Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models.

L. Galtchouk *   S. Pergamenshchikov†

Abstract

An adaptive nonparametric estimation procedure is constructed for heteroscedastic regression when the noise variance depends on the unknown regression. A non-asymptotic upper bound for a quadratic risk (oracle inequality) is obtained.

Keywords: Adaptive estimation; Heteroscedastic regression; Nonasymptotic estimation; Nonparametric estimation; Oracle inequality.

AMS 2000 Subject Classifications: Primary: 62G08; Secondary: 62G05, 62G20

*Department of Mathematics, Strasbourg University 7, rue Rene Descartes, 67084, Strasbourg, France, e-mail: galtchou@math.u-strasbg.fr

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

1 Introduction

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

$$y_j \,=\, S(x_j) + \sigma_j \xi_j\,, \tag{1.1}$$

where design points $x_j = j/n$, $S(\cdot)$ is an unknown function to be estimated, $(\xi_j)_{1\le j\le n}$ is a sequence of i.i.d. random variables, $\mathbf{E}\xi_1 = 0\,,\quad \mathbf{E}\,\xi_1^2 = 1\,$, $(\sigma_j)_{1\le j\le n}$ are unknown volatility coefficients, which may depend on design points and on unknown regression function S.

The models of type (1.1) with $\sigma_j = \sigma_j(x_j)$ were introduced in [1] as a generalisation of the nonparametric ANCOVA model of [18]. It should be noted that heteroscedastic regressions with this type of volatility coefficients have been encountered in econometric studies, namely, in consumer budget studies utilizing observations on individuals with diverse incomes and in analyses of the investment behavior of firms of different sizes (see [12]). For example, for consumer budget problems one uses there (see p. 83) some parametric version of model (1.1) with the volatility coefficient defined as

$$\sigma_j^2 = c_0 + c_1 x_j + c_2 S^2(x_j)\,, \tag{1.2}$$

where c_0, c_1 and c_2 are some nonnegative unknown constants.

Moreover, this regression model appears in the drift estimation problem for stochastic differential equations when one passes from continuous time to discrete time model by making use of sequential kernel estimators having asymptotically minimal variances (see [6],[8]-[10]).

The volatility coefficient estimation in heteroscedastic regression was considered in a few papers (see, for example, [3] and the references therein). By making use of the squared first-order differences of the observations the initial problem in that paper was reduced to the regression function estimation in the model of type (1.1).

In this paper we develop the approach proposed in [7]. The first goal of the research is to construct an adaptive procedure for estimating the function S which does not use any smoothness information of S and which is based on observations $(y_j)_{1\le j\le n}$ and further to obtain a sharp non-asymptotic upper bound (oracle inequality) for a quadratic risk in the case when the smoothness of S is unknown. The second goal is to prove that the constructed procedure is efficient also in the asymptotic setup.

Problems of constructing a nonparametric estimator and proving a non-asymptotic upper bound for a risk in homoscedastic model, that is when

$\sigma_j \equiv \sigma$, were studied in few papers. A non-asymptotic upper bound for a quadratic risk over thresholding estimators is given in [13]. In papers [2], [15] an adaptive model selection procedure has been constructed. It is based on least squares estimators and a non-asymptotic upper bound has been obtained for a quadratic risk which is best in the principal term for the given class of estimators when the noise vector $(\xi_1 \ldots, \xi_n)$ is gaussian. This type of upper bounds is called the *oracle inequality*. In [5] the oracle inequality has been obtained for a model selection procedure based on any estimators in the case when the noise vector $(\xi_1, \ldots, \xi_n)$ has a spherically symmetric distribution. Moreover, some sharp oracle inequalities have been obtained also for homoscedastic regression with gaussian noises, see, for example, [14]. Here the adjective "sharp" means that the coefficient of the principal term may be chosen as close to unity as desired.

In the paper for heteroscedastic regression an adaptive procedure is constructed for which the sharp non-asymptotic oracle inequality is proved. It should be noted that the methods used in former papers to obtain the sharp oracle inequality in regression models are limited by the homoscedastic case since they are based on the fact that an orthogonal transformation of a noise gaussian vector $(\xi_1, \ldots, \xi_n)$ gives a gaussian vector. In heteroscedastic regression models under consideration these methods are not valid since the noise vector is not gaussian. To obtain sharp non-asymptotic oracle inequalities in the heteroscedastic case the authors develop a new mathematical tools based on "penalty" methods and Pinsker's type weights.

Moreover, in [11] we show that the given adaptive estimator is efficient in the asymptotic sense, that is, the sharp asymptotic lower bound is proved for a quadratic risk and it is attained over this estimator. The sharp non-asymptotic oracle inequality plays a cornerstone role in proving the asymptotic efficiency. To obtain the optimal upper bound for the risk one should use a weighted least squares estimator with weights depending on the smoothness of the unknown regression function. The smoothness being unknown, one can't use directly this weighted least squares estimator giving the minimal upper bound. The given sharp non-asymptotic oracle inequality allow us to replace this unknown weighted least squares estimator with an adaptive estimator. The risk of the adaptive estimator is less than the risk of optimal (unknown) weighted least squares estimator up to additive and multiplicative constants. Taking in account that the multiplicative constant tends to one and the order of the additive constant is less then the order of the convergence rate, we obtain that the risk of the given adaptive procedure asymptotically coincides with the risk of the optimal (unknown) weighted least squares estimator. Therefore, given the optimal lower bound, we obtain the asymptotic efficiency of the adaptive procedure satisfying the sharp non-asymptotic oracle inequality.

The paper is organized as follows. In Section 2 we construct an adaptive estimation procedure based on weighted least squares estimators and we obtain a non-asymptotic upper bound for the quadratic risk. In Section 3 we propose an estimator for the integrated noise variance and give the oracle inequality in the case of Sobolev space, $S \in W_r^k$. The proofs are given in Section 4. Section 5 contains a numerical comparison of the given procedure with an adaptive procedure proposed in [4]. The Appendix contains some technical results.

2 Oracle inequality

In this paper we study the non-asymptotic estimation problem of the function S in the model (1.1) by observations $(y_j)_{1\le j\le n}$ with odd sample number n. We assume that in (1.1) the sequence $(\xi_j)_{1\le j\le n}$ is i.i.d. with

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

Moreover, we assume that $(\sigma_l)_{1\le l\le n}$ is a sequence of positive random variables independent of $(\xi_i)_{1\le i\le n}$ and bounded away from $+\infty$, i.e. there exists some nonrandom unknown constant $\sigma_* > 0$ such that

$$\max_{1\le l\le n} \sigma_l^2 \,\le\, \sigma_*\,. \tag{2.2}$$

For any estimate $\widehat{S}_n$ of S based on observations $(y_j)_{1\le j\le n}$, the estimation accuracy is measured by the mean integrated squared error (MISE)

$$\mathbf{E}_S\,\|\widehat{S}_n - S\|_n^2\,, \tag{2.3}$$

where

$$\|\widehat{S}_n - S\|_n^2 = (\widehat{S}_n - S, \widehat{S}_n - S)_n = \frac{1}{n}\sum_{l=1}^n (\widehat{S}_n(x_l) - S(x_l))^2\,.$$

We make use of the trigonometric basis $(\phi_j)_{j\ge1}$ in $\mathcal{L}_2[0,1]$ with

$$\phi_1 = 1\,, \quad \phi_j(x) = \sqrt{2}\,Tr_j(2\pi[j/2]x)\,,\ j\ge 2\,, \tag{2.4}$$

where the function $Tr_j(x) = \cos(x)$ for even j and $Tr_j(x) = \sin(x)$ for odd j; $[x]$ denotes the integer part of x. Note that if n is odd, then this basis is orthonormal for the empirical inner product generated by the sieve $(x_j)_{1\le j\le n}$, that is for any $1 \le i, j \le n$,

$$(\phi_i\,,\,\phi_j)_n = \frac{1}{n}\sum_{l=1}^{n}\phi_i(x_l)\phi_j(x_l) = \delta_{ij}\,, \tag{2.5}$$

where δ_{ij} is Kronecker's symbol.

By making use of this basis we define the discrete Fourier transformation in (1.1) and obtain the Fourier coefficients

$$\widehat{\theta}_{j,n} = (Y, \phi_j)_n \quad \text{and} \quad \theta_{j,n} = (S, \phi_j)_n\,. \tag{2.6}$$

Here $Y = (y_1, \ldots, y_n)'$ and $S = (S(x_1), \ldots, S(x_n))'$. The prime denotes the transposition.

From (1.1) it follows directly that these Fourier coefficients satisfy the following equation

$$\widehat{\theta}_{j,n} = \theta_{j,n} + \frac{1}{\sqrt{n}}\xi_{j,n} \quad \text{with} \quad \xi_{j,n} = \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\sigma_l\xi_l\phi_j(x_l)\,. \tag{2.7}$$

We estimate the function S by the weighted least squares estimator

$$\widehat{S}_\lambda(x) = \sum_{j=1}^{n}\lambda(j)\widehat{\theta}_{j,n}\phi_j(x)\,, \tag{2.8}$$

where $x \in [0,1]$, the weight vector $\lambda = (\lambda(1), \ldots, \lambda(n))'$ belongs to some finite set Λ from $[0,1]^n$. We denote by ν the cardinal number of the set Λ.

Now we need to write a cost function to choose a weight $\lambda \in \Lambda$. Of course, it is obvious, that the best way is to minimize the cost function which is equal to the empirical squared error

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

which in our case is equal to

$$\mathrm{Err}_n(\lambda) \,=\, \sum_{j=1}^{n}\lambda^2(j)\widehat{\theta}_{j,n}^2 \,-\, 2\sum_{j=1}^{n}\lambda(j)\widehat{\theta}_{j,n}\,\theta_{j,n} \,+\, \sum_{j=1}^{n}\theta_{j,n}^2\,. \tag{2.9}$$

Since coefficients $\theta_{j,n}$ are unknown, we need to replace the term $\widehat{\theta}_{j,n}\theta_{j,n}$ by some estimator which we choose as

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{1}{n}\widehat{\varsigma}_n \,,$$

where $\widehat{\varsigma}_n$ is some estimator of the integrated noise variance

$$\varsigma_n = n^{-1}\sum_{l=1}^{n}\sigma^2_l \,. \tag{2.10}$$

Such type of estimators is given in (3.5).

Moreover, for this substitution to the empirical squared error one needs to pay a penalty. Finally, we define the cost function by the following way

$$J_n(\lambda) \,=\, \sum_{j=1}^{n}\lambda^2(j)\widehat{\theta}^2_{j,n} \,-2\,\sum_{j=1}^{n}\lambda(j)\,\widetilde{\theta}_{j,n} \,+\, \rho\widehat{P}_n(\lambda)\,, \tag{2.11}$$

where ρ is some positive coefficient which will be chosen later. The penalty term we define as

$$\widehat{P}_n(\lambda) = \frac{|\lambda|^2\widehat{\varsigma}_n}{n} \quad \text{with} \quad |\lambda|^2 = \sum_{j=1}^{n}\lambda^2(j)\,. \tag{2.12}$$

Note that in the case when the sequence $(\sigma_l)_{1\le l\le n}$ is known, i.e. $\widehat{\varsigma}_n = \varsigma_n$, we obtain

$$P_n(\lambda) = \frac{|\lambda|^2\varsigma_n}{n}\,. \tag{2.13}$$

We set

$$\widehat{\lambda} = \text{argmin}_{\lambda\in\Lambda}\, J_n(\lambda) \tag{2.14}$$

and define an estimator of S as

$$\widehat{S}_* = \widehat{S}_{\widehat{\lambda}}\,. \tag{2.15}$$

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.

To formulate the oracle inequality we introduce, for $0 < \rho < 1/3$, the following function

$$\Upsilon^*_n(\rho) = \frac{16\nu}{\rho} + 4u_{1,n}\left(1+\nu\frac{\sqrt{\xi^*}}{\sqrt{n}}\right) + 4\nu v_n\frac{\sqrt{\xi^*}}{\sqrt{n}}\,. \tag{2.16}$$

Here and thereafter we make use of the following notations: for $i = 1, 2$,

$$v_n = \max_{\lambda \in \Lambda} \sum_{j=1}^{n} \lambda(j) \quad \text{and} \quad u_{i,n} = \max_{\lambda \in \Lambda} \sup_{1 \le l \le n} |\sum_{j=1}^{n} \lambda^i(j) \, (\phi_j^2(x_l) - 1)| \,, \qquad (2.17)$$

.

Theorem 2.1. *Let Λ be any finite set in $[0,1]^n$. For any $n \ge 3$ and $0 < \rho < 1/3$, the estimator $\widehat{S}_*$ satisfies the oracle inequality*

$$\mathbf{E}_S \|\widehat{S}_* - S\|_n^2 \le \frac{1 + 3\rho - 2\rho^2}{1 - 3\rho} \min_{\lambda \in \Lambda} \mathbf{E}_S \|\widehat{S}_\lambda - S\|_n^2 + \frac{1}{n} \mathcal{B}_n(\rho) \,, \qquad (2.18)$$

where $\mathcal{B}_n(\rho) = \Psi_n(\rho) + \kappa(\rho) v_n \, \mathbf{E}_S |\widehat{\varsigma}_n - \varsigma_n|$ with

$$\Psi_n(\rho) = \frac{\rho(1-\rho)\Upsilon_n^*(\rho) + 2\nu + 2\rho^2(1-\rho)u_{2,n}}{\rho(1-3\rho)} \sigma_*$$

and $\kappa(\rho) = 4(1-\rho^2)/(1-3\rho)$.

If in the model (1.1) the volatility coefficients $(\sigma_l)_{1 \le l \le n}$ are known, then $\widehat{\varsigma}_n = \varsigma_n$ and inequality (2.18) has the following form

$$\mathbf{E}_S \|\widehat{S}_* - S\|_n^2 \le \frac{1 + 3\rho - 2\rho^2}{1 - 3\rho} \min_{\lambda \in \Lambda} \mathbf{E}_S \|\widehat{S}_\lambda - S\|_n^2 + \frac{1}{n} \Psi_n(\rho) \,. \qquad (2.19)$$

Remark 2.1. *Note that the principal term in the right-hand side of (2.18)–(2.19) is best in the class of estimators $(\widehat{S}_\lambda \,, \, \lambda \in \Lambda)$. Inequalities of such type are called* the sharp non-asymptotic oracle inequalities. *The inequality is sharp in the sense that the coefficient of the principal term may be chosen as close to 1 as desired. Similar inequalities for homoscedastic models (1.1) with $\sigma_l = \sigma$ were given, for example, in [14]. The methods used there cannot be extended to the heteroscedastic case since, after the Fourier transformation, the random variables $(\xi_{i,n})$ in model (2.7) are dependent contrary to the homoscedastic case, where these random variables are independent (see, for example,[17]).*

Remark 2.2. *If one would like to obtain the asymptotically minimal MISE of the estimator $\widehat{S}_*$, then the secondary term $\mathcal{B}_n(\rho)$ in (2.18) should be slowly varing, i.e. for any $\gamma > 0$,*

$$\mathcal{B}_n(\rho)/n^\gamma \to 0\,, \quad as \quad n \to \infty\,. \tag{2.20}$$

Indeed, since usually the optimal rate is of order $n^{2k/(2k+1)}$ for some $k \geq 1$, then after multiplying the inequality (2.18) by this rate the principal term gives the optimal constant and the secondary one is of the type

$$\mathcal{B}_n(\rho)/n^{1/(2k+1)}\,.$$

Therefore the property (2.20) provides the asymptotic vanishing, as $n \to \infty$, the secondary term $\mathcal{B}_n(\rho)$ for $k \geq 1$. To obtain the property (2.20), it suffices that, for any $\gamma > 0$,

$$\rho n^\gamma \to +\infty\,,$$

and

$$\frac{u_{1,n} + u_{2,n} + v_n \mathbf{E}_S|\widehat{\varsigma}_n - \varsigma_n|}{n^\gamma} \to 0\,, \quad as \quad n \to \infty\,, \tag{2.21}$$

thanks to definitions of $\mathcal{B}_n(\rho)$ and $\Psi_n(\rho)$. To obtain the first convergence it suffices to take the parameter ρ as $\rho \geq \varrho_n$, where ϱ_n is a slowly decreasing function, i.e.

$$\lim_{n\to\infty} \varrho_n \,=\, 0 \quad and\ for\ any \quad \gamma > 0 \quad \lim_{n\to\infty} n^\gamma\, \varrho_n \,=\, +\infty\,, \tag{2.22}$$

for example, $\varrho = 1/\ln n$. For the second convergence the choice of $u_{1,n}, u_{2,n}, v_n$ and of the estimator $\widehat{\varsigma}_n$ is proposed below.

Consider now the order of the termes $v_n, u_{1,n}, u_{2,n}$ and the function $\Psi_n(\rho)$ in the case when the finite set Λ is formed by a special version of Pinsker's weights (see, for example, [16]). To this end, we define the sieve

$$\mathcal{A}_\varepsilon = \{1, \ldots, k^*\} \times \{t_1, \ldots, t_m\}\,,$$

where $t_i = i\varepsilon$ and $m = [1/\varepsilon^2]$. We suppose that the parameters $k^* \geq 1$ and $0 < \varepsilon \leq 1$ are functions of n, i.e. $k^* = k^*_n$ and $\varepsilon = \varepsilon_n$, such that,

$$\left\{ \begin{array}{l} \lim_{n\to\infty} k^*_n = +\infty\,, \quad \lim_{n\to\infty} k^*_n/\ln n = 0\,, \\[2ex] \lim_{n\to\infty} \varepsilon_n \,=\, 0 \quad \text{and} \quad \lim_{n\to\infty} n^\gamma\, \varepsilon_n \,=+\infty\,, \end{array} \right. \tag{2.23}$$

for any $\gamma > 0$. For example, one can take for $n \geq 3$

$$\varepsilon_n = 1/\ln n \quad \text{and} \quad k^*_n = \overline{k} + \sqrt{\ln n}\,, \tag{2.24}$$

where $\overline{k}$ is any nonnegative constant.

For any $\alpha = (\beta, t) \in \mathcal{A}_\varepsilon$ we define the weight vector $\lambda_\alpha = (\lambda_\alpha(1), \ldots, \lambda_\alpha(n))'$ as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + \left(1 - (j/\omega_\alpha)^\beta\right)\, \mathbf{1}_{\{j_0 < j \leq \omega_\alpha\}}\,. \tag{2.25}$$

Here $j_0 = j_0(\alpha) = [\omega_\alpha/\ln n]$ with

$$\omega_\alpha = \overline{\omega} + (A_\beta\, t\, n)^{1/(2\beta+1)}\,,$$

where $\overline{\omega}$ is any nonnegative constant and

$$A_\beta = (\beta+1)(2\beta+1)/(\pi^{2\beta}\beta)\,.$$

Hence,

$$\Lambda \,=\, \{\lambda_\alpha\,,\, \alpha \in \mathcal{A}_\varepsilon\} \tag{2.26}$$

and $\nu = \nu_n = k^*_n m_n$. Note that in this case, in view of (2.23), for any $\gamma > 0$

$$\lim_{n\to\infty} \nu_n/n^\gamma = 0\,.$$

Moreover, by (2.25)

$$\sum_{j=1}^n \lambda_\alpha(j) = \mathbf{1}_{\{j_0 \geq 1\}}\, j_0 + \mathbf{1}_{\{\omega_\alpha \geq 1\}} \sum_{j=j_0+1}^{[\omega_\alpha]} \left(1 - (j/\omega_\alpha)^\beta\right) \leq \omega_\alpha\,.$$

Therefore, taking into account that $A_\beta \leq A_1 < 1$ for $\beta \geq 1$ we find that

$$\upsilon_n \leq \overline{\omega} + (n/\varepsilon)^{1/3}\,,$$

i.e.

$$\limsup_{n\to\infty} \frac{\upsilon_n}{\sqrt{n}} \,<\, \infty\,. \tag{2.27}$$

Moreover, note that for any $1 \leq l \leq n$, we get

$$\sum_{j=1}^n \lambda_\alpha(j)\,(\phi^2_j(x_l) - 1) = \mathbf{1}_{\{j_0 \geq 1\}} \sum_{j=1}^{j_0} (\phi^2_j(x_l) - 1)$$
$$+\, \mathbf{1}_{\{\omega_\alpha \geq 1\}} \sum_{j=j_0+1}^{[\omega_\alpha]} \left(1 - (j/\omega_\alpha)^\beta\right)(\phi^2_j(x_l) - 1)\,.$$

Thus Lemma A.2 implies that

$$u_{1,n} \leq \ 1 + 2^{\beta+1} \leq 1 + 2^{k^*+1} \, .$$

Due to the condition for k_n^* in (2.23) this function is slowly varying, i.e. for any $\gamma > 0$,

$$\lim_{n\to\infty} u_{1,n}/n^{\gamma} = 0 \, . \tag{2.28}$$

By the same way we obtain that

$$u_{2,n} \leq 1 + 2^{k^*+2} + 2^{2k^*+1}$$

and, therefore, for any $\gamma > 0$

$$\lim_{n\to\infty} u_{2,n}/n^{\gamma} = 0 \, . \tag{2.29}$$

Then for any sequence $(\varrho_n)_{n\geq 1}$ satisfying properties (2.22) and for any $\gamma > 0$,

$$\lim_{n\to\infty} \ \sup_{\varrho_n \leq \rho \leq 1/6} \ \Psi_n(\rho)/n^{\gamma} = 0 \, .$$

Remark 2.3. *As we shall see in the proof of Theorem 2.1, the oracle inequality is true for any basis and any design* $(x_k)_{1\leq k\leq n}$ *which possesse the orthonormality property* (2.5)*. Moreover, if the sequences* (2.17) *and an estimator of the unknown integrated variance* $\widehat{\varsigma}_n$ *satisfy the property* (2.21)*, then the secondary term in the inequality* (2.18) *possesses the property* (2.20)*. In the next section we give an estimator* $\widehat{\varsigma}_n$ *in the case of the trigonometric basis and the equidistant design.*

3 Oracle inequality for $S \in W_r^k$

Assume that $S : \mathbb{R} \to \mathbb{R}$ is a k times differentiable 1-periodic function such that

$$\sum_{j=0}^{k} \|S^{(j)}\|^2 \leq r \, , \tag{3.1}$$

where

$$\|f\|^2 = \int_0^1 f^2(t)\mathrm{d}t\,. \tag{3.2}$$

We denote by W_r^k the set of all such functions. Moreover, we suppose that $r>0$ and $k\ge 1$ are unknown parameters.

Note that, the space W_r^k can be represented as an ellipses in the Hilbert space, i.e.

$$W_r^k = \{S\in\mathcal{L}_2[0,1]\,:\, S=\sum_{j=1}^\infty \theta_j\phi_j \quad \text{such that} \quad \sum_{j=1}^\infty a_j\theta_j^2 \le r\}\,, \tag{3.3}$$

where the basis functions $(\phi_j)_{j\ge1}$ are defined in (2.4); $(\theta_j)_{j\ge1}$ are the Fourier coefficients, i.e.

$$\theta_j = (S,\phi_j) = \int_0^1 S(t)\phi_j(t)\mathrm{d}t\,. \tag{3.4}$$

The coefficients $(a_j)_{j\ge1}$ are defined as

$$a_j = \sum_{l=0}^k \|\phi_j^{(l)}\|^2 = \sum_{l=0}^k (2\pi[j/2])^{2l}\,.$$

To estimate ς_n, we make use of the following estimator:

$$\widehat{\varsigma}_n = \sum_{j=d_n+1}^n \widehat{\theta}^2_{j,n}\,, \tag{3.5}$$

where the parameter d_n, $1\le d_n\le n-1$, will be chosen later.

In Section 4 we show the following result.

Lemma 3.1. *For any $n\ge 2$, $r>0$ and $S\in W_r^1$*

$$\mathbf{E}_S\,|\widehat{\varsigma}_n - \varsigma_n| \le \frac{2\left(\sqrt{\xi^*}+\sqrt{2}\right)\sigma_* + \varsigma_n^*(r)}{\sqrt{n}}\,, \tag{3.6}$$

where

$$\varsigma_n^*(r) = \frac{4r\sqrt{n}}{d_n^2} + 4\sqrt{r\sigma_*}\frac{1}{d_n} + \frac{(2+d_n)\sigma_*}{\sqrt{n}}\,.$$

If we choose the parameter d_n in (3.5) such that

$$\lim_{n\to\infty} d_n/\sqrt{n} = 0 \quad \text{and} \quad \lim_{n\to\infty} d_n^2/\sqrt{n} = \infty\,, \tag{3.7}$$

we obtain that

$$\lim_{n\to\infty} \varsigma_n^*(r) = 0\,.$$

Theorem 2.1 and inequality (3.6) imply immediately the following result.

Theorem 3.2. *Let Λ be any finite set in $[0,1]^n$. Assume that in the model* (1.1) *the function S belongs to W_r^1. Then, for any $n \geq 3$ and $0 < \rho < 1/3$, the procedure $\widehat{S}_*$ from* (2.15) *with $\widehat{\varsigma}_n$ defined by* (3.5) *and* (3.7) *satisfies the following oracle inequality*

$$\mathbf{E}_S \|\widehat{S}_* - S\|_n^2 \leq \frac{1+3\rho-2\rho^2}{1-3\rho} \min_{\lambda\in\Lambda} \mathbf{E}_S \|\widehat{S}_\lambda - S\|_n^2 + \frac{1}{n}\mathcal{D}_n(\rho, r)\,, \tag{3.8}$$

where

$$\mathcal{D}_n(\rho,r) = \Psi_n(\rho) + \kappa(\rho)\,\left(2\left(\sqrt{\xi^*}+\sqrt{2}\right)\sigma_* + \varsigma_n^*(r)\right)\frac{\upsilon_n}{\sqrt{n}}\,.$$

Moreover, if the sequencies (2.17) *satisfy the properties* (2.27)–(2.29) *then, for any $\gamma > 0$,*

$$\lim_{n\to\infty} \sup_{\varrho_n \leq \rho \leq 1/6} \mathcal{D}_n(\rho,r)/n^\gamma = 0\,,$$

where ϱ_n is any slowly decreasing sequence, i.e. satisfying (2.22).

Remark 3.1. *The inequality* (3.8) *is used to prove the asymptotic efficiency of the estimator* (2.15) *(see, [11]). To obtain the minimal asymptotic quadratic risk, one has to take a weighted least squares estimator* (2.8) *with weights of type* (2.25)*, where the parameter α depends on unknown smoothness of unknown function S. So one can't use this estimator directly. The oracle inequality allows us to overcome this difficulty because the upper bound is majorized up to a multiplicative and a additive constants by the minimal quadratic risk over all weighted estimators including the optimal one. Taking into account that the multiplicative constant tends to unity, as $n \to \infty$, and the additive constant is negligible in comparison with any degree of n and, in particular, with the optimal convergence rate and then multiplying the inequality* (3.8) *by the optimal convergence rate, one obtains the asymptotically minimal upper bound for the quadratic risk of the estimator* (2.15)*. The last result means that this estimator is asymptotically efficient.*

4 Proofs

4.1 Proof of Theorem 2.1

First of all, note that we can represent the empirical squared error $\mathrm{Err}_n(\lambda)$ by the following way

$$\mathrm{Err}_n(\lambda) = J_n(\lambda) + 2\sum_{j=1}^{n} \lambda(j)\theta'_{j,n} + \|S\|_n^2 - \rho\,\widehat{P}_n(\lambda) \tag{4.1}$$

with $\theta'_{j,n} = \widetilde{\theta}_{j,n} - \theta_{j,n}\widehat{\theta}_{j,n}$. The second term is most difficult to handle in the right-hand part of (4.1). To estimate, we decompose it in the sum of terms and we apply appropriate technique to each of them. By setting

$$\varsigma_{j,n} = \mathbf{E}_S\,\xi^2_{j,n} = \frac{1}{n}\sum_{l=1}^{n} \sigma^2_l \phi^2_j(x_l) \quad \text{and} \quad \mu_{j,n} = \xi^2_{j,n} - \varsigma_{j,n}\,, \tag{4.2}$$

we find that

$$\theta'_{j,n} = \frac{1}{\sqrt{n}}\theta_{j,n}\xi_{j,n} + \frac{1}{n}\mu_{j,n} + \frac{1}{n}\left(\varsigma_{j,n} - \widehat{\varsigma}_n\right). \tag{4.3}$$

Moreover, we can represent $\mu_{j,n}$ as

$$\mu_{j,n} = \frac{1}{n}\sum_{l=1}^{n} \sigma^2_l \phi^2_j(x_l)\eta_l + 2\sum_{l=2}^{n} \tau_{j,l}\,\xi_l = \mu'_{j,n} + 2\mu''_{j,n} \tag{4.4}$$

with $\eta_l = \xi^2_l - 1$ and

$$\tau_{j,l} = \frac{1}{n}\sigma_l\phi_j(x_l)\sum_{k=1}^{l-1} \sigma_k\phi_j(x_k)\,\xi_k\,.$$

Now we set

$$N'(\lambda) = \sum_{j=1}^{n} \lambda(j)\,\mu'_{j,n} \quad \text{and} \quad N''(\lambda) = \frac{1}{\sqrt{n\varsigma_n}}\sum_{j=1}^{n} \overline{\lambda}(j)\,\mu''_{j,n}\,\mathbf{1}_{\{\varsigma_n>0\}}\,, \tag{4.5}$$

where $\overline{\lambda}(j) = \lambda(j)/|\lambda|$. In the Appendix we show that

$$\sup_{\lambda\in\Lambda} \mathbf{E}_S\,|N'(\lambda)| \,\le\, \sigma_*(v_n + u_{1,n})\frac{\sqrt{\xi^*}}{\sqrt{n}} \tag{4.6}$$

and

$$\sup_{\lambda\in\mathbb{R}^n} \mathbf{E}_S (N''(\lambda))^2 \le \frac{2\sigma_*}{n}\,. \tag{4.7}$$

Now, for any $\lambda\in\Lambda$, we rewrite (4.1) as

$$\begin{aligned}\mathrm{Err}_n(\lambda) = J_n(\lambda) + \frac{2}{n} N'(\lambda) + 4\sqrt{P_n(\lambda)} N''(\lambda) \\ + 2M(\lambda) + \frac{2}{n}\,\Delta(\lambda) + \|S\|_n^2 - \rho \widehat{P}_n(\lambda)\,,\end{aligned}$$

where $P_n(\lambda)$ is defined in (2.13),

$$\Delta(\lambda) = \sum_{j=1}^n \lambda(j)\,(\varsigma_{j,n} - \widehat{\varsigma}_n) \quad \text{and} \quad M(\lambda) = n^{-1/2} \sum_{j=1}^n \lambda(j)\theta_{j,n}\xi_{j,n}\,. \tag{4.8}$$

Further we estimate the term $\Delta(\lambda)$. Setting

$$\overline{\varsigma}_{j,n} = \varsigma_{j,n} - \varsigma_n = \frac{1}{n}\sum_{l=1}^n \sigma_l^2(\phi_j^2(x_l) - 1)\,, \tag{4.9}$$

we obtain that

$$\begin{aligned}|\Delta(\lambda)| &\le \,|\sum_{j=1}^n \lambda(j)\overline{\varsigma}_{j,n}\,| + v_n|\widehat{\varsigma}_n - \varsigma_n| \\ &\le \sigma_* u_{1,n} + v_n|\widehat{\varsigma}_n - \varsigma_n|\,.\end{aligned} \tag{4.10}$$

Now from (4.1) we obtain that, for some fixed $\lambda_0\in\Lambda$,

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) - \mathrm{Err}_n(\lambda_0) = J(\widehat{\lambda}) - J(\lambda_0) + 2M(\widehat{\vartheta}) + \frac{2}{n}N'(\widehat{\vartheta}) \\ + 4\sqrt{P_n(\widehat{\lambda})}N''(\widehat{\lambda}) - 4\sqrt{P_n(\lambda_0)}N''(\lambda_0) \\ - \rho\widehat{P}_n(\widehat{\lambda}) + \rho\widehat{P}_n(\lambda_0) + \frac{2}{n}\left(\Delta(\widehat{\lambda}) - \Delta(\lambda_0)\right)\,,\end{aligned}$$

where $\widehat{\vartheta} = \widehat{\lambda} - \lambda_0$.

By the definition of $\widehat{\lambda}$ in (2.14) and by (4.10) we get

$$\begin{aligned}\mathrm{Err}_n(\widehat{\lambda}) - \mathrm{Err}_n(\lambda_0) \le 2M(\widehat{\vartheta}) + \frac{4\sigma_* u_{1,n} + 4v_n|\widehat{\varsigma}_n - \varsigma_n|}{n} \\ + \frac{2}{n}N'(\widehat{\vartheta}) + 4\sqrt{P_n(\widehat{\lambda})}N''(\widehat{\lambda}) - \rho\widehat{P}_n(\widehat{\lambda}) \\ + \rho\widehat{P}_n(\lambda_0) - 4\sqrt{P_n(\lambda_0)}N''(\lambda_0)\,.\end{aligned}$$

Moreover, making use of the inequality

$$2|ab| \le \epsilon a^2 + \epsilon^{-1} b^2 \tag{4.11}$$

with $\epsilon = \rho/4$ and taking into account the definition of the penalty term in (2.12) we deduce, for any $\lambda \in \Lambda$,

$$\begin{aligned} 4\sqrt{P_n(\lambda)}|N''(\lambda)| &\le \rho P_n(\lambda) + 4\frac{(N''(\lambda))^2}{\rho} \\ &\le \rho \widehat{P}_n(\lambda) + \rho\frac{|\lambda|^2|\widehat{\varsigma}_n - \varsigma_n|}{n} + \frac{4(N''(\lambda))^2}{\rho}\,. \end{aligned}$$

Thus from here it follows that

$$\mathrm{Err}_n(\widehat{\lambda}) \le \mathrm{Err}_n(\lambda_0) + 2M(\widehat{\vartheta}) + \Upsilon_n + 2\rho \widehat{P}_n(\lambda_0)\,, \tag{4.12}$$

where

$$\Upsilon_n = \frac{4}{n}N_1^* + \frac{8}{\rho}(N_2^*)^2 + \frac{4\sigma_* u_{1,n}}{n} + \frac{4+2\rho}{n} v_n |\widehat{\varsigma}_n - \varsigma_n|$$

with $N_1^* = \sup_{\lambda\in\Lambda}|N'(\lambda)|$ and $N_2^* = \sup_{\lambda\in\Lambda}|N''(\lambda)|$. Moreover, note that the bounds (4.6), (4.7) and (4.10) imply that

$$\mathbf{E}_S \Upsilon_n \le \Upsilon_n^*(\rho)\frac{\sigma_*}{n} + \frac{4+2\rho}{n} v_n \mathbf{E}_S|\widehat{\varsigma}_n - \varsigma_n|\,, \tag{4.13}$$

where the function $\Upsilon_n^*(\rho)$ is defined in (2.16).

Now we study the second term in (4.8). First, note that for any nonrandom vector $\vartheta = (\vartheta(1), \ldots, \vartheta(n))' \in \mathbb{R}^n$ Lemma A.4 implies

$$\mathbf{E}_S M^2(\vartheta) \le \frac{\sigma_*}{n}\sum_{j=1}^n \vartheta^2(j)\theta_{j,n}^2 = \sigma_* \frac{\|S_\vartheta\|_n^2}{n}\,, \tag{4.14}$$

where

$$S_\vartheta = \sum_{j=1}^n \vartheta(j)\theta_{j,n}\phi_j\,.$$

We set now

$$Z^* = \sup_{\vartheta\in\Lambda_1}\frac{nM^2(\vartheta)}{\|S_\vartheta\|_n^2} \quad \text{with} \quad \Lambda_1 = \Lambda - \lambda_0\,.$$

To estimate this term we apply the inequality (4.14), i.e.

$$\mathbf{E}_S\, Z^* \,\le\, \sum_{\vartheta\in\Lambda_1} \frac{n\mathbf{E}_S\, M^2(\vartheta)}{\|S_\vartheta\|_n^2} \le \nu\sigma_*\,. \tag{4.15}$$

Moreover, making use of inequality (4.11) with $\epsilon=\rho\|S_\vartheta\|_n$, we get

$$2|M(\vartheta)| \le \rho\|S_\vartheta\|_n^2 + \frac{Z^*}{n\rho}\,. \tag{4.16}$$

Now we estimate $\|S_\vartheta\|_n^2$. We have

$$\|S_\vartheta\|_n^2 - \|\widehat{S}_\vartheta\|_n^2 = \sum_{j=1}^n \vartheta^2(j)(\theta_{j,n}^2 - \widehat{\theta}_{j,n}^2) \le -2M_1(\vartheta) \tag{4.17}$$

with

$$M_1(\vartheta) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \vartheta^2(j)\theta_{j,n}\xi_{j,n}\,.$$

Now, taking into account that $|\vartheta(j)|\le 1$ for any $\vartheta\in\Lambda_1$, we obtain

$$\mathbf{E}_S M_1^2(\vartheta) \le \sigma_* \frac{\|S_\vartheta\|_n^2}{n}\,.$$

Putting

$$Z_1^* = \sup_{\vartheta\in\Lambda_1} \frac{nM_1^2(\vartheta)}{\|S_\vartheta\|_n^2}\,,$$

we get

$$\mathbf{E}_S\, Z_1^* \,\le \nu\sigma_*\,. \tag{4.18}$$

Therefore, applying inequality (4.16) for $M_1(\vartheta)$ in (4.17) we deduce the upper bound for $\|S_\vartheta\|_n^2$, i.e.

$$\|S_\vartheta\|_n^2 \le \frac{1}{1-\rho}\|\widehat{S}_\vartheta\|_n^2 + \frac{Z_1^*}{n\rho(1-\rho)}\,. \tag{4.19}$$

Taking into account this inequality in (4.16) we obtain that

$$\begin{aligned} 2M(\vartheta) &\le \frac{\rho}{1-\rho}\|\widehat{S}_\vartheta\|_n^2 + \frac{Z^*+Z_1^*}{n\rho(1-\rho)} \\ &\le \frac{2\rho(\mathrm{Err}_n(\lambda)+\mathrm{Err}_n(\lambda_0))}{1-\rho} + \frac{Z^*+Z_1^*}{n\rho(1-\rho)}\,. \end{aligned}$$

Therefore (4.12) implies that

$$\mathrm{Err}_n(\widehat{\lambda}) \le \frac{1+\rho}{1-3\rho}\mathrm{Err}_n(\lambda_0) + \frac{1-\rho}{1-3\rho}\Upsilon_n + \frac{Z^* + Z_1^*}{n\rho(1-3\rho)} + \frac{2\rho(1-\rho)}{1-3\rho}\widehat{P}_n(\lambda_0)\,,$$

Now by inequalities (4.15)–(4.18) we get that

$$\mathbf{E}_S\mathrm{Err}_n(\widehat{\lambda}) \le \frac{1+\rho}{1-3\rho}\mathbf{E}_S\mathrm{Err}_n(\lambda_0) + \frac{1-\rho}{1-3\rho}\mathbf{E}_S\,\Upsilon_n + \frac{2\nu\sigma_*}{n\rho(1-3\rho)} + \frac{2\rho(1-\rho)}{1-3\rho}\,\mathbf{E}_S\widehat{P}_n(\lambda_0)\,.$$

By making use of inequality (4.13) and Lemma A.1 we come to Theorem 2.1.
□

4.2 Proof of Lemma 3.1

First notice that from (2.7) we obtain that

$$\begin{aligned}\widehat{\varsigma}_n - \varsigma_n \;=\; & \sum_{j=d_n+1}^{n} \theta^2_{j,n} + \frac{2}{\sqrt{n}} \sum_{j=d_n+1}^{n} \theta_{j,n}\,\xi_{j,n} \\ & + n^{-1} \sum_{j=d_n+1}^{n} \mu_{j,n} + n^{-1} \sum_{j=d_n+1}^{n} \overline{\varsigma}_{j,n} - \frac{d_n}{n}\,\varsigma_n \\ := & \,\Delta_1 + \frac{2}{\sqrt{n}}\Delta_2 + \frac{1}{n}\Delta_3 + \frac{1}{n}\Delta_4 - \frac{d_n}{n}\,\varsigma_n\,,\end{aligned}$$

where $\mu_{j,n}$ and $\overline{\varsigma}_{j,n}$ are defined in (4.3) and (4.9) respectively.

We estimate the first term by Lemma A.3 for $S \in W^1_r$. We have

$$\Delta_1 \le \frac{4r}{d^2_n}\,.$$

The next term we estimate with the help of Lemma A.4. We get that

$$\mathbf{E}_S(\Delta_2)^2 \le \sigma_*\Delta_1 \le \sigma_*\frac{4r}{d^2_n}\,.$$

By (4.4) and (4.5) we can represent Δ_3 as

$$\Delta_3 = N'(\lambda_I) + 2|\lambda_I|\sqrt{n\varsigma_n}N''(\lambda_I)$$

with the vector $\lambda_I = (\lambda_I(1), \ldots, \lambda_I(n))'$ having the indicator components, i.e. $\lambda_I(j) = \mathbf{1}_{\{j>d_n\}}$. By estimating in (A.1) ϕ_j^2 by 2 we obtain

$$\mathbf{E}_S\,|N'(\lambda_I)| \le\, 2\sigma_*\,\sqrt{\xi^*}\,\sqrt{n}\,.$$

Thus the upper bound (4.7) implies

$$\mathbf{E}_S|\Delta_3| \le 2\sigma_*(\sqrt{\xi^*} + \sqrt{2})\sqrt{n} = \overline{\sigma}\sqrt{n}\,.$$

Moreover, due to Lemma A.2 with $k = 0$, one has

$$\begin{aligned}|\Delta_4| \,&=\, \left|n^{-1}\sum_{l=1}^{n}\sigma_l^2\sum_{j=d_n+1}^{n}(\phi_j^2(x_l)-1)\right|\\ &\le \frac{\sigma_*}{n}\sum_{l=1}^{n}\left|\sum_{j=1}^{n}(\phi_j^2(x_l)-1)\right| + \frac{\sigma_*}{n}\sum_{l=1}^{n}\left|\sum_{j=1}^{d_n}(\phi_j^2(x_l)-1)\right| \le 2\sigma_*\,.\end{aligned}$$

Hence Lemma 3.1. $\square$

5 Numerical example

In this Section, we compare via simulations the adaptive procedure proposed for the heteroscedastic regression in [4] (section 4.1) with that of (2.15).

Consider the model (1.1) with

$$S(x) \,=\, x\,\sin(2\pi x) \,+\, x^2\,(1-x)\,\cos(4\pi x) \quad \text{and} \quad \sigma_j^2 = 1 + S^2(x_j)\,,$$

assuming that $(\xi_k)_{k\ge1}$ follow the gaussian distribution with zero mean and unit variance.

In the procedure (2.15) we take the weight vectors defined in (2.25) with $k^* = 100 + \sqrt{\ln n}$, $\varepsilon = 1/\ln n$, $m = \ln^2 n$, $\rho = 1/(3 + \ln^2 n)$ and

$$\omega_\beta = 10 + (A_\beta t n)^{1/(2\beta+1)}\,.$$

Moreover, we make use of the estimate (3.5) with $d_n = n^{1/3}$.

The results of simulations are reported in Table 1.

Table 1

$\hat{E}\,\|S^* - S\|_n^2$	$\hat{E}\,\|\tilde{S} - S\|_n^2$	n
0.260	0.410	21
0.148	0.427	41
0.058	0.476	101
0.034	0.430	201
0.019	0.448	401

The columns of Table 1 with the headings $\hat{E}\,\|S^* - S\|_n^2$, $\hat{E}\,\|\tilde{S} - S\|_n^2$ and n report, respectively, the empirical quadratic risk for the procedure (2.25), the empirical quadratic risk for the procedure from [4] and the sample size. To calculate the empirical risks 50 independent Monte Carlo simulations were performed.

Table 1 shows that in comparison with the procedure from [4] our adaptive estimator performs resonably well for the small sample sizes.

6 Appendix

A.1 Proof of (4.6)

First note that we can represent the term $N'(\lambda)$ as

$$N'(\lambda) = \sum_{l=1}^{n} g_{l,n}\,\eta_l \quad \text{with} \quad g_{l,n} = \frac{\sigma_l^2}{n} \sum_{j=1}^{n} \lambda(j)\phi_j^2(x_l)\,.$$

Recalling that $\mathbf{E}\,\eta_1^2 = \xi^* - 1$ we calculate

$$\mathbf{E}_S\,(N'(\lambda))^2 = \frac{\xi^* - 1}{n^2} \sum_{l=1}^{n} \mathbf{E}_S \sigma_l^4 \left(\sum_{j=1}^{n} \lambda(j)\phi_j^2(x_l) \right)^2 .$$

Therefore for any vector $\lambda \in \mathbb{R}^n$

$$\mathbf{E}_S\,|N'(\lambda)| \le \sigma_* \frac{\sqrt{\xi^*}}{\sqrt{n}} \max_{1\le l\le n} \,|\sum_{j=1}^{n} \lambda(j)\phi_j^2(x_l)\,|\,. \tag{A.1}$$

Thus taking into account here definitions (2.17) we come to inequality (4.6).
□

A.2 Proof of (4.7)

By putting $g_l = \sum_{j=1}^{n} \overline{\lambda}(j)\tau_{j,l}$ and taking into account that the random variables $(\xi_k)_{1\le k\le n}$ are independent of $(\sigma_k)_{1\le k\le n}$ we obtain that

$$\mathbf{E}_S\left((N''(\lambda))^2 \,|\, \sigma_1,\ldots,\sigma_n\right) = \mathbf{1}_{\{\varsigma_n>0\}} \left(\sum_{k=1}^{n} \sigma_k^2\right)^{-1} \sum_{l=2}^{n} \widehat{g}_l\,, \tag{A.2}$$

where

$$\widehat{g}_l = \mathbf{E}(g_l^2 \,|\, \sigma_1,\ldots,\sigma_n) = \frac{\sigma_l^2}{n^2} \sum_{k=1}^{l-1} \sigma_k^2 \left(\sum_{j=1}^{n} \overline{\lambda}(j)\phi_j(x_l)\phi_j(x_k)\right)^2 .$$

Therefore the orthonormality property (2.5) implies that for any $\lambda \in \mathbb{R}^n$

$$\begin{aligned}\widehat{g}_l &\le \sigma_* \frac{\sigma_l^2}{n^2} \sum_{k=1}^{n} \left(\sum_{j=1}^{n} \overline{\lambda}(j)\phi_j(x_l)\phi_j(x_k)\right)^2 \\ &= \sigma_* \frac{\sigma_l^2}{n} \sum_{j=1}^{n} \overline{\lambda}^2(j)\phi_j^2(x_l) \;\le\; \frac{2\sigma_*}{n}\,\sigma_l^2\,.\end{aligned}$$

Now by making use of this inequality in (A.2) we get (4.7).
□

A.3 Technical lemma

Lemma A.1. *For any $n \ge 1$ and $\lambda \in \Lambda$,*

$$\mathbf{E}_S \widehat{P}_n(\lambda) \le \mathbf{E}_S\, Err_n(\lambda) + \frac{\upsilon_n}{n}\mathbf{E}_S|\widehat{\varsigma}_n - \varsigma_n| + \frac{\sigma_* u_{2,n}}{n}\,.$$

Proof. Indeed, by the definition of $\mathrm{Err}_n(\lambda)$ we have

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

Therefore,

$$\mathbf{E}_S \mathrm{Err}_n(\lambda) \geq \mathbf{E}_S \frac{1}{n} \sum_{j=1}^{n} \lambda^2(j)\, \xi_{j,n}^2 = \mathbf{E}_S \frac{1}{n} \sum_{j=1}^{n} \lambda^2(j)\, \varsigma_{j,n}\,,$$

where the sequence $(\varsigma_{j,n})$ is defined in (4.2). Moreover, note that the last term can be estimated as

$$\left| \sum_{j=1}^{n} \lambda^2(j) \varsigma_{j,n} - |\lambda|^2 \varsigma_n \right| = \left| \frac{1}{n} \sum_{l=1}^{n} \sigma_l^2 \sum_{j=1}^{n} \lambda^2(j)\, (\phi_j^2(x_l) - 1) \right| \leq \sigma_* u_{2,n}\,.$$

We recall that the definition of the set Λ and the definition of v_n in (2.17) imply that $|\lambda|^2 \leq v_n$ for $\lambda \in \Lambda$. Therefore for any $\lambda \in \Lambda$

$$\begin{aligned}\sum_{j=1}^{n} \lambda^2(j)\, \varsigma_{j,n} &\geq |\lambda|^2 \widehat{\varsigma}_n - \sigma_* u_{2,n} - |\lambda|^2 |\widehat{\varsigma}_n - \varsigma_n| \\ &\geq |\lambda|^2 \widehat{\varsigma}_n - \sigma_* u_{2,n} - v_n |\widehat{\varsigma}_n - \varsigma_n|\,.\end{aligned}$$

Hence the desired inequality. □

A.4 Properties of trigonometric basis

Lemma A.2. *For any $k \geq 0$,*

$$\sup_{N \geq 2} \sup_{x \in [0,1]} N^{-k} \left| \sum_{l=2}^{N} l^k\, (\phi_l^2(x) - 1) \right| \leq 2^k\,. \tag{A.3}$$

Proof. Due to the properties of the trigonometric functions, we get

$$\begin{aligned}\sum_{l=2}^{N} l^k\, (\phi_l^2(x) - 1) \;=\; &\sum_{1 \leq l \leq N/2} (2l)^k\, \cos(4\pi l x) \\ &- \sum_{1 \leq l \leq (N-1)/2} (2l+1)^k\, \cos(4\pi l x)\,.\end{aligned}$$

This yields

$$\left|\sum_{l=2}^{N} l^k \left(\phi_l^2(x)-1\right)\right| \le \left|\sum_{1\le l\le (N-1)/2} \left((2l+1)^k-(2l)^k\right)\cos(4\pi l x)\right| + N^k$$
$$\le \sum_{1\le l\le (N-1)/2} \left((2l+1)^k-(2l)^k\right) + N^k$$
$$= \sum_{1\le l\le (N-1)/2} \sum_{j=0}^{k-1} \binom{k}{j} (2l)^j + N^k\,.$$

This implies (A.3).

□

Lemma A.3. *For any function $S\in W_r^k$,*

$$\sup_{n\ge 1}\sup_{1\le m\le n-1} m^{2k}\left(\sum_{j=m+1}^{n}\theta_{j,n}^2\right) \le \frac{4r}{\pi^{2(k-1)}}\,. \tag{A.4}$$

Proof. First, note that any function S from W_r^k can be represented by its Fourier series, i.e. $S=\sum_{j=1}^{\infty}\theta_j\phi_j$ with the coefficients defined by (3.4). 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

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

Moreover, it is easy to deduce that

$$\|\Delta_m\|_n^2 = n^{-1}\sum_{k=1}^{n}\Delta_m^2(x_k) = \sum_{k=1}^{n}\int_{x_{k-1}}^{x_k}\Delta_m^2(x_k)\mathrm{d}x$$
$$\le 2\int_0^1\Delta_m^2(x)\mathrm{d}x + 2\sum_{k=1}^{n}\int_{x_{k-1}}^{x_k}(\Delta_m(x_k)-\Delta_m(x))^2\mathrm{d}x\,.$$

The last term in this inequality we estimate as

$$(\Delta_m(x_k) - \Delta_m(x))^2 = \left(\int_x^{x_k} \dot{\Delta}_m(z) \mathrm{d}z \right)^2 \\ \le n^{-1} \int_{x_{k-1}}^{x_k} (\dot{\Delta}_m(z))^2 \mathrm{d}z \,.$$

Therefore,

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

Now note that by the representation of the set W_r^k in the form (3.3) we can estimate the first term in the last inequality as

$$\sum_{j=m+1}^{\infty} \theta_j^2 = \sum_{j=m+1}^{\infty} \theta_j^2 \frac{a_j}{a_j} \le \frac{r}{a_{m+1}} \le \frac{r}{(\pi m)^{2k}} \,.$$

Similarly, we find that

$$\sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2 \le \sup_{j \ge m+1} \frac{\|\dot{\phi}_j\|^2}{a_j} r \le \sup_{j \ge m+1} \frac{\|\dot{\phi}_j\|^2}{\|\phi_j^{(k)}\|^2} r \le \frac{r}{(\pi m)^{2(k-1)}} \,.$$

Therefore, for $m \le n$ we get that

$$\frac{1}{n^2} \sum_{j=m+1}^{\infty} \theta_j^2 \|\dot{\phi}_j\|^2 \le \frac{r}{\pi^{2(k-1)} m^{2k}} \,.$$

This implies (A.4). $\square$

Lemma A.4. *Let $\xi_{j,n}$ be defined in* (2.7) *for the model* (1.1)*. Then, for any real numbers $f_1, \ldots, f_n$,*

$$\mathbf{E} \left(\sum_{j=1}^{n} f_j \, \xi_{j,n} \right)^2 \le \sigma_* \sum_{j=1}^{n} f_j^2 \,. \tag{A.5}$$

Proof. Due to the definition of $\xi_{j,n}$, one has

$$\sum_{j=1}^{n} f_j^2 \, \xi_{j,n} = \sum_{l=1}^{n} \sigma_l \, \widetilde{f}_l \, \xi_l \quad \text{with} \quad \widetilde{f}_l = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} f_j \, \phi_j(x_l) \,.$$

Moreover

$$\begin{aligned}\mathbf{E} \left(\sum_{j=1}^{n} f_j \, \xi_{j,n} \right)^2 &= \sum_{l=1}^{n} \sigma_l^2 \widetilde{f}_l^2 \, \le \, \sigma_* \sum_{l=1}^{n} \widetilde{f}_l^2 \\ &= \sigma_* \sum_{i,j=1}^{n} f_i \, f_j \, (\phi_i, \phi_j)_n \,.\end{aligned}$$

The orthogonality of the basis (ϕ_j) implies inequality (A.5). Hence Lemma A.4. □

Acknowledgements

The authors are grateful to the anonymous Reviewers for a careful reading of manuscript and helpful comments which led to an improved presentation of the paper.

References

[1] M.G.Akritas and I.Van Keilegom *ANCOVA methods for heteroscedastic nonparametric regression models.* J. Amer. Statist. Assoc. **96**(2001) 220-232.

[2] A.Barron, L.Birgé and P.Massart *Risk bounds for model selection via penalization.* Probab. Theory Related Fields. **113**(1999), 301-413.

[3] T.Cai and L.Wang *Adaptive variance function estimation in heteroscedactic nonparametric regression.* Annals of Statistics.(2008), 1-38 accepted.

[4] S.Efromovich and M.Pinsker *Sharp - optimal and adaptive estimation for heteroscedastic nonparametric regression.* Statistica Sinica. **6**(1996), 925-942.

[5] D.Fourdrinier and S.Pergamenshchikov *Improved selection model method for the regression with dependent noise.* Annals of the Institute of Statistical Mathematics. **59** (3)(2007), 435-464.

[6] L.Galtchouk and S.Pergamenshchikov *Nonparametric sequential estimation of the drift in diffusion processes.* Mathematical Methods of Statistics. **13**, 1(2004), 25-49.

[7] L.Galtchouk and S.Pergamenshchikov *Efficient adaptive nonparametric estimation in heteroscedastic regression models.* Preprint of the Strasbourg Louis Pasteur University, IRMA, 2005/020(2005) available at *http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchtchikov*

[8] L.Galtchouk and S.Pergamenshchikov *Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes.* Statistical Inference for Stochastic Process. **9**(2006), 1-16.

[9] L.Galtchouk and S.Pergamenshchikov *Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 1. Sharp non-asymptotic oracle inequalities.* Mathematical Methods of Statistics.(2007a) submitted.

[10] L.Galtchouk and S.Pergamenshchikov *Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 2: Asymptotic efficiency.* Mathematical Methods of Statistics.(2007b), submitted.

[11] L.Galtchouk and S.Pergamenshchikov *Adaptive nonparametric estimation in heteroscedastic regression models. Asymptotic efficiency.* Journal of the Korean Statistical Society.(2007c), submitted.

[12] S.M.Goldfeld and R.E.Quandt *Nonlinear Methods in Econometrics.* North-Holland, London (1972).

[13] J.Kalifa and S.Mallat *Thresholding estimators for linear inverse problems and deconvolutions.* Ann. Statist. **31**(2003), 58-109.

[14] A.Kneip *Ordered linear smoothers.* Ann. Statist. **22**(1994), 835-866.

[15] P.Massart *A non-asymptotic theory for model selection.* 4ECM Stockholm (2004). 309-323.

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

[17] A.Rohde *On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise.* Statistics & Decisions. **22**(2004), 235-243.

[18] S.G.Young and A.W.Bowman *Non-parametric Analysis of Covariance.* Biometrics. **51**(1995), 920–931.