=1}^N \frac{\alpha_j(\eta)}{(1 + o(1)) \beta_j(\eta) + (\kappa_j)^{-2} i_d} \\ &= \frac{(1 + o(1)) \varsigma(0)}{2 \gamma_q^*(\varepsilon) n^{2q/(2q+1)}} \frac{1}{L_n} \sum_{j=1}^N \frac{\alpha_j(\eta)}{(1 + o(1)) \beta_j(\eta) + (\kappa_j)^{-2} i_d} , \end{aligned}$$

where we denote $\alpha_j(\eta) = (e_j^2, I_\eta)^2$ and $\beta_j(\eta) = (e_j^2, I_\eta^2)$.

In Appendix 6.9 we show that

$$\lim_{n \rightarrow \infty} \varphi_n \omega_n = 0 . \quad (6.10)$$

This result and the previous inequality imply

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\hat{S}_n} \varphi_n \mathcal{R}_0(\hat{S}_n) &\geq \frac{\varsigma(0) N_\varepsilon^*}{2 \gamma_q^*(\varepsilon)} \Omega_*(\eta, d) \liminf_{n \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \frac{\kappa_j^2}{\kappa_j^2 + 1} \\ &= \frac{\varsigma(0) N_\varepsilon^*}{2 \gamma_q^*(\varepsilon)} \Omega_*(\eta, d) \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (1 - (j/N)^q) \\ &= \frac{\varsigma(0) N_\varepsilon^*}{2 \gamma_q^*(\varepsilon)} \frac{q}{q+1} \Omega_*(\eta, d) \\ &= (1 - \varepsilon)^{1/(2q+1)} \frac{\gamma_q^0}{\gamma_q^*(\varepsilon)} \Omega_*(\eta, d) , \end{aligned} \quad (6.11)$$

where $\gamma_q^0 = \gamma_q(0)$, $\Omega_*(\eta, d) = \liminf_{N \rightarrow \infty} \Omega_N(\eta, d)$ and

$$\Omega_N(\eta, d) = \frac{\sum_{j=1}^N (\alpha_j(\eta) \kappa_j^2) (\beta_j(\eta) \kappa_j^2 + i_d)^{-1}}{\sum_{j=1}^N \kappa_j^2 (\kappa_j^2 + 1)^{-1}} . \quad (6.12)$$

In Appendix 6.8 we show that

$$\lim_{\eta \rightarrow 0} \lim_{d \rightarrow \infty} \Omega_*(\eta, d) = 1 . \quad (6.13)$$

The condition \mathbf{H}_1) implies $\gamma_q^*(\varepsilon) \rightarrow \gamma_q^0$ as $\varepsilon \rightarrow 0$. Therefore limiting $\varepsilon \rightarrow 0, \eta \rightarrow 0, d \rightarrow \infty$ in (6.11) yields

$$\liminf_{n \rightarrow \infty} \inf_{\hat{S}_n} \varphi_n \tilde{\mathcal{R}}_0(\hat{S}_n) \geq 1.$$

Thus (6.8) implies the inequality (6.3) from which it follows (6.1) through Lemma 5.1. \square

6 Appendix

6.1 Properties of trigonometric basis

Lemma 6.1. *Let $\theta_j, \theta_{j,n}$ be coefficients from (3.3) and (4.2), respectively. Then for $1 \leq j \leq n$ and $n \geq 2$*

$$\sup_{S \in W_r^q} |\theta_{j,n} - \theta_j| \leq 2\pi \sqrt{r} \frac{j}{n}. \quad (\text{A.1})$$

Proof. Indeed, we have

$$\begin{aligned} |\theta_{j,n} - \theta_j| &= \left| \sum_{l=1}^n \int_{x_{l-1}}^{x_l} (S(x_l)\phi_j(x_l) - S(x)\phi_j(x)) dx \right| \\ &\leq \frac{1}{n} \sum_{l=1}^n \int_{x_{l-1}}^{x_l} (|\dot{S}(z)\phi_j(z)| + |S(z)\dot{\phi}_j(z)|) dz \\ &= \frac{1}{n} \int_0^1 (|\dot{S}(z)| |\phi_j(z)| + |S(z)| |\dot{\phi}_j(z)|) dz. \end{aligned}$$

By making use of the Bounyakovskii-Cauchy-Schwartz inequality we get

$$\begin{aligned} |\theta_{j,n} - \theta_j| &\leq \frac{1}{n} \left(\|\dot{S}\| \|\phi\| + \|\dot{\phi}\| \|S\| \right) \\ &\leq \frac{1}{n} \left(\|\dot{S}\| + \pi j \|S\| \right). \end{aligned}$$

The definition of class W_r^q in (2.2) implies (A.1). Hence Lemma 6.1. \square

Lemma 6.2. *For any $m \geq 0$,*

$$\sup_N \sup_{x \in [0,1]} N^{-m} \left| \sum_{l=2}^N l^m (\phi_l^2(x) - 1) \right| \leq 2^m. \quad (\text{A.2})$$

Proof. By the properties of the trigonometric functions we get

$$\begin{aligned}
\sum_{l=2}^N l^m (\phi_l^2(x) - 1) &= \sum_{1 \leq l \leq N/2} (2l)^m [2 \cos^2(2\pi lx) - 1] \\
&+ \sum_{1 \leq l \leq (N-1)/2} (2l+1)^m (2 \sin^2(2\pi lx) - 1) \\
&= \sum_{1 \leq l \leq N/2} (2l)^m \cos(4\pi lx) \\
&- \sum_{1 \leq l \leq (N-1)/2} (2l+1)^m \cos(4\pi lx).
\end{aligned}$$

This yields

$$\begin{aligned}
\left| \sum_{l=2}^N l^m (\phi_l^2(x) - 1) \right| &\leq \left| \sum_{1 \leq l \leq (N-1)/2} ((2l+1)^m - (2l)^m) \cos(4\pi lx) \right| + N^m \\
&\leq \sum_{1 \leq l \leq (N-1)/2} |(2l+1)^m - (2l)^m| + N^m \\
&= \sum_{1 \leq l \leq (N-1)/2} \sum_{j=0}^{m-1} \binom{m}{j} (2l)^j + N^m.
\end{aligned}$$

This implies (A.2). \square

Lemma 6.3. For any function $S \in W_r^q$,

$$\sup_{n \geq 1} \sup_{1 \leq m \leq n-1} m^{2q} \left(\sum_{j=m+1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2} \right) \leq \frac{2r}{\pi^{2(q-1)}}, \quad (\text{A.3})$$

where ν_n is defined by (3.2).

Proof. First, note that any function S from W_r^q we can be represented by its Fourier series, i.e. $S = \sum_{j=1}^{\infty} \theta_j \phi_j$ with the coefficients defined by (4.2). By denoting the residual term for S as

$$\Delta_m(x) = S - \sum_{j=1}^m \theta_j \phi_j = \sum_{j=m+1}^{\infty} \theta_j \phi_j(x),$$

we obtain that

$$\sum_{j=m+1}^n \frac{\theta_{j,n}^2}{\|\phi_j\|_n^2} = \inf_{\alpha_1, \dots, \alpha_m} \|S - \sum_{j=1}^m \alpha_j \phi_j\|_n^2 \leq \|\Delta_m\|_n^2.$$

Moreover

$$\begin{aligned}\|\Delta_m\|_n^2 &= \frac{1}{n} \sum_{k=1}^n \Delta_m^2(x_k) = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} \Delta_m^2(x_k) dx \\ &\leq 2 \int_0^1 \Delta_m^2(x) dx + 2 \sum_{k=1}^n \int_{x_{k-1}}^{x_k} (\Delta_m(x_k) - \Delta_m(x))^2 dx.\end{aligned}$$

The last term in this inequality we estimate as

$$\begin{aligned}(\Delta_m(x_k) - \Delta_m(x))^2 &= \left(\int_x^{x_k} \dot{\Delta}_m(z) dz \right)^2 \\ &\leq \frac{1}{n} \int_{x_{k-1}}^{x_k} (\dot{\Delta}_m(z))^2 dz.\end{aligned}$$

Therefore

$$\begin{aligned}\|\Delta_m\|_n^2 &\leq 2\|\Delta_m\|^2 + 2\frac{1}{n^2}\|\dot{\Delta}_m\|^2 \\ &= 2 \sum_{j=m+1}^{\infty} \theta_j^2 + 2\frac{1}{n^2} \sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2.\end{aligned}$$

Taking here into account the inequality (4.3) we obtain (A.3). Hence Lemma 6.3. \square

Lemma 6.4. *Let $\xi_{j,n}$ be defined in (3.3) for the model (1.1)–(2.1). Then, for any real numbers v_1, \dots, v_n ,*

$$\begin{aligned}\mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^4 &\leq \xi^* \left(\mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^2 \right)^2, \\ \mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^2 &\leq \sigma_* \nu_n \sum_{j=1}^n v_j^2,\end{aligned}\tag{A.4}$$

where $\sigma_* = \max_{1 \leq j \leq n} \sigma_j^2$ and ν_n is defined in (3.2).

Proof. By the definition of $\xi_{j,n}$ we obtain that $\sum_{j=1}^n v_j \xi_{j,n} = \sum_{l=1}^n \sigma_l \tilde{v}_l \xi_l$ with $\tilde{v}_l = \frac{1}{\sqrt{n}} \sum_{j=1}^n v_j \phi_j(x_l)$. Therefore, this implies immediately the first inequality in (A.4). Moreover

$$\begin{aligned}\mathbf{E} \left(\sum_{j=1}^n v_j \xi_{j,n} \right)^2 &= \sum_{l=1}^n \sigma_l^2 \tilde{v}_l^2 \leq \sigma_* \sum_{l=1}^n \tilde{v}_l^2 \\ &= \sigma_* \sum_{i,j=1}^n v_i v_j (\phi_i, \phi_j)_n.\end{aligned}$$

The orthogonal properties of the basis (ϕ_j) and (3.2) imply the second inequality in (A.4). Hence Lemma 6.4. \square

6.2 Proof of (3.21)

Indeed, we have

$$\begin{aligned} \mathbf{E}_S V_n^*(\tau) &\leq \sum_{i=1}^3 \sum_{\tau \neq \tau_1} \mathbf{E}_S |N_{i,n}(\tau) - N_{i,n}(\tau_1)| \\ &\leq v_\varepsilon \sum_{i=1}^3 \sup_{\tau \in \mathcal{A}_\varepsilon} \mathbf{E}_S |N_{i,n}(\tau) - N_{i,n}(\tau_1)|. \end{aligned} \quad (\text{A.5})$$

We start with the first term in the right-hand part. We have

$$\begin{aligned} \mathbf{E}_S |N_{1,n}(\tau) - N_{1,n}(\tau_1)| &= \frac{1}{n} \mathbf{E}_S \left| \sum_{j=1}^n (\rho_\tau(j) - \rho_{\tau_1}(j)) \tilde{\xi}_{j,n} \right| \\ &\leq \frac{1}{n} \sum_{j=1}^n (\rho_\tau(j) + \rho_{\tau_1}(j)) (\mathbf{E} \tilde{\xi}_{j,n}^2)^{1/2}, \end{aligned}$$

where $\tilde{\xi}_{j,n}$ is defined in (3.13). Moreover,

$$\begin{aligned} \mathbf{E} \tilde{\xi}_{j,n}^2 &= \mathbf{E} \left(\xi_{j,n}^2 - m_{j,n} \right)^2 = \mathbf{E} \xi_{j,n}^4 - m_{j,n}^2 \\ &= \frac{1}{n^2} \sum_{r,s=1}^n \sigma_r^2 \sigma_s^2 \phi_j^2(x_r) \phi_j^2(x_s) \mathbf{E} \xi_r^2 \xi_s^2 - m_{j,n}^2 \\ &= (\xi^* - 1) \frac{1}{n^2} \sum_{r=1}^n \sigma_r^4 \phi_j^4(x_r) \leq \frac{2(\xi^* - 1) \sigma_*^2 \nu_n}{n}. \end{aligned} \quad (\text{A.6})$$

Thus, for any fixed τ and τ_1 from \mathcal{A}_ε ,

$$\mathbf{E}_S |N_{1,n}(\tau) - N_{1,n}(\tau_1)| \leq \frac{\sqrt{2(\xi^* - 1) \nu_n \sigma_*}}{n \sqrt{n}} (w_\tau + w_{\tau_1}),$$

where $w_\tau = n^{1/(2\beta+1)} c_\tau \leq c^* n^{1/3}$ since $\beta \geq 1$. Therefore

$$\sup_{\tau \in \mathcal{A}_\varepsilon} \mathbf{E}_S |N_{1,n}(\tau) - N_{1,n}(\tau_1)| \leq \frac{2\sqrt{2(\xi^* - 1) \nu_n \sigma_*} c^*}{n^{7/6}}. \quad (\text{A.7})$$

To estimate the second term in the right-hand part of (A.5), note that

$$\begin{aligned} N_{2,n}(\tau) &= \frac{1}{n^2} \sum_{d=1}^n \sigma_d^2 \sum_{j=1}^n \rho_\tau(j) \bar{\phi}_j(x_d) \\ &\quad - \frac{1}{n^2} \sum_{d=1}^n \sigma_d^2 \rho_\tau(n) \left(1 - \frac{1}{\nu_n}\right) \bar{\phi}_n(x_d). \end{aligned}$$

Let us show that, for any $n \geq 2$,

$$\sup_x \left| \sum_{j=1}^n \rho_\tau(j) \bar{\phi}_j(x) \right| \leq 3 \cdot 2^\beta. \quad (\text{A.8})$$

Indeed, by definition (3.4), $\rho_\tau(j) = 1$ for $j \leq k_0(\tau)$. Therefore the inequality (A.8) follows immediately from Lemma 6.2, for $n \leq k_0(\tau)$. Let us show (A.8), for $n > k_0(\tau)$. In this case we can represent the sum in the left-hand part of (A.8) as

$$\begin{aligned} \sum_{j=1}^n \rho_\tau(j) \bar{\phi}_j(x) &= \sum_{j=2}^n \rho_\tau(j) \bar{\phi}_j(x) \\ &= \sum_{2 \leq j \leq w_\tau \wedge n} \bar{\phi}_j(x) - \frac{1}{(w_\tau)^\beta} \sum_{2 \leq j \leq w_\tau \wedge n} j^\beta \bar{\phi}_j(x) \\ &\quad + \frac{1}{(w_\tau)^\beta} \sum_{2 \leq j \leq k_0(\tau)} j^\beta \bar{\phi}_j(x), \end{aligned}$$

where $a \wedge b = \min(a, b)$. Now it is easy to see that inequality (A.8) follows directly from Lemma 6.2. Therefore by (A.8)

$$\begin{aligned} \sup_{\tau \in \mathcal{A}_\varepsilon} \mathbf{E} |N_{2,n}(\tau) - N_{2,n}(\tau_1)| &\leq \frac{6 \cdot 2^{\beta_*}}{n^2} \sum_{d=1}^n \sigma_d^2 + \frac{2}{n^2} \sum_{d=1}^n \sigma_d^2 \left(1 - \frac{1}{\nu_n}\right) \\ &\leq \frac{7 \sigma_* 2^{\beta_*}}{n}. \end{aligned}$$

For the last addend in (A.5), in view of (3.4), we get

$$\begin{aligned} \mathbf{E} |N_{3,n}(\tau) - N_{3,n}(\tau_1)| &\leq \frac{w_\tau + w_{\tau_1}}{n} \mathbf{E} |\hat{\varsigma}_n - \varsigma_n| \\ &\leq \frac{c^* (n^{1/(2\beta+1)} + n^{1/(2\beta_1+1)})}{n} \mathbf{E} |\hat{\varsigma}_n - \varsigma_n| \\ &\leq \frac{2c^*}{n^{2/3}} \mathbf{E} |\hat{\varsigma}_n - \varsigma_n|. \end{aligned}$$

□

6.3 Proof of Proposition 3.2

First notice that taking into account (3.13) in (3.11) we obtain that

$$\begin{aligned}\hat{\varsigma}_n - \varsigma_n &= \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \theta_{j,n}^2 + \frac{2}{\sqrt{n}} \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \theta_{j,n} \xi_{j,n} \\ &\quad + \frac{1}{n} \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \tilde{\xi}_{j,n} + \frac{1}{n} \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \tilde{m}_{j,n} - \frac{(l_n+1)\nu_n - 1}{n\nu_n} \varsigma_n.\end{aligned}$$

Therefore by applying here (A.6) and Lemma 6.4 we come to the following upper bound

$$\begin{aligned}\mathbf{E}_S |\hat{\varsigma}_n - \varsigma_n| &\leq \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \theta_{j,n}^2 + \frac{2\sqrt{\sigma_*\nu_n}}{\sqrt{n}} \left(\sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^4} \theta_{j,n}^2 \right)^{1/2} \\ &\quad + \frac{\sqrt{2(\xi^* - 1)\nu_n\sigma_*}}{\sqrt{n}} + \frac{(l_n+1)\sigma_*}{n} + \frac{1}{n} \left| \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \tilde{m}_{j,n} \right|.\end{aligned}$$

Moreover, by Lemma 6.2 we have

$$\begin{aligned}\left| \sum_{j=l_n+1}^n \frac{1}{\|\phi_j\|_n^2} \tilde{m}_{j,n} \right| &= \left| \frac{1}{n} \sum_{d=1}^n \sigma_d^2 \sum_{j=l_n+1}^{n-1} \bar{\phi}_j(x_d) + \frac{1}{\nu_n} \tilde{m}_{n,n} \right| \\ &\leq \sigma_* \sum_{d=1}^n \left| \sum_{j=l_n+1}^{n-1} \bar{\phi}_j(x_d) \right| + \frac{1}{\nu_n} |\tilde{m}_{n,n}| \\ &\leq 3\sigma_*.\end{aligned}$$

This proposition follows now directly from Lemma 6.3 \square

6.4 Mean forecast bound

Lemma 6.5. *Let α, ξ be two integrable random variables. Let β be a positive real number and $\{\Gamma_x, x > 0\}$ be a family of events such that, for any $x > 0$,*

$$\mathbf{P}(\xi > \alpha + \beta x, \Gamma_x) = 0.$$

Assume also that there exists some positive integrable on \mathbb{R}_+ function $M(x)$ for which $\mathbf{P}(\Gamma_x^c) \leq M(x)$, for all $x \in \mathbb{R}_+$. Then

$$\mathbf{E} \xi \leq \mathbf{E} \alpha + \beta M^*,$$

where $M^* = \int_0^\infty M(x) dx$.

Proof. Denote $\eta = \xi - \alpha$. Then

$$\xi = \alpha + \eta \leq \alpha + \eta_+$$

and

$$\mathbf{E} \xi \leq \mathbf{E} \alpha + \mathbf{E} \eta_+.$$

Moreover,

$$\begin{aligned} \mathbf{E} \eta_+ &= \int_0^\infty \mathbf{P}(\eta_+ > z) dz = \beta \int_0^\infty \mathbf{P}(\eta_+ > \beta x) dx \\ &= \beta \int_0^\infty (\mathbf{P}(\eta_+ > \beta x, \Gamma_x) + \mathbf{P}(\eta_+ > \beta x, \Gamma_x^c)) dx \\ &= \beta \int_0^\infty (\mathbf{P}(\eta > \beta x, \Gamma_x) + \mathbf{P}(\eta_+ > \beta x, \Gamma_x^c)) dx \\ &\leq \beta \int_0^\infty \mathbf{P}(\Gamma_x^c) dx \leq \beta \int_0^\infty M(x) dx \leq \beta M^*. \end{aligned}$$

□

6.5 Proof of Lemma 5.1

First notice that, for any $S \in W_r^q$, one has

$$\begin{aligned} \|\hat{S}_n - S\|_n^2 &= \frac{1}{n} \sum_{j=1}^n (\hat{S}_n(x_j) - S(x_j))^2 = \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (\hat{S}_n(x_j) - S(x_j))^2 dx \\ &= \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (\hat{S}_n(x_j) - S(x))^2 dx \\ &\quad + 2 \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (\hat{S}_n(x_j) - S(x))(S(x) - S(x_j)) dx \\ &\quad + \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (S(x) - S(x_j))^2 dx = \|T(\hat{S})_n - S\|^2 + D_n + \Delta_n, \end{aligned}$$

where

$$\begin{aligned} D_n &= 2 \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (\hat{S}_n(x_j) - S(x))(S(x) - S(x_j)) dx, \\ \Delta_n &= \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (S(x) - S(x_j))^2 dx. \end{aligned}$$

For any $0 < \varepsilon < 1$, by making use of the elementary inequality $2ab \leq \varepsilon a^2 + \varepsilon^{-1}b^2$, one gets

$$D_n \leq \varepsilon \|T(\hat{S})_n - S\|^2 + \varepsilon^{-1} \Delta_n.$$

Moreover, for any $S \in W_r^q$ with $q \geq 1$, by the Bounyakovskii-Cauchy-Schwartz inequality we obtain that

$$\Delta_n = \sum_{j=1}^n \int_{x_{j-1}}^{x_j} \left(\int_x^{x_j} \dot{S}(t) dt \right)^2 dx \leq \frac{1}{n} \sum_{j=1}^n \int_{x_{j-1}}^{x_j} (\dot{S}(t))^2 dt = \frac{1}{n^2} \|\dot{S}\|^2 \leq \frac{r}{n^2}.$$

Hence Lemma 5.1. \square

6.6 Construction of the density ρ_d

We set

$$\rho_d(t) = \frac{1}{\varsigma_d} f_d \left(\frac{t}{\varsigma_d} \right), \quad \varsigma_d^2 = \left(\int_{-d}^d z^2 f_d(z) dz \right)^{-1}, \quad (\text{A.9})$$

where f_d is a truncated Gaussian density which is defined by the following way. First we need to choose the smooth indicator function. We choose this function for $d \geq 3$ as

$$\Lambda_d(t) = \int_{\mathbb{R}} \mathbf{1}_{(|z| \leq d-2)} G(z-t) dz,$$

where

$$G(t) = G_* \exp\{-1/(1-t^2)\} I(t), \quad G_* = \left(\int_{-1}^1 \exp\{-1/(1-t^2)\} dt \right)^{-1}$$

and $I(t) = \mathbf{1}_{\{|t| \leq 1\}}$. Notice that this smoothing density G is infinitely times differentiable. Now we set

$$f_d(t) = \frac{f(t)\Lambda_d(t)}{f_d^*}, \quad f(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}} \quad \text{and} \quad f_d^* = \int_{\mathbb{R}} f(t)\Lambda_d(t) dt.$$

The properties of the density ρ_d are given in the following Lemma

Lemma 6.6. *The density ρ_d satisfies the following properties:*

1. For any $d \geq 3$,

$$\int_{\mathbb{R}} t \rho_d(t) dt = 0, \quad \int_{\mathbb{R}} t^2 \rho_d(t) dt = 1.$$

2. The Ficher information i_d for ρ_d goes to 1 as $d \rightarrow \infty$.

Proof. Notice that the first part of this Lemma follows immediately from the definition (A.9). Let us show the second one. Due to properties of Λ_d ,

$$f_d(z) = f(z)\mathbf{1}_{(|z|\leq d-3)} + f(z)\Lambda_d(z)\mathbf{1}_{(|z|\geq d-3)}.$$

Therefore

$$f_d^* = \int_{\mathbb{R}} f(z)\mathbf{1}_{(|z|\leq d-3)} dz + \int_{\mathbb{R}} f(z)\Lambda_d(z)\mathbf{1}_{(|z|\geq d-3)} dz \rightarrow \int_{\mathbb{R}} f(z) dz = 1,$$

since $\sup_{x \in \mathbb{R}} |\Lambda_d(x)| \leq 1$. Similarly $\zeta_d^2 \rightarrow 1$ as $d \rightarrow \infty$.

Consider the Fisher information i_d . We have

$$\begin{aligned} i_d &= \int_{-d}^d \frac{(\dot{\rho}_d)^2(z)}{\rho_d(z)} dz = \frac{1}{f_d^* \zeta_d^2} \int_{-d/\zeta_d}^{d/\zeta_d} \frac{(\dot{f}_d(u))^2}{f_d(u)} du \\ &= \frac{1}{f_d^* \zeta_d^2} \int_{-d/\zeta_d}^{d/\zeta_d} \frac{(\dot{f}(u))^2}{f(u)} \Lambda_d(u) du + 2 \frac{1}{f_d^* \zeta_d^2} \int_{-d/\zeta_d}^{d/\zeta_d} \dot{f}(u) \dot{\Lambda}_d(u) du \\ &\quad + \frac{1}{f_d^* \zeta_d^2} \int_{-d/\zeta_d}^{d/\zeta_d} \frac{(\dot{\Lambda}_d(u))^2}{\Lambda_d(u)} f(u) du. \end{aligned}$$

From the definition of Λ_d it is easy to deduce that, for any integrable on \mathbb{R} function f ,

$$\lim_{d \rightarrow \infty} \int_{\mathbb{R}} f(u) \Lambda_d(u) du = \int_{\mathbb{R}} f(u) du, \quad \lim_{d \rightarrow \infty} \int_{\mathbb{R}} f(u) \dot{\Lambda}_d(u) du = 0$$

and $\dot{\Lambda}_d(u) = 0$ for $u \leq d-3$. Therefore we obtain the following asymptotic relationship for i_d as $d \rightarrow \infty$

$$i_d = 1 + o(1) + \int_{-d/\zeta_d}^{d/\zeta_d} \frac{(\dot{\Lambda}_d(u))^2}{\Lambda_d(u)} f(u) \mathbf{1}_{\{|u|\geq d-3\}} du. \quad (\text{A.10})$$

To estimate the last integral, notice that by the Bounyakovskii-Cauchy-Schwartz inequality

$$\begin{aligned} (\dot{\Lambda}_d(u))^2 &= \left(\int_{\mathbb{R}} \mathbf{1}_{(|z|\leq d-2)} \dot{G}(u-z) dz \right)^2 \\ &\leq \int_{\mathbb{R}} \mathbf{1}_{(|z|\leq d-2)} dz \int_{\mathbb{R}} \mathbf{1}_{(|z|\leq d-2)} \left(\dot{G}(u-z) \right)^2 dz \\ &\leq 2d \int_{\mathbb{R}} \mathbf{1}_{(|z|\leq d-2)} \left(\dot{G}(u-z) \right)^2 dz, \end{aligned}$$

where

$$(\dot{G}(t))^2 = \frac{16t^2}{(1-t^2)^4} G_* e^{-(1-t^2)^{-1}} G(t) \leq c^* G(t),$$

with $c^* = 16G_* \sup_{z \geq 1} z^4 e^{-z} < \infty$. Therefore

$$(\dot{\Lambda}_d(u))^2 \leq 2c^* d \int_{\mathbb{R}} \mathbf{1}_{(|z| \leq d-2)} G(u-z) dz = 2c^* d \Lambda_d(u).$$

From here we estimate the last integral in (A.10) by

$$2c^* d \int_{\mathbb{R}} f(u) \mathbf{1}_{(|u| > d-3)} du \rightarrow 0 \text{ as } d \rightarrow \infty.$$

Hence we get Lemma 6.6. \square

6.7 The van Trees inequality for regression models

Let $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P}_\vartheta, \vartheta \in \Theta \subseteq \mathbb{R}^\iota)$ be a statistical model relative the observations $(y_j)_{1 \leq j \leq n}$ governed by the regression equation

$$y_j = S_\vartheta(x_j) + \sigma_j(\vartheta) \xi_j, \quad (\text{A.11})$$

where ξ_1, \dots, ξ_n are i.i.d. $\mathcal{N}(0, 1)$ random variables, $\vartheta = (\vartheta_1, \dots, \vartheta_\iota)$ is an unknown vector parameter and $\sigma_j(\vartheta) = g(x_j, S_\vartheta)$, where the function $g(x, S)$ is defined in the condition \mathbf{H}_1). Assume that a prior distribution μ of the parameter ϑ in \mathbb{R}^ι is defined by the density $\Phi(\vartheta)$ of the following form

$$\Phi(\vartheta) = \Phi(\vartheta_1, \dots, \vartheta_\iota) = \prod_{j=1}^{\iota} \Phi_j(\vartheta_j),$$

where Φ_j is a continuous density on \mathbb{R} with a compact support such that

$$\mathcal{I}_l = \int_{\mathbb{R}} \frac{(\dot{\Phi}_l(z))^2}{\Phi_l(z)} dz < \infty.$$

Let $\lambda(\cdot)$ be a real-valued function, $\lambda : \mathbb{R}^\iota \rightarrow \mathbb{R}$. We assume that this function is differentiable with respect to ϑ_l , for any $1 \leq l \leq \iota$, such that the partial derivatives are integrable with respect to the density $\Phi(\cdot)$ and we put

$$\Lambda_l = \int_{\mathbb{R}^\iota} \frac{\partial \lambda(\vartheta)}{\partial \vartheta_l} \Phi(\vartheta) d\vartheta.$$

Let now $\hat{\lambda}_n$ be an estimator of $\lambda = \lambda(\vartheta)$ based on observations $(y_j)_{1 \leq j \leq n}$. For any $\mathcal{B}(\mathbb{R}^n \times \mathbb{R}^\iota)$ - measurable integrable function $\zeta = \zeta(x, \vartheta)$, $x \in \mathbb{R}^n$, $\vartheta \in \mathbb{R}^\iota$, we set

$$\tilde{\mathbf{E}} \zeta(Y, \vartheta) = \int_{\mathbb{R}^\iota} \mathbf{E}_\vartheta \zeta(Y, \vartheta) \Phi(\vartheta) d\vartheta,$$

where \mathbf{E}_ϑ is the expectation with respect to the distribution \mathbf{P}_ϑ of the vector $Y = (y_1, \dots, y_n)$. Note that in this case

$$\mathbf{E}_\vartheta \zeta(Y, \vartheta) = \int_{\mathbb{R}^n} \zeta(v, \vartheta) f(v, \vartheta) dv,$$

where

$$f(v, \vartheta) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_j(\vartheta)}} \exp \left\{ -\frac{(v_j - S_\vartheta(x_j))^2}{2\sigma_j^2(\vartheta)} \right\}.$$

We prove the following result (cp. Gill and Levit (1995)).

Lemma 6.7. *Assume that the function $g(\cdot, \cdot)$ satisfies the condition \mathbf{H}_1) – \mathbf{H}_2) and the unknown function $S_\vartheta(\cdot)$ is differentiable with respect to ϑ_l , for any $1 \leq l \leq \iota$, moreover,*

$$\frac{\partial S_\vartheta(\cdot)}{\partial \vartheta_l} \in \mathbf{L}_1[0, 1].$$

Then for any square integrable estimator $\hat{\lambda}_n$ of λ and any $1 \leq l \leq \iota$,

$$\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda)^2 \geq \frac{\Lambda_l^2}{A_l + B_l + \mathcal{I}_l},$$

where

$$A_l = \sum_{j=1}^n \int_{\mathbb{R}^\iota} \frac{1}{\sigma_j^2(\vartheta)} \left(\frac{\partial S_\vartheta(x_j)}{\partial \vartheta_l} \right)^2 \Phi(\vartheta) d\vartheta,$$

$$B_l = 2 \sum_{j=1}^n \int_{\mathbb{R}^\iota} \left(\frac{\tilde{L}_l(x_j, S_\vartheta)}{g(x_j, S_\vartheta)} \right)^2 \Phi(\vartheta) d\vartheta, \quad \tilde{L}_l(x, S) = L_{x,S} \left(\frac{\partial S_\vartheta}{\partial \vartheta_l} \right),$$

the operator $L_{x,S}$ is defined in the condition \mathbf{H}_2).

Proof. We put

$$\varrho_l(v, \vartheta) = \frac{\frac{\partial}{\partial \vartheta_l} (f(v, \vartheta)\Phi(\vartheta))}{f(v, \vartheta)\Phi(\vartheta)}.$$

Integrating by parts and taking into account that the density Φ_l has a finite compact support yields

$$\begin{aligned}
\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda) \varrho_l &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^l} (\hat{\lambda}_n(v) - \lambda(\vartheta)) \frac{\partial}{\partial \vartheta_l} (f(v, \vartheta) \Phi(\vartheta)) \, d\vartheta \, dv \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^{l-1}} \left(\int_{\mathbb{R}} (\hat{\lambda}_n(v) - \lambda(\vartheta)) \frac{\partial}{\partial \vartheta_l} (f(v, \vartheta) \Phi(\vartheta)) \, d\vartheta_l \right) \prod_{j \neq l}^l d\vartheta_j \, dv \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^{l-1}} \left(\int_{\mathbb{R}} \left(\frac{\partial}{\partial \vartheta_l} \lambda(\vartheta) \right) f(v, \vartheta) \Phi(\vartheta) \, d\vartheta_l \right) \prod_{j \neq l}^l d\vartheta_j \, dv \\
&= \int_{\mathbb{R}^l} \left(\frac{\partial}{\partial \vartheta_l} \lambda(\vartheta) \right) \Phi(\vartheta) \left(\int_{\mathbb{R}^n} f(v, \vartheta) \, dv \right) \, d\vartheta = \Lambda_l.
\end{aligned}$$

Therefore by the Bouniakovskii-Cauchy-Schwartz inequality

$$\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda(\vartheta))^2 \geq \frac{\Lambda_l^2}{\tilde{\mathbf{E}}\varrho_l^2}.$$

To evaluate the denominator in the last ratio, note that

$$\varrho_l(v, \vartheta) = \frac{\frac{\partial}{\partial \vartheta_l} (f(v, \vartheta))}{f(v, \vartheta)} + \frac{\dot{\Phi}_l(\vartheta_l)}{\Phi_l(\vartheta_l)} = \frac{\partial}{\partial \vartheta_l} \ln f(v, \vartheta) + \frac{\dot{\Phi}_l(\vartheta_l)}{\Phi_l(\vartheta_l)}.$$

By (A.11) we find that

$$\frac{\partial}{\partial \vartheta_l} \ln f(Y, \vartheta) = \sum_{j=1}^n (\xi_j^2 - 1) \frac{1}{\sigma_j(\vartheta)} \frac{\partial}{\partial \vartheta_l} \sigma_j(\vartheta) + \sum_{j=1}^n \xi_j \frac{1}{\sigma_j(\vartheta)} \frac{\partial}{\partial \vartheta_l} S_\vartheta(x_j).$$

Thus

$$\tilde{\mathbf{E}} \frac{\partial}{\partial \vartheta_l} \ln f(Y, \vartheta) = 0.$$

Moreover, the condition \mathbf{H}_2) implies the relationship

$$\frac{\partial}{\partial \vartheta_l} \sigma_j(\vartheta) = \frac{\partial}{\partial \vartheta_l} g(x_j, S_\vartheta) = \tilde{L}_l(x_j, S_\vartheta)$$

from which it follows

$$\tilde{\mathbf{E}} \left(\frac{\partial}{\partial \vartheta_l} \ln f(Y, \vartheta) \right)^2 = A_l + B_l.$$

Hence Lemma 6.7. \square

6.8 Proof of (6.13)

Notice that to prove (6.13) it suffices to show that

$$\limsup_{\eta \rightarrow 0, d \rightarrow \infty} \sup_{N \geq 1} |\Omega_N(\eta, d) - 1| = 0, \quad (\text{A.12})$$

where $\Omega_N(\eta, d)$ is defined in (6.12). Indeed, by the direct calculation we find that, for any $N \geq 1$,

$$|\Omega_N(\eta, d) - 1| \leq \frac{(\max_{j \geq 1} (|\alpha_j(\eta) - 1| + |\beta_j(\eta) - 1|) + |i_d - 1|)}{\min(\beta_*(\eta), i_d)},$$

where $\beta_*(\eta) = \min_{j \geq 1} \beta_j(\eta)$. Therefore the definitions of $\alpha_j(\eta)$ and $\beta_j(\eta)$ and the properties (6.2) and (6.7) imply (A.12). \square

6.9 Proof of (6.10)

In this proof we use the following Lemmas.

Lemma 6.8. *Let S_ϑ be defined in (6.4) and $S_\vartheta^{(l)}$ be its derivative of order l , $1 \leq l \leq q$. Then, for sufficiently large n and any $0 < \varepsilon < 1$,*

$$\sum_{l=1}^q \|S_\vartheta^{(l)}\|^2 \leq (1 - \varepsilon/2)r + (1 + \mu) \sum_{l=1}^q \hat{\zeta}_l + (1 + \mu^{-1}) L_n^{-2},$$

where $\mu = \varepsilon/(2 - 2\varepsilon)$ and

$$\hat{\zeta}_l = \sum_{m=1}^M g_0^2(a_m) \sum_{j=1}^N \kappa_j^2 \frac{1}{n} \left(\frac{j\pi}{2h} \right)^{2l} \zeta_{m,j}, \quad \zeta_{m,j} = \hat{\xi}_{m,j}^2 - 1. \quad (\text{A.13})$$

Proof. From the definition of S_ϑ in (6.4) we obtain that

$$S_\vartheta^{(l)}(x) = \sum_{m=1}^M \sum_{j=1}^N \vartheta_{m,j} D_{m,j}^{(l)}(x),$$

where

$$D_{m,j}^{(l)}(x) = \frac{1}{h^l} Q_{0,j}^{(l)}(v_m(x)) + \frac{1}{h^l} Q_{1,j}^{(l)}(v_m(x)),$$

$$Q_{0,j}^{(l)}(u) = e_j^{(l)}(u) I_\eta(u), \quad Q_{1,j}^{(l)}(u) = \sum_{i=1}^l \binom{l}{i} e_j^{(l-i)}(u) I_\eta^{(i)}(u).$$

Setting

$$\overline{Q}_{m,i}^{(l)}(u) = \sum_{j=1}^N \vartheta_{m,j} Q_{i,j}^{(l)}(u), \quad i = 0, 1,$$

we find that

$$\begin{aligned} \|S_\vartheta^{(l)}\|^2 &= \frac{1}{h^{2l-1}} \sum_{m=1}^M \int_{-1}^1 \left(\overline{Q}_{0,j}^{(l)}(v) + \overline{Q}_{1,j}^{(l)}(v) \right)^2 dv \\ &= \frac{1}{h^{2l-1}} \sum_{m=1}^M \left(\|\overline{Q}_{m,0}^{(l)}\|^2 + 2(\overline{Q}_{m,0}^{(l)}, \overline{Q}_{m,1}^{(l)}) + \|\overline{Q}_{m,1}^{(l)}\|^2 \right). \end{aligned}$$

Therefore, for any $\mu > 0$, we get that

$$\|S_\vartheta^{(l)}\|^2 \leq \frac{1+\mu}{h^{2l-1}} \sum_{m=1}^M \|\overline{Q}_{m,0}^{(l)}\|^2 + \frac{1+\mu^{-1}}{h^{2l-1}} \sum_{m=1}^M \|\overline{Q}_{m,1}^{(l)}\|^2. \quad (\text{A.14})$$

Moreover

$$\begin{aligned} \|\overline{Q}_{m,0}^{(l)}\|^2 &= \int_{-1}^1 \left(\sum_{j=1}^N \vartheta_{m,j} Q_{0,j}(u) \right)^2 du \\ &= \int_{-1}^1 \left(\sum_{j=1}^N \vartheta_{m,j} e_j^{(l)}(u) \right)^2 I_\eta^2(u) du \leq \int_{-1}^1 \left(\sum_{j=1}^N \vartheta_{m,j} e_j^{(l)}(u) \right)^2 du \\ &= \sum_{j=1}^N \vartheta_{m,j}^2 \int_{-1}^1 \left(e_j^{(l)}(u) \right)^2 du \leq \sum_{j=1}^N \vartheta_{m,j}^2 (\pi j/2)^{2l}. \end{aligned}$$

Taking into account the definition of $\vartheta_{m,j}$ yields the following bound for the first sum in the right-hand term of (A.14)

$$\begin{aligned} \frac{1}{h^{2l-1}} \sum_{m=1}^M \|\overline{Q}_{m,0}^{(l)}\|^2 &\leq \sum_{m=1}^M g_0^2(a_m) \sum_{j=1}^N \kappa_j^2 \hat{\zeta}_{m,j}^2 \frac{1}{n} \left(\frac{j\pi}{2h} \right)^{2l} \\ &= H_l + \hat{\zeta}_l, \end{aligned}$$

where

$$H_l = \frac{1}{n} \left(\frac{\pi}{2h} \right)^{2l} \sum_{m=1}^M g_0^2(a_m) \sum_{j=1}^N \kappa_j^2 j^{2l}.$$

Moreover we can rewrite H_l as

$$\begin{aligned} H_l &\leq (1 + o(1)) \frac{\pi^{2l}}{n(2h)^{2l+1}} \varsigma_0 \sum_{j=1}^N \kappa_j^2 j^{2q} \\ &\leq (1 + o(1)) \frac{N^{2q+1} \pi^{2l} q}{(q+1)(2q+1)n(2h)^{2l+1}} \varsigma_0. \end{aligned}$$

This yields that

$$H_l = \begin{cases} (1 + o(1))(1 - \varepsilon)r, & \text{if } l = q, \\ o(n^{-1/(2q+1)}), & \text{if } l < q. \end{cases}$$

Let us consider now the second term in the right-hand part of (A.14). By putting

$$\psi_{m,i}(u) = \sum_{j=1}^N \vartheta_{m,j} e_j^{(l-i)}(u)$$

we can rewrite $\overline{Q}_{m,1}^{(l)}$ as $\overline{Q}_{m,1}^{(l)}(u) = \sum_{i=1}^l \binom{l}{i} I_\eta^{(i)}(u) \psi_{m,i}(u)$. Applying the Bouniakovskii-Cauchy-Schwartz inequality yields

$$\begin{aligned} \|\overline{Q}_{m,1}^{(l)}\|^2 &\leq \int_{-1}^1 \sum_{i=1}^l \binom{l}{i}^2 (I_\eta^{(i)}(u))^2 \sum_{s=1}^l \psi_{m,s}^2(u) du \\ &\leq C(\eta) \sum_{i=1}^l \int_{-1}^1 \psi_{m,i}^2(u) du = C(\eta) \sum_{i=1}^l \int_{-1}^1 \left(\sum_{j=1}^N \vartheta_{m,j} e_j^{(l-i)}(u) \right)^2 du \\ &= C(\eta) \sum_{i=1}^l \sum_{j=1}^N \vartheta_{m,j}^2 \|e_j^{(l-i)}\|^2 \leq C(\eta) \sum_{i=1}^l \sum_{j=1}^N \vartheta_{m,j}^2 (\pi j)^{2(q-i)} \\ &\leq C_1(\eta) \sum_{j=1}^N \vartheta_{m,j}^2 j^{2(q-1)}, \end{aligned}$$

where $C(\eta) = \max_{1 \leq l \leq q} \max_{-1 \leq u \leq 1} \sum_{i=1}^l \binom{l}{i}^2 (I_\eta^{(i)}(u))^2$ and $C_1(\eta) = qC(\eta) \pi^{2q}$.

From this it follows that

$$\begin{aligned} \frac{1}{h^{2l-1}} \sum_{m=1}^M \|\overline{Q}_{m,1}^{(l)}\|^2 &\leq \frac{C_1(\eta)}{h^{2l-1}} \sum_{m=1}^M \sum_{j=1}^N \vartheta_{m,j}^2 j^{2(q-1)} \\ &\leq \frac{d^2 C_1(\eta)}{nh^{2l}} \sum_{m=1}^M g(a_m) \sum_{j=1}^N \kappa_j^2 j^{2(q-1)} \\ &\leq C \frac{N^{2q}}{L_n^{2l+1} n^{2(q-l)/(2q+1)}}, \end{aligned}$$

where C is a constant. Notice that, if $l = q$ then

$$\frac{1}{h^{2q-1}} \sum_{m=1}^M \|\overline{Q}_{m,1}^{(l)}\|^2 \leq C \frac{1}{L_n^{2q+1}} N^q \sum_{j=1}^N j^{q-2} \leq C \frac{N^{2q-1}}{L_n^{2q+1}} \leq C \frac{1}{L_n^2}.$$

If $l < q$ then

$$\frac{1}{h^{2l-1}} \sum_{m=1}^M \|\overline{Q}_{m,1}^{(l)}\|^2 = O(L_n^{2q} n^{-2/(2q+1)}) = O(L_n^{-1}).$$

Therefore, putting in (A.14) $\mu = \varepsilon/2(1 - \varepsilon)$, we come to the assertion of this Lemma. \square

Lemma 6.9. *Let $\hat{\zeta}_l$ be defined in (A.13). Then, for any $p \geq 1/2$,*

$$\mathbf{E} \left(\sum_{l=1}^q \hat{\zeta}_l \right)^{2p} \leq \frac{C(p, d) L_n^p}{n^{p/(2q+1)}}, \quad (\text{A.15})$$

where $C(p, d)$ is a constant depending on p and d .

Proof. From (A.13) it follows that

$$\sum_{l=1}^q \hat{\zeta}_l = \frac{1}{n} \sum_{m=1}^M g_0^2(a_m) \tilde{\zeta}_m, \quad (\text{A.16})$$

where

$$\tilde{\zeta}_m = \sum_{j=1}^N \kappa_j^2 \zeta_{m,j} \sum_{l=1}^q \left(\frac{\pi j}{2h} \right)^{2l}.$$

The process $(\tilde{\zeta}_m, m \geq 1)$ is a square-integrable martingale-difference. Applying the Burkholder-Davis-Gundy inequality to the right-hand part of (A.16) yields, for $p \geq 1$,

$$\mathbf{E} \left(\sum_{l=1}^q \hat{\zeta}_l \right)^{2p} \leq \frac{C(p)}{n^{2p}} \mathbf{E} \left(\sum_{m=1}^M g_0^4(a_m) \tilde{\zeta}_m^2 \right)^p, \quad (\text{A.17})$$

where $C(p)$ is a constant depending on p . Moreover,

$$\begin{aligned} |\tilde{\zeta}_m| &\leq (d+1)^2 q \sum_{j=1}^N \left(\left(\frac{N}{j} \right)^q - 1 \right) \left(\frac{\pi j}{h} \right)^{2q} \\ &= \frac{q(d+1)^2 \pi^{2q}}{h^{2q}} N^{2q+1} \leq C L_n^{2q+1} \frac{1}{h^{2q}} = C L_n n^{2q/(2q+1)}. \end{aligned}$$

Thus (A.17) implies

$$\begin{aligned} \mathbf{E} \left(\sum_{l=1}^q \hat{\zeta}_l \right)^{2p} &\leq C \frac{L_n^{2p} n^{4pq/(2q+1)}}{n^{2p}} \left(\sum_{m=1}^M g^4(a_m) \right)^p \leq C \frac{L_n^{2p} n^{4pq/(2q+1)}}{n^{2p} h^p} \\ &= C \frac{L_n^p}{n^{p/(2q+1)}} \end{aligned}$$

and therefore (A.15). \square

Let us prove now (6.10). To end this, note that

$$\begin{aligned} \|S_\vartheta\|^2 &= \int_0^1 \left(\sum_{m=1}^M \sum_{j=1}^N \vartheta_{m,j} D_{m,j}(x) \right)^2 dx \\ &\leq h \sum_{m=1}^M \sum_{j=1}^N \vartheta_{m,j}^2 = \frac{d}{n} \sum_{m=1}^M g_0^2(a_m) \sum_{j=1}^N \kappa_j^2 \\ &\leq \frac{\text{Const} \cdot d}{\varphi_n^2 L_n} \sum_{j=1}^N \kappa_j^2 = o(1) \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Therefore, for sufficiently large n ,

$$|\omega_n| \leq (r+1) \mu_\vartheta(z : S_z \notin W_r^q) = (r+1) \mathbf{P}(S_\vartheta \notin W_r^q).$$

To obtain (6.10), it suffices to prove the relationship

$$\lim_{n \rightarrow \infty} \varphi_n \mathbf{P} \left(\sum_{l=1}^q \|S_\vartheta^{(l)}\|^2 > r \right) = 0.$$

Indeed, Lemma 6.8 implies that, for sufficiently large n ,

$$\sum_{l=1}^q \|S_\vartheta^{(l)}\|^2 \leq (1 - \varepsilon/2)r + (1 + \mu) \sum_{l=1}^q \hat{\zeta}_l + (1 + \mu^{-1})L_n^{-2},$$

where $\mu = \frac{\varepsilon}{2(1-\varepsilon)}$ and $\hat{\zeta}_l$ is defined by (A.13). Thus, for sufficiently large n ,

$$\mathbf{P} \left(\sum_{l=1}^q \|S_\vartheta^{(l)}\|^2 > r \right) \leq \mathbf{P} \left(\sum_{l=1}^q \hat{\zeta}_l > r_1 \right) \leq (r_1)^{-2p} \mathbf{E} \left(\sum_{l=1}^q \hat{\zeta}_l \right)^{2p},$$

where $r_1 = \varepsilon r / (2(1 + \mu))$. By making use of Inequality (A.15) with $p > 2q$ we obtain that

$$\varphi_n^2 \mathbf{P} \left(\sum_{l=1}^q \|S_\vartheta^{(l)}\|^2 > r \right) \leq \frac{c L_n^p n^{2q/(2q+1)}}{n^{p/(2q+1)}} = \frac{c L_n^p}{n^{(p-2q)/(2q+1)}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence (6.10). \square

References

- [1] Akritas, M.G. and Van Keilegom, I. (2001) ANCOVA methods for heteroscedastic nonparametric regression models. *J. American Statist. Assoc.*, **96**, n453, p. 220-232.
- [2] Barron, A., Birgé, L. and Massart, P.(1999) Risk bounds for model selection via penalization. *Probab. Theory Rel. Fields*, **113**, p.301-413.
- [3] Donoho, D.L. (1994) Statistical estimation and optimal recovery. *Ann. Statist.*, **22**, p. 238-270.
- [4] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage : asymptotia? *J. Roy. Statist. Soc. Ser. B*, **57**, p. 301-369.
- [5] Donoho, D.L. and Liu, R. (1991) Geometrizing rates of convergence, III. *Ann. Statist.*, **19**, p. 668-701.
- [6] Donoho, D.L. and Johnstone, I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, p. 879-921.
- [7] Efroimovich, S.Yu. (1999) *Nonparametric curve estimation. Methods, theory and applications*. Springer-Verlag, Berlin, New York.
- [8] Galtchouk, L. and Pergamenschikov, S. (2005) Asymptotically efficient estimates for nonparametric regression models. *Statistics and Probability Letters*, to appear.
- [9] Gill, R.D. and Levit, B.Y. (1995) Application of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1**, p. 59-79.
- [10] Goldfeld, S.M. and Quandt, R.E. (1972) *Nonlinear Methods in Econometrics*. North-Holland, London.
- [11] Golubev, G.K. (1992) Asymptotic minimax estimation of regression function in additive model. *Problems Inform. Transmission*, **28**(2), p. 101-112.
- [12] Golubev, G.K. and Nussbaum, M. (1993) Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.*, **37**, p. 521-529.
- [13] Gunst, R.F. and Mason, R.L. (1980) *Regression analysis and its applications: a data oriented approach*. Marcel Dekker, New-York.

- [14] Ibragimov, I.A. and Hasminskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. Springer, Berlin, New York.
- [15] Ibragimov, I.A. and Hasminskii, R.Z. (1982) Bounds for the risk of nonparametric regression estimates. *Theory Probab. Appl.*, **27**, p. 84-99
- [16] Kalifa, J. and Mallat, S. (2003) Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.*, **31**, p. 58-109.
- [17] Korostelev, A.P. (1993) Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.* **38**, p.737-743.
- [18] Massart, P.(2004) A non-asymptotic theory for model selection. *4ECM Stockholm 2004*, p. 309-323.
- [19] Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.*, **13**, p. 984-997.
- [20] Nemirovski, A. (2000) Topics in non-parametric statistics. *Lectures on probability theory and statistics, in Ecole de Saint-Flour 1998, Lecture Notes in Math.*, **1738**, Springer, p. 85-277.
- [21] Speckman, P. (1985) Spline smoothing and optimal rate of convergence in nonparametric regression models. *Ann. Statist.*, **13**, p. 970-983.
- [22] Stone, C. J. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, p. 1040-1053.